

Physics and Applications of Semiconductor Quantum Structures

**Proceedings of the International Workshop on Physics and
Applications of Semiconductor Quantum Structures
(Asian Science Seminar), Cheju Island, Korea,
October 18–23 1998**

Edited by Takafumi Yao and Jong-Chun Woo

IoP

**Institute of Physics Publishing
Bristol and Philadelphia**

© IOP Publishing Ltd 2001

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency under the terms of its agreement with the Committee of Vice-Chancellors and Principals.

British Library Cataloguing-in-Publication Data

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

ISBN 0 7503 0637 8

Library of Congress Cataloging-in-Publication Data are available

Commissioning Editor: Tom Spicer
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Victoria Le Billon
Marketing Executive: Colin Fenton

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK
US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 1035, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset in Te_X using the IOP Bookmaker Macros
Printed in the UK by MPG Books Ltd, Bodmin

Contents

Preface	xv
PART 1	
Plenary lecture	1
1 The evolution of semiconductor superlattices and quantum nanostructures	
<i>L Esaki</i>	3
1.1 Introduction	3
1.2 Quantum structures	5
References	9
PART 2	
Fabrication and characterization of quantum structures	11
2 Formation and characterization of semiconductor nanostructures	
<i>J Motohisa and T Fukui</i>	13
2.1 Introduction	14
2.2 InGaAs quantum wires on GaAs multatomic steps	15
2.2.1 Quantum nanostructures utilizing atomic steps	15
2.2.2 Formation of GaAs multatomic steps on GaAs (001) vicinal surfaces	17
2.2.3 Fabrication of InGaAs quantum wire structures using multatomic steps	20
2.2.4 Optical characterization of InGaAs quantum wires	24
2.2.5 InGaAs quantum wire lasers	28
2.3 GaAs and InAs quantum dots and related nanostructures by selective area growth	33
2.3.1 MOVPE selective area growth for quantum nanostructures	33
2.3.2 Selective area MOVPE of GaAs and InAs quantum dots on pyramidal structures	34

2.3.3	Fabrication and characterization of high-density quantum dot arrays and quantum dot network	39
2.3.4	Position controlled InAs quantum dots on GaAs pyramidal structures	44
2.3.5	Fabrication of single electron transistors and their transport properties	51
2.3.6	Resistance-load SET inverter circuit	55
2.4	Summary and outlook	58
	Acknowledgments	60
	References	61
3	New quantum wire and quantum dot structures by selective MBE on patterned high-index substrates	
<i>K H Ploog and R Nötzel</i>		65
3.1	Introduction	65
3.2	Results	66
3.2.1	Quantum wires on stripe-patterned GaAs(311)A	66
3.2.2	Dot-like structures on square- and triangular-patterned GaAs(311)A	67
3.2.3	Coupled wire-dot arrays on stripe-patterned GaAs(311)A	68
3.2.4	Uniform quantum dot arrays by hydrogen-enhanced step bunching on stripe-patterned GaAs(311)A	71
3.3	Concluding remarks	72
	Acknowledgments	73
	References	74
4	New fabrication techniques and optical properties of GaN and Si quantum dots	
<i>Y Aoyagi, S Tanaka and X Zhao</i>		75
4.1	Introduction	75
4.2	Formation of self-assembled GaN quantum dots using Si antisurfactant and optical properties of the dots	76
4.2.1	Experimental procedure	76
4.2.2	Experimental results	77
4.2.3	Discussion	80
4.3	Formation of nanocrystalline Si by Er doping and optical properties of the nanocrystal	81
4.3.1	Experimental procedure	82
4.3.2	Experimental results	82
4.4	Conclusion	86
	References	86
5	Dislocation filtering techniques for MBE large mismatched heteroepitaxy	
<i>J M Zhou, Q Huang, H Chen and C S Peng</i>		88

5.1	Introduction	88
5.2	Survey of several dislocation filtering methods	89
5.2.1	Constant-composition layer	89
5.2.2	Strained-layer superlattices (SLS)	90
5.2.3	Composition graded layer	91
5.2.4	Patterned substrate	92
5.3	Low-temperature Si and/or GeSi buffer for GeSi alloy strain relaxation	92
5.4	New initial growth method for pure cubic GaN on GaAs (001)	97
5.5	InAs growth on GaAs using new prelayer technology	100
5.6	Conclusion	103
	References	104

6 Growth of widegap II-VI quantum structures and their optical properties

<i>T Yao</i>	106	
6.1	Introduction	106
6.2	ZnS/(ZnSe) _n /ZnS quantum dots	108
6.2.1	Fabrication of ZnS/(ZnSe) _n /ZnS ultrathin quantum structures and their characterization	108
6.2.2	Temperature dependence of PL of ZnS/ZnSe quantum dot structures	114
6.2.3	Time resolved PL	118
6.2.4	The effect of substrate temperature on the formation of quantum structures	120
6.2.5	Mn doped ZnSe quantum dots	123
6.3	CdSe/ZnSe quantum dots	127
6.3.1	Alloying at the heterointerface	127
6.3.2	Self-organized formation of CdSe quantum dots	130
6.3.3	Spectral diffusion in CdSe QDs	132
6.4	Summary	135
	Acknowledgments	135
	References	136

PART 3

Transport properties in quantum structures

7 Theory of quantum transport in mesoscopic systems: antidot lattices	139	
<i>T Ando, S Uryu, T Nakanishi and S Ishizaka</i>	141	
7.1	Introduction	141
7.2	Antidot lattices	142
7.3	Commensurability peaks	144
7.4	Aharonov-Bohm-type oscillation	149
7.5	Triangular antidot lattices	153

7.6	Altshuler–Aronov–Spivak oscillation	156
7.7	Scattering matrix formalism	158
7.7.1	Quantum-wire junction	158
7.7.2	Energy bands and density of states	160
7.7.3	Commensurability peaks and Aharonov–Bohm oscillation	163
7.8	Anderson localization	166
7.8.1	Experiments	166
7.8.2	Thouless-number method	168
7.8.3	Numerical results	171
7.8.4	Localization oscillation	172
7.9	Summary and conclusion	174
	Acknowledgments	175
	References	175
8	Edge states in magnetic quantum structures and composite fermion systems	
<i>G Ihm, H-S Sim, K J Chang and S J Lee</i>		178
8.1	Introduction	178
8.2	Magnetic quantum dot	180
8.3	Composite fermions in an antidot: application of magnetic quantum dot	184
8.4	Composite fermion edge channels	188
8.5	Conclusion	191
	Acknowledgments	192
	References	192
9	Electronic states in circular and ellipsoidally deformed quantum dots	
<i>S Tarucha, D G Austing, S Sasaki, L P Kouwenhoven, S M Reimann, M Koskinen and M Manninen</i>		194
9.1	Introduction	195
9.2	Experimental details	196
9.3	Electronic properties of circular quantum dots	198
9.3.1	Atom-like properties: shell structure and Hund's first rule	198
9.3.2	Magnetic field dependence	199
9.3.3	Spin triplet for the four-electron ground state	201
9.4	Electronic properties of elliptical quantum dots	203
9.4.1	Deformed dots in rectangular mesa devices	203
9.4.2	Magnetic field dependence	206
9.4.3	Study of Zeeman effect on the spin states	207
9.5	Comparison to model calculations	209
9.6	Conclusions	213
	Acknowledgments	213
	References	214

PART 4	
Optical properties	217
10 Electron–hole and exciton systems in low dimensions	
<i>T Ogawa</i>	219
10.1 Introduction	219
10.2 Remarks on low-dimensional excitons	220
10.3 Single-exciton problem	221
10.3.1 One-photon absorption process	223
10.3.2 Two-photon absorption process	226
10.3.3 Dielectric confinement effect	228
10.4 A few exciton problems	230
10.4.1 An excitonic molecule in one dimension	230
10.4.2 Optical nonlinearity due to excitons	233
10.5 Excitonic many-body problems	237
10.5.1 Exciton Bose–Einstein condensation and density waves in one dimension	237
10.5.2 Optical responses of the Tomonaga–Luttinger liquid: the Mahan exciton	241
10.6 Conclusions and prospect	247
Acknowledgments	250
References	250
11 Size quantization and electron dynamics in nanometre-size semiconductors	
<i>T Goto</i>	253
11.1 Introduction	253
11.2 Two-dimensional size confinement in ultrathin PbI ₂ nanocrystals	254
11.3 Size confinement of an exciton internal motion in ultrathin PbI ₂ nanocrystals	258
11.4 Nonlinear optical processes in CuCl nanocrystals	259
11.5 Photoinduced phenomena in luminescence spectra of a single CdSe nanocrystal	267
References	271
12 Electronic properties of InAs/GaAs quantum dots	
<i>R Heitz, I Mukhametzhanov, O Stier, A Hoffmann, A Madhukar and D Bimberg</i>	273
12.1 Introduction	273
12.2 Experimental	275
12.3 Experimental results and discussion	277
12.3.1 Energy transfer processes	278
12.3.2 Temperature dependence	282
12.3.3 Local-equilibrium emission	286
12.3.4 Excited-state transitions	291

12.3.5 Phonon-assisted exciton recombination	293
12.4 Conclusions	296
Acknowledgments	297
References	297
13 Graded and coupled quantum wells for emission of radiation by intersubband emission	
<i>K D Maranowski, A C Gossard and K L Chapman</i>	300
13.1 Introduction	300
13.2 Review of intersubband emission	302
13.3 Emission from parabolic quantum wells	307
13.4 Optical excitation of parabolic quantum wells	313
13.5 Quantum interference effects	315
13.5.1 Absorption in coupled quantum wells	317
13.5.2 Interference of intersubband transitions	321
13.6 Conclusions	331
Acknowledgments	333
References	333
PART 5	
Spin-dependent properties	335
14 Spin-dependent properties of magnetic III–V semiconductors	
<i>H Ohno</i>	337
14.1 Introduction	337
14.2 Preparation and basic properties of (Ga,Mn)As	338
14.2.1 Molecular beam epitaxy	339
14.2.2 Lattice constant and local lattice configuration	339
14.3 Magnetic and magnetotransport properties	340
14.3.1 Magnetic properties	340
14.3.2 Magnetotransport properties	341
14.4 Origin of ferromagnetism	345
14.5 Heterostructures	347
14.5.1 Ferromagnetic/non-magnetic/ferromagnetic trilayers	347
14.5.2 Resonant tunnelling structures	351
14.6 Conclusion	352
Acknowledgments	352
References	352
15 Zeeman separation in GaAs quantum structures	
<i>J-C Woo and K-H Yoo</i>	355
15.1 Introduction	355
15.2 Zeeman effects of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ bulk	356
15.3 GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum well	357
15.4 GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum dot and quantum wire	362

15.5 Concluding remarks	366
References	366
PART 6	
Device applications	369
16 Self-assembling of silicon quantum dots and its application to novel nanodevices	
<i>M Hirose</i>	371
16.1 Introduction	371
16.2 Self-assembling of silicon quantum dots	372
16.3 Quantum confinement in silicon dot	375
16.3.1 Valence band spectra and charging effect	375
16.3.2 Comparison with theory	379
16.4 Resonant tunnelling through a Si quantum dot	380
16.4.1 Device fabrication	380
16.4.2 Resonant tunnelling characteristics	381
16.5 Silicon quantum dot memory	382
16.5.1 Device fabrication	383
16.5.2 Memory characteristics of Si quantum dot floating gate MOS structures	383
16.6 Summary	385
Acknowledgment	385
References	385
17 Quantum devices based on III–V compound semiconductors	
<i>H Hasegawa, H Fujikura and H Okada</i>	387
17.1 Introduction	388
17.2 Brief general overview on compound semiconductor quantum devices	389
17.2.1 Wave-particle duality and devices	389
17.2.2 Quantum wave devices	389
17.2.3 Single electron devices	391
17.3 Formation of GaAs- and InP-based quantum structure by molecular beam epitaxy	395
17.3.1 Approaches for formation of quantum structures	395
17.3.2 Selective MBE growth of InGaAs quantum wires, dots and wire-dot coupled structures	396
17.4 Control of surfaces and interfaces of nanostructures	405
17.4.1 A key issue for nanofabrication	405
17.4.2 Formation technology of damage-free Schottky gates for quantum structures	406
17.5 Compound semiconductor quantum wave devices	413

17.5.1 Various lateral quantum wave devices investigated at RCIQE	413
17.5.2 Schottky IPG single and coupled quantum wire transistors	413
17.6 Compound semiconductor single electron devices	418
17.6.1 Key issues for SEDs and efforts at RCIQE	418
17.6.2 Voltage gain in GaAs-based lateral SETs having novel Schottky gates	420
17.6.3 Single electron memory devices utilizing IPG quantum wire transistor and metal nanodots	425
17.7 Conclusion	429
Acknowledgments	429
References	429
18 Quantum interference in corrugated conducting wire transistors	433
K Park	433
18.1 Issues and challenges in semiconductor nanostructures	433
18.2 Quantum transport; electron interference	436
18.3 Review on lateral quantum interference devices	438
18.4 Various characteristics of quantum interference devices	440
18.4.1 Uneven gate conducting wire	440
18.4.2 Electrostatic Aharonov–Bohm ring transistor	442
18.4.3 Electron diffraction transistor	446
18.4.4 One-dimensional corrugated quantum channel	448
18.5 Summary	449
Acknowledgments	450
References	450
19 Non-Markovian optical gain of strained-layer quantum-well lasers with many-body effects	452
D Ahn	452
19.1 Introduction	452
19.2 The model	453
19.3 Numerical results and discussion	459
19.4 Conclusions	463
Acknowledgment	463
References	463
20 Generation of non-classical lights from semiconductor light emitters	465
M Yamanishi, H Sumitomo, M Kobayashi and Y Kadoya	465
20.1 Introduction	465
20.2 Squeezing in photon-number fluctuations under constant-current operation	466
20.2.1 Basic physics for constant-current operation	466
20.2.2 Experimental arrangement and results (constant-current operation)	468

<i>Contents</i>	xiii
20.2.3 Pump-current-dependence of squeezing bandwidth	471
20.2.4 Framework of backward-pump model	471
20.2.5 Theoretical bandwidth for noise suppression in comparison with experimental results	474
20.3 Squeezing in photon-number fluctuations due to the backward pump process	476
20.3.1 Basic physics for constant-voltage operation	476
20.3.2 Experimental arrangements and results (constant-voltage operation)	477
20.3.3 New squeezing mechanism based on the backward pump process	480
20.4 Conclusions	482
References	483

Preface

Since the revolutionary idea of ‘artificial’ semiconductor superlattice structures proposed by Esaki and Tsu in 1969, semiconductor quantum structures have opened up a new era not only in the research of physics and materials science, but also in the development of electronic and optical devices. In the initial stage of this field, study was limited to the two-dimensionally confined electron system, which led to the discovery of many novel phenomena such as quantum Hall effects, quantum confined Stark effects, Aharonov–Bohm effects. It is obvious that these discoveries are due to advances in materials science, in particular the atomic-scale growth techniques such as molecular beam epitaxy and metal organic chemical vapour deposition. Well-known examples of application are quantum well lasers, high electron mobility transistors and resonant tunnelling diodes, which are based on the two-dimensional confinement of electrons.

In recent years, efforts in this field have been devoted to the fabrication and characterization of quantum structures with reduced dimensionality, namely one- and zero-dimensional structures, with remarkable advancement in material processing and micro-fabrication technology. The successful fabrication of quantum wire and quantum dot structures have enabled scientists to explore novel properties and new-concept devices. Some of the outstanding examples are Coulomb blockade effects, microcavity lasers, exciton-based nonlinear optical effects, and single-electron transistors.

In view of the rapid progress in this multidisciplinary area related to semiconductor quantum structures, the researchers in this field felt a need to hold a forum, where scientists in various backgrounds could get together to review recent achievements and to discuss the future directions of development. In order to fulfil such a demand, the 1998 Asian Science Seminar entitled the International Workshop on Physics and Application of Semiconductor Quantum Structures has been formed under the sponsorship of the Japan Society for the Promotion of Science (JSPS) and the Korea Science and Engineering Foundation (KOSEF).

In this workshop, experts and leading scientists were invited to cover the overall spectrum on the research activities in this field. This book is comprised of the invited lectures of this workshop and a number of reviews. This book starts with a perspective review on the evolution of semiconductor superlattices and quantum nanostrctures ([part 1](#)) followed by the fabrication and characterization

of quantum structures ([part 2](#)), transport properties ([part 3](#)), optical properties ([part 4](#)), spin dependent properties ([part 5](#)), and device applications ([part 6](#)).

We would like to extend our sincere gratitude to the contributing authors of the articles, the organizing members of the workshop, and especially the sponsoring organizations of the workshop for their kind cooperation and support.

Takafumi Yao and Jong-Chun Woo

February 2000

PART 1

PLENARY LECTURE

Chapter 1

The evolution of semiconductor superlattices and quantum nanostructures

*Leo Esaki
Tsukuba, Ibaraki 305-0047, Japan*

In the early twentieth century, encounters with physical phenomena which require detailed analyses in nanoscale, such as electron motion, prompted the advent of quantum mechanics, since Newtonian mechanics could not possibly provide an adequate explanation for them. Electron tunnelling through nanoscale barriers is the most direct consequence of the law of quantum mechanics, for which the Esaki tunnel diode gave most convincing experimental evidence in 1957. Following the evolutionary path of quantum nanostructures, significant milestones are presented, including the birth of semiconductor superlattices, resonant tunnel diodes, quantum wires and dots.

1.1 Introduction

The twentieth century will be characterized by the fact that science and technology have made remarkable progress, including the establishment of quantum mechanics, the development of semiconductor devices with the invention of the transistor and the evolution of computers/telecommunications.

In the early century, encounters with physical phenomena such as the electron's motion or the photon's behaviour for which Newtonian mechanics could not possibly provide an adequate explanation, prompted the advent of quantum mechanics. The framework of quantum mechanics was established in the superb work of Werner Heisenberg, Erwin Schrödinger, Paul Dirac and Max Born in the period 1925–6.

During the infancy of the quantum theory, de Broglie [1] introduced a new fundamental hypothesis that matter was endowed with a dualistic nature—particles may also have the characteristics of waves. This hypothesis found

expression, in the hands of Schrödinger [2], in the definite form now known as the Schrödinger wave equation, whereby an electron is assumed to be represented by a solution to this equation. The continuous non-zero nature of such solutions, even in classically forbidden regions of negative kinetic energy, implies an ability to penetrate such forbidden regions and a probability of tunnelling from one classically allowed region to another. The concept of tunnelling itself arises from this quantum-mechanical result, and has no analogy in classical mechanics. The subsequent experimental manifestations of that concept can be regarded as one of the early triumphs of the quantum theory. For instance, in 1928, Fowler and Nordheim [3] explained, on the basis of electron tunnelling, the main features of the phenomenon of electron emission from cold metals by high external electric fields, which had been unexplained since its observation by Lilienfeld in 1922.

In 1932, Wilson [4], Frenkel and Joffe [5], and Nordheim [6] applied quantum mechanical tunnelling to the interpretation of metal–semiconductor contact rectifiers such as those made from selenium or cuprous oxide. Apparently, this theory was accepted for a number of years until it was finally discarded after it was realized that it predicted rectification in the wrong direction for ordinary, practical diodes. It is now clear that, in the usual circumstances, the surface barriers met by semiconductors in contact with metals are far too thick to observe any tunnelling current.

In 1934, the development of the energy-band theory of solids prompted Zener [7] to propose interband tunnelling as an explanation for dielectric breakdown. He calculated the rate of transitions from a filled band to a next-higher unfilled band by the application of an electric field. In effect, he showed that an energy gap could be treated in the manner of a potential barrier. The Zener mechanism in dielectric breakdown, however, has never been proved to be important in reality. If a high electric field is applied to the bulk crystal of a dielectric or semiconductor, avalanche breakdown (electron–hole pair generation) generally precedes tunnelling, and thus the field never reaches a critical value for tunnelling.

With the invention of the transistor in 1947 came a renewed interest in the tunnelling process. Around 1950, the technology of Ge p–n junction diodes was developed, and efforts were made to understand the junction properties. In explaining the reverse-bias characteristic, McAfee *et al* [8] applied a modified Zener theory and asserted that low-voltage breakdown in Ge diodes resulted from interband tunnelling. Results of later studies, however, indicated that most Ge junctions broke down by avalanche, but by that time the name ‘Zener diodes’ had already been given to the low-breakdown Si diodes. Actually, Zener diodes are almost always avalanche diodes.

In these circumstances, in 1956, the investigation of interband tunnelling was initiated with heavily-doped Ge p–n junctions, where the junction width was successfully reduced to the range of nanometres.

We first obtained a backward diode which was more conductive in the reverse direction than in the forward direction. In this respect it agreed with

the rectification direction predicted by the previously-mentioned old tunnelling rectifier theory. The estimated junction width at zero bias was approximately 20 nm. In this junction, the possibility of an avalanche was completely excluded because the breakdown occurs at a voltage much lower than the threshold for electron–hole pair production.

By further narrowing the junction width to less than 10 nm (thereby further decreasing the length of the tunnelling path), through a further increase in the doping level, in 1957, the Esaki tunnel diode [9, 10] with a negative resistance was obtained. The characteristic of this diode gave convincing evidence for tunnelling electrons in solids, a phenomenon which had been clouded by questions for decades. This diode, together with the backward diode, constituted the first semiconductor quantum electron structure.

1.2 Quantum structures

In 1969, research on artificially structured materials was initiated when Esaki and Tsu [11, 12] proposed an engineered semiconductor superlattice with a one-dimensional periodic potential, where the period in the range of nanometres is substantially longer than the lattice constant of host crystals but shorter than the electron mean free path or the electron phase-coherent length.

Before arriving at the superlattice concept, we had been examining the feasibility of structural formation of potential barriers and wells that were thin enough to exhibit resonant tunnelling [13]. A resonant tunnel diode [14, 15] appeared to have more spectacular characteristics than the single tunnel barrier Esaki diode. It was thought that advanced technologies using semiconductors might be ready for demonstration of the de Broglie electron waves. Actually, resonant tunnelling can be compared to the transmission of an electromagnetic wave through a Fabry–Perot resonator; the equivalent of a Fabry–Perot resonant cavity is formed by the semiconductor potential-well sandwiched between two potential barriers.

The idea of the superlattice occurred to us as a natural extension of double-, triple- and multiple-barrier structures: the superlattice consists of a series of potential wells coupled by resonant tunnelling. An important parameter for the observation of quantum effects in the structure is the phase-coherence length, which approximates to the electron mean free path. This depends on the bulk quality as well as the interface quality of crystals, and also on the temperatures and the effective mass. As schematically illustrated in [figure 1.1](#), if the characteristic dimensions such as superlattice periods or well widths are reduced below the phase-coherent length, the entire electron system will enter a mesoscopic quantum regime of low dimensionality, on a scale between the macroscopic and microscopic. Our proposal was to explore quantum effects in the mesoscopic regime.

[Figure 1.2](#) illustrates schematically the evolutionary path, starting with the

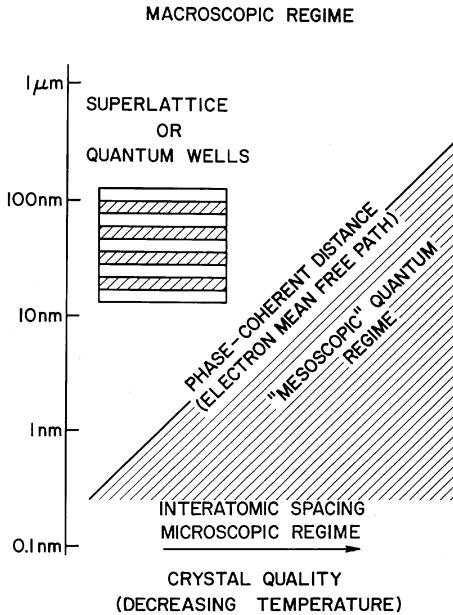


Figure 1.1. Schematic illustration of a ‘mesoscopic’ quantum regime (shaded) with superlattice or quantum wells (inset).

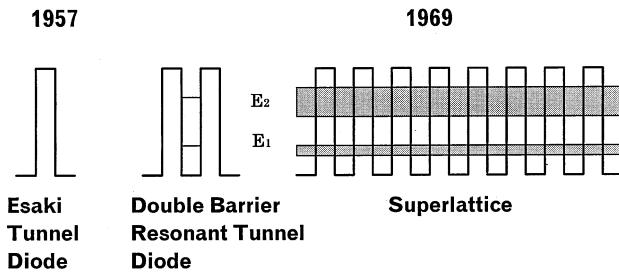


Figure 1.2. From left to right, the Esaki tunnel diode, a resonant tunnel diode and a superlattice.

Esaki diode of a single potential barrier, then moving to the double-barrier resonant tunnel diode and, finally, reaching the superlattice. The introduction of the one-dimensional superlattice potential perturbs the band structure of the host materials, yielding a series of narrow subbands and forbidden gaps which arise from the subdivision of the Brillouin zone into a series of minizones. Thus, the superlattice was expected to exhibit unprecedented electronic properties. This could be considered to be modern alchemy, for it is intended to transform ‘common’ semiconductors into ‘super’ semiconductors.

At the inception of the superlattice idea, it was recognized that the utilization of long, tailor-made lattice periods provided a unique opportunity to exploit electric field-induced effects. The electron dynamics in the superlattice direction was analysed for conduction electrons in a narrow subband of a highly perturbed energy–wavevector relationship. The result led to the prediction of a negative differential resistance at a modestly high electric field, which could be a precursor of Bloch oscillations. The superlattice allows us to enter the regime of electric-field-induced quantization: the formation of Stark ladders [16, 17], for example, can be proved in a (one-dimensional) superlattice [18], whereas in natural (three-dimensional) crystals the existence and nature of these localized states in a high electric field have been controversial [19, 20].

This was, perhaps, the first proposal which advocated engineering a new semiconductor material by applying the principles of quantum theory and using the most advanced crystal-growth techniques of the day. The proposal was made to the US Army Research Office (ARO), a funding agency, in 1969. It was daringly stated, but with little confidence in a successful outcome at the time: ‘...the study of superlattices and observations of quantum mechanical effects on a new physical scale may provide a valuable area of investigation in the field of semiconductors’.

This attempt is considered to be modern alchemy which transforms common semiconductors to super man-made materials in use of quantum mechanics with advanced techniques, whereas traditional alchemy transforms base metals to noble metals with magic or witchcraft. It is of significance that materials prepared by modern alchemy, indeed, exhibit salient properties which do not exist in nature.

Although this proposal was favourably received by ARO, the original version of the paper [11] was rejected for publication by *Physical Review* on a referee’s unimaginative assertion that it was ‘too speculative’ and involved ‘no new physics’. The shortened version published in the *IBM Journal of Research and Development* [12] was subsequently selected as a Citation Classic by the Institute for Scientific Information (ISI) in July 1987. Our 1969 proposal was also cited as one of the most innovative ideas at the ARO 40th Anniversary Symposium in Durham, North Carolina, in 1991.

In any case, with the proposal we launched a program to make a ‘Gedanken-experiment’ a reality. In some circles, the proposed material was criticized as being close to impossible to prepare. One of the objections was that a man-made structure with compositional variations of the order of several nanometres could not be thermodynamically stable because of interdiffusional effects. Fortunately, however, it turned out that interdiffusion was negligible at the temperatures involved.

In 1970, Esaki *et al* [21] studied a GaAs–GaAs_{0.5}P_{0.5} superlattice with a period of 20 nm synthesized by CVD (chemical vapour deposition) by Blakeslee and Aliotta [22]. Although transport measurements failed to reveal any predicted effect, the specimen probably constituted the first strained-layer superlattice having a relatively large lattice mismatch. Early efforts in our

group to obtain epitaxial growth of $\text{Ge}_{1-x}\text{Si}_x$ and $\text{Cd}_{1-x}\text{Hg}_x\text{Te}$ superlattices were soon abandoned because of rather serious technical problems at that time. Instead, we focused our research effort on compositional $\text{GaAs}-\text{Ga}_{1-x}\text{Al}_x\text{As}$ superlattices grown by MBE (molecular beam epitaxy). In 1972, we found a negative resistance in such superlattices [23], which was interpreted in terms of the superlattice effect.

Following the derivation of the voltage dependence of resonant tunnel currents [15], Chang *et al* [24] observed current-voltage characteristics with a negative resistance. Subsequently, Esaki and Chang [25] measured quantum transport properties in a superlattice with a narrow bandwidth, which exhibited an oscillatory behaviour. Tsu *et al* [26] also performed photocurrent measurements on superlattices subjected to an electric field perpendicular to the plane layers using a semitransparent Shottky contact, which revealed their miniband configurations.

Heteroepitaxy is of great interest for the growth of compositional superlattices. Innovations and improvements in epitaxial techniques such as MBE and MOCVD (metal-organic chemical vapour deposition) have made it possible to prepare high-quality heterostructures with predesigned potential profiles and impurity distributions having dimensional control close to interatomic spacing. This great precision has created our freedom of access to the mesoscopic quantum regime [27, 28].

Since a one-dimensional potential can be introduced along the direction of growth, famous examples in the history of one-dimensional mathematical physics, including the above-mentioned resonant tunnelling [13], Kronig–Penney bands [29], Tamm surface states [30], Zener band-to-band tunnelling [7] and Stark ladders including Bloch oscillations [16–18], all of which had been only textbook exercises, could, for the first time, be practiced in a laboratory. Thus, do-it-yourself quantum mechanics is now possible, since its principles dictate the details of semiconductor structures [31].

Our original proposal [11, 12] and pioneering experiments have triggered a wide spectrum of experimental and theoretical investigations on superlattices, quantum wells and quantum wires and dots [32] over the last two decades. A variety of engineered structures now exhibit extraordinary transport and optical properties which do not exist in any known natural crystal. This new degree of freedom offered in semiconductor research through advanced materials engineering has inspired many ingenious experiments, resulting in observations not only of predicted effects but also of phenomena hitherto totally unknown.

The number of papers related to the subject presented at the biennial International Conference on the Physics of Semiconductors has increased tremendously. After 1972, when the first paper [23] was presented, the field went through a short period of incubation before experiencing a phenomenal expansion in the 1980s. It appears that nearly half of all semiconductor physicists in the world are working in this area. Activity at this new frontier of semiconductor physics has in turn given immeasurable stimulus to device physics, provoking

innumerable new ideas for applications. Thus, a new class of transport and optoelectronic devices has emerged.

At the opening address of the 23rd ICPS (Berlin, 1996), G Landwehr [33] stated, ‘...the impact of two-dimensional systems was highly underestimated by most people at that time except, of course, Leo Esaki who proposed—together with Raphael Tsu—the superlattice concept which has given semiconductor physics an enormous impetus’.

References

- [1] de Broglie L 1923 Waves and quanta *Nature* **112** 540
- [2] Schrödinger E 1926 Quantisierung als eigenwertproblem *Ann. Phys., Lpz.* **4**(79) 489
- [3] Fowler R H and Nordheim L 1928 Electron emission in intense electric fields *Proc. R. Soc. A* **119** 173
- [4] Wilson A H 1932 A note on the theory of rectification *Proc. R. Soc. A* **136** 487
- [5] Frenkel J and Joffe A 1932 On the electric and photoelectric properties of contacts between a metal and a semiconductor *Phys. Z. Sowjetunion* **1** 60
- [6] Nordheim L 1932 Zur theorie der detektorwirkung *Z. Phys.* **75** 434
- [7] Zener C 1934 A theory of the electrical breakdown of solid dielectrics *Proc. R. Soc.* **145** 523
- [8] McAfee K B, Ryder E J, Shockley W and Sparks M 1951 Observations of Zener current in germanium p–n junction *Phys. Rev.* **83** 650
- [9] Esaki L 1958 New phenomenon in narrow germanium p–n junctions *Phys. Rev.* **109** 603
- [10] Esaki L 1960 Properties of heavily-doped germanium and narrow p–n junctions *Solid State Physics in Electronics and Telecommunications (Proc. Int. Conf., Brussels, 1958) Vol. I Semiconductors part I*, ed M Desirant and J L Michels (New York: Academic) p 514
- [11] Esaki L and Tsu R 1969 Superlattice and negative conductivity in semiconductors *IBM Research Notes RC-2418*
- [12] Esaki L and Tsu R 1970 Superlattice and negative differential conductivity in semiconductors *IBM J. Res. Dev.* **14** 61
- [13] Bohm D 1951 *Quantum Theory* (Englewood Cliffs, NJ: Prentice-Hall) p 283
- [14] Iogansen L V 1964 The possibility of resonance transmission of electrons in crystals through a system of barriers *Sov. Phys.–JETP* **18** 146
Iogansen L V 1963 *Zh. Eksp. Teor. Fiz.* **45** 207
- [15] Tsu R and Esaki L 1973 Tunnelling in a finite superlattice *Appl. Phys. Lett.* **22** 562
- [16] James H M 1949 Electronic states in perturbed periodic systems *Phys. Rev.* **76** 1611
- [17] Wannier G H 1959 *Elements of Solid State Theory* (Cambridge: Cambridge University Press) p 190
Wannier G H 1960 Wave functions and effective Hamiltonian for Bloch electrons in an electric field *Phys. Rev.* **117** 432
- [18] Shockley W 1972 Stark ladders for finite, one-dimensional models of crystals *Phys. Rev. Lett.* **28** 349
- [19] Zak J 1968 Stark ladder in solids? *Phys. Rev. Lett.* **20** 1477

- Zak J 1991 Comment on the existence proofs of the Wannier–Stark ladder *Phys. Rev. B* **43** 4519
- [20] Rabinovitch A and Zak J 1971 Electrons in crystals in a finite-range electric field *Phys. Rev. B* **4** 2358
- [21] Esaki L, Chang L L and Tsu R 1970 A one-dimensional ‘superlattice’ in semiconductors *Proc. 12th Int. Conf. Low Temp. Phys. (Kyoto)* p 551
- [22] Blackslee A E and Aliotta C F 1970 Man-made superlattice crystals *IBM J. Res. Dev.* **14** 686
- [23] Esaki L, Chang L L, Howard W E and Rideout V L 1972 Transport properties of a GaAs–GaAlAs superlattice *Proc. 11th Int. Conf. on the Physics of Semiconductors (Warsaw)* p 431
- [24] Chang L L, Esaki L and Tsu R 1974 Resonant tunnelling in semiconductor double barriers *Appl. Phys. Lett.* **24** 593
- [25] Esaki L and Chang L L 1974 New transport phenomenon in a semiconductor superlattice *Phys. Rev. Lett.* **33** 495
- [26] Tsu R, Chang L L, Sai-Halasz G A and Esaki L 1975 Effects of quantum states on the photocurrent in a ‘superlattice’ *Phys. Rev. Lett.* **34** 1509
- [27] Esaki L 1986 A bird’s-eye view on the evolution of semiconductor superlattices and quantum wells *IEEE J. Quantum Electron.* **QE-22** 1611
- [28] Esaki L 1991 Implications of semiconductor superlattice research *Highlights in Condensed Matter Physics and Future Prospects* ed L Esaki (New York: Plenum) p 55
- [29] Kronig R de L and Penney W G 1931 Quantum mechanics of electrons in crystal lattices *Proc. R. Soc. A* **130** 499
- [30] Tamm I 1932 Über eine mögliche art der elektronen-bindung *Phys. Z. Sowjetunion* **1** 733
- [31] Esaki L 1992 Do-it-yourself quantum mechanics in low-dimensional structures *Phys. Scr. T* **42** 102
- [32] Tarucha S, Austing D G, Honda T, van der Hage R J and Kouwenhoven L P 1996 Shell filling and spin effects in a few electron quantum dot *Phys. Rev. Lett.* **77** 3613
- [33] Landwehr G 1996 Opening address *Proc. 23rd Int. Conf. on the Physics of Semiconductors (Berlin)* (Singapore: World Scientific) p xi

PART 2

FABRICATION AND CHARACTERIZATION OF QUANTUM STRUCTURES

Chapter 2

Formation and characterization of semiconductor nanostructures

Junichi Motohisa¹ and Takashi Fukui

Research Center for Interface Quantum Electronics, Hokkaido University, North 13 West 8, Sapporo 060-8628, Japan

¹ *E-mail: motohisa@rciqe.hokudai.ac.jp*

We describe the fabrication of quantum wire (QWR) and quantum dot (QD) structures by metal-organic vapour phase epitaxial (MOVPE) growth, their characterization of optical and electronic properties and their application to quantum devices. Firstly, multiaatomic steps formed during MOVPE growth on vicinal (001) GaAs surfaces were used to introduce lateral confinement potential and to realize QWRs. Based on the understanding of the formation process of GaAs and InGaAs multiaatomic steps on vicinal (001) GaAs surfaces, InGaAs/GaAs QWRs were successfully fabricated on multiaatomic steps and their one-dimensional nature was verified by comprehensive study of their optical properties. Application of QWRs to lasers was also demonstrated. Next, GaAs and InAs quantum dot and related nanostructures were fabricated by selective area (SA)-MOVPE growth. Zero-dimensional character of QDs formed by SA-MOVPE was demonstrated by the ultra-narrow PL peak from InAs QDs. Quantum dots connecting with GaAs quantum wires were also fabricated and they were applied to GaAs single electron transistors (SETs) and resistance-load single electron inverter circuit. Clear Coulomb blockade type conductance oscillations were observed near the pinch-off voltage of SETs at low temperatures. The inverter operation of a circuit using this SET together with load resistance was also demonstrated.

2.1 Introduction

Semiconductor quantum nanostructures where the carriers are confined quantum mechanically in the nanometre regions have been attracting much attention. This is because the semiconductor nanostructures exhibit unique electrical and optical properties which are completely different from conventional bulk semiconductors. Their application is also expected to realize, high-performance novel quantum functional devices.

To study and explore their physics and, especially, the possibility of practical device applications, one of the important issues is how one can realize quantum nanostructures which satisfy some requirements, such as in their size, uniformity, electrical or optical quality, and so on, so that quantum effects arising from the wave nature of electrons can clearly be observed. This is partly proved in two-dimensional electron systems (2DES) and quantum well (QW) or superlattice (SL) structures, where quantum mechanical confinement of electrons and holes are realized in one direction. That is, stacks of thin semiconductor films (semiconductor heterostructures) has opened up a new paradigm in the device application as well as the physics of 2DES and QW structures, and realization of high-quality semiconductor heterostructures are due greatly to advanced epitaxial growth techniques, such as molecular beam epitaxy (MBE) and metal-organic vapour phase epitaxy (MOVPE). Now, prompted by the predictions of Sakaki [2] and Sakaki and Arakawa [3], the interests have been taken to achieve carrier confinement in more than two dimensions and to realize quantum wires (QWRs), quantum dots (QDs), and related nanostructures such as lateral surfaces superlattices (LSSL). Their fabrication issue has become more and more important as the requirements are much more strict than those in QWs.

Various kinds of techniques have been reported and demonstrated to realize quantum nanostructures. One of the approaches is to utilize micro- or nanofabrication techniques with lithography, etching, and/or ion implantation. Here, quantum wells or heterostructures are used as starting materials for the nanofabrication, and lateral confinement in wire or dot structures are realized by etching after those patterns are defined, for example, by electron beam (e-beam) lithography. Although this approach has disadvantages that the size for the lateral confinement is limited by the lithography, and that the process-induced damages particularly at the etched-sidewalls are problematic for the quality of the devices, some of the unique electronic states and transport properties in low-dimensional electron systems are disclosed [4]. Manipulation of atoms and atomic scale control of their chemical reactions has also been attempted recently using scanned probe techniques to go beyond the size limit in the conventional lithography.

There is also a completely different kind of approach in which quantum nanostructures are directly formed during crystal growth. This has become possible owing to the advancement of MBE and MOVPE growth techniques and the understanding of their growth mechanism. Up to now, a variety of quantum nanostructures are realized by using such crystal growth techniques, which can be

classified as follows.

- (1) Quantum wire arrays using atomic steps as templates for the lateral compositional or thickness modulation [5–13].
- (2) Selective area growth of QWRs and QDs on substrates partially masked by SiO₂ or SiN_x [14–27].
- (3) Growth on substrates with patternings, such as V-grooves or mesa-patterns [28–39].
- (4) Self-assembled formation of QDs, for example, by droplet epitaxy [40], by using islanding in Stranski–Krastanow growth mode in strained layer growth [41–51], by self-organization of strained layer on high-index surfaces [52], or by strain-induced phase separation [53–55].

In all these methods, high-quality quantum nanostructures can be realized with sufficiently small size and high density. In addition, it is possible to realize heterostructures so that electrons and holes can be confined with abrupt interfaces. Thus, such direction formation methods using crystal growth are thought to be one of the most promising techniques for the formation of semiconductor nanostructures, and in fact, one- and zero-dimensional properties of carriers in QWRs and QDs and their device applications are investigated and reported with nanostructures formed by the above methods [28, 56–70].

In this paper, we will describe our two approaches, namely (1) and (2) in the above categories, by MOVPE growth for the formation of high-density and high-quality QWRs, QDs, and related quantum nanostructures. We will show that the variety of nanostructures can be realized by MOVPE growth. We will also describe their characterization and device application. In section 2.2, growth, characterization, and application of InGaAs QWRs utilizing GaAs multiaatomic steps in GaAs (001) vicinal surfaces will be described. We first briefly explain the formation of multiaatomic steps during MOVPE growth on vicinal GaAs (001) substrates. Then we will describe the fabrication and optical properties of InGaAs QWRs using multiaatomic steps. Quantum wire lasers using InGaAs QWRs in their active regions are also demonstrated and will be described. Quantum nanostructures formed by selective area (SA) MOVPE growth are described in section 2.3. In this section, we will first describe the growth process of SA-MOVPE, and formation and optical properties of GaAs and InAs QDs on pyramidal structures. Fabrication and basic properties of single electron transistors and circuits will be described. Section 2.4 gives a summary and an outlook of our study.

2.2 InGaAs quantum wires on GaAs multiaatomic steps

2.2.1 Quantum nanostructures utilizing atomic steps

The first approach we are going to describe is to use atomic steps as templates to form QWRs. One of the advantages in this kind of approach is its simplicity

that high-density arrays of QWRs can be formed spontaneously during growth in a self-organized way at the atomic step edges. Therefore, once we prepare vicinal substrates, we do not need any artificial lithographic processes, and can realize nanostructures free from process-induced damages. Furthermore, the size and density of these structures can be controlled by only adjusting the crystal growth conditions and the misorientation angle of the vicinal substrates.

This kind of approach was first proposed and attempted by Petroff *et al* [5] with MBE, and later with MOVPE by Fukui *et al* [6], and formation of fractional layer superlattices or quantum wire arrays was reported. Ideally, atomic steps do not exist on low-index surfaces. However, if the substrates are slightly mis-cut from exact (001) orientation, the surface consists of atomic terraces and steps. For *monoatomic* steps, the average terrace width L is given by $a \cot \alpha$, where a is the thickness of one monolayer (ML, 0.283 nm for (001) GaAs) and α is the misorientation angle. This gives us $L = 16\text{--}8$ nm for $\alpha = 1\text{--}2^\circ$. If we can control the growth mode so that the growth proceeds from the step edges (step-flow growth mode), lateral compositional modulation of GaAs and AlAs can be realized. Thus, it is effective to introduce lateral periodic structures in the nanometre scale in the direction normal to the growth. Although the use of *monoatomic* step is quite difficult as it requires precise control of growth rate as well as growth mode, anisotropic electronic states were confined both in transport [71, 72] and optical properties [73, 74], and devices such as lasers [56, 73] and electron wave interference transistors [75] were reported.

While formation of those quantum nanostructures utilizing monoatomic steps were attempted, it was found that monoatomic steps on vicinal GaAs (001) surfaces got together during MOVPE growth, and as a result, coherent array of *multiaatomic* steps with fairly regular spacings were formed [8]. Then we proposed and demonstrated QWR structures using multiaatomic steps. Figure 2.1 schematically shows such QWR structures, where thickness modulated InGaAs layers are formed on GaAs multiaatomic steps which give rise to lateral quantum confinement. We have reported on the spontaneous formation of the multiaatomic steps during metal-organic vapour phase epitaxial (MOVPE) growth on vicinal GaAs (001) surfaces [76], and their application to GaAs/AlGaAs quantum wires [11, 77] and electron wave interference devices [78, 79]. In addition, by growing InGaAs on GaAs multiaatomic steps, we also have reported on the self-organized formation of InGaAs QWRs of figure 2.1.

There are also some attempts to form quantum nanostructures using steps on different surfaces from vicinal (001) GaAs. In these kinds of attempts, surfaces after step bunching or faceting takes place are used in a similar way to grow a thickness modulated well layer, for example, on vicinal (001) InP surfaces [12], vicinal (110) GaAs [9, 10] and (111)B GaAs [13] surfaces. Step bunching or faceting are also observed in some high-indexed surfaces, for example, (311)A GaAs [81], (331)A [82], and (775)B GaAs [83], which is thought to be originated from the instability of the high-indexed surfaces. These phenomena are also used to form high-density arrays of quantum wires and quantum dots.

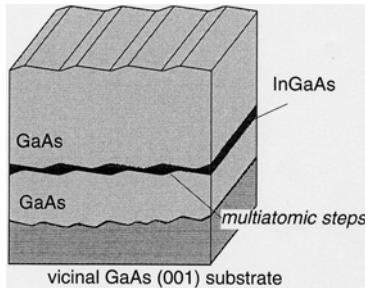


Figure 2.1. Schematic illustration of quantum wires formed on GaAs multiaatomic steps.

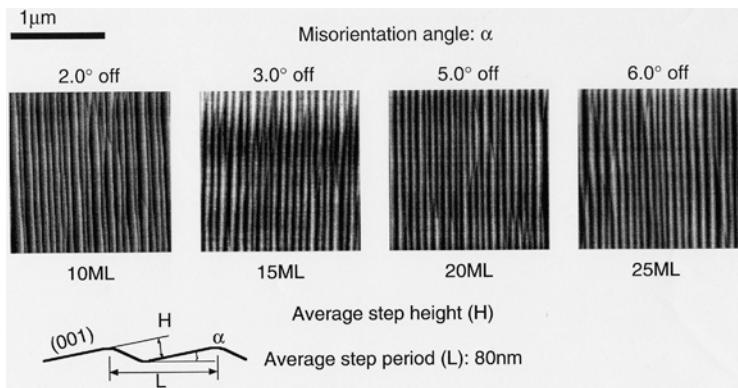


Figure 2.2. Typical atomic force microscope images of GaAs surfaces after MOVPE growth.

2.2.2 Formation of GaAs multiaatomic steps on GaAs (001) vicinal surfaces

When GaAs is grown on a vicinal GaAs (001) surface, monoatomic steps on the initial surface start to get together, and as a result, multiaatomic steps are formed. This step bunching takes place very rapidly after the initiation of the growth, and the period of multiaatomic steps saturate to a certain value after several nanometres of growth. Figure 2.2 shows typical atomic force microscope (AFM) images of GaAs surfaces after MOVPE growth on vicinal (001) GaAs substrates. Here, the misorientation of the substrates is toward the [110], and the misorientation angle is varied from 2° to 6°. One can see the surfaces are periodically corrugated, as shown in the schematic cross section of figure 2.2. One also can see that the step edges are extremely straight over 1 μm. Furthermore, the average period L of multiaatomic step are almost the same for all substrate misorientation angles, and is 80 nm in these examples.

We here briefly summarize the main features of GaAs multiaatomic steps on GaAs vicinal (001) surfaces, some of which are already observed in figure 2.2.

- (1) The multiaatomic steps on vicinal (001) surfaces are spontaneously formed in a wide range of growth conditions during MOVPE growth of GaAs at a growth temperature of 600 °C [76].
- (2) They are also formed during thermal treatment in ultra-high vacuum [84] and AsH₃/H₂ ambient [85, 86].
- (3) If we look at the time-evolution of the multiaatomic steps from the initial surface with monoatomic steps, their period L increases, saturates to the same value very quickly and maintains the periodicity after the few nanometres of growth.
- (4) The saturated value of period L and height H of multiaatomic steps depend on growth conditions such as growth temperature, growth rate and AsH₃ partial pressures, and substrate misoriented angle. For example, the period L can be controlled in the range from about 30–100 nm [76], and tends to be shorter for higher arsine partial pressures.
- (5) When the misorientation angle is changed, the period of multiaatomic steps does not change so much, and the height changes in accordance with the misorientation angle (see [figure 2.2](#)).
- (6) Uniformity of the period of multiaatomic steps $\Delta L/L$ of $\sim 13\%$ is achieved with reasonable reproducibility, and the step edges are straight over 1 μm as we can see in figure 2.2.

The formation process of multiaatomic steps on GaAs (001) is schematically shown in [figure 2.3](#) and its main mechanism is explained by the Schwoebel effect [87–89]. This effect is caused by the difference of incorporation rate of atoms that diffuse on a certain terrace to either side of the steps, that is, to the step edges at down-steps (indicated by D in figure 2.3) or up-steps (U). Before the growth, the surface is composed of monoatomic steps and terraces, and their spacings have some fluctuation. The growth species are supplied from the gas phase, thus, the longer terraces have more chance to have species. Therefore, if the incorporation rate of grown species are larger at down-steps than that at the up-steps, the growth rate of the step front is larger for longer terraces, or smaller for shorter terraces. This results in the step bunching as shown in the figure. We note here that step-ordering can also take place if the incorporation rate becomes larger at the up-steps than at the down-steps, and as a result, an array of *monoatomic steps* are formed [90, 91].

To confirm this Schwoebel effect we carried out a Monte Carlo simulation of this step bunching process based on a microscopic modelling ([figure 2.4\(a\)](#)) of diffusion with appropriate diffusion barriers for hopping [89]. Here we introduced different barrier energy for the diffusion and incorporation processes between T (diffusion on terrace regions E_{terrace}), U (incorporation from up-steps, E_{up}), and D (incorporation from down-steps, E_{down}), being $E_{\text{up}} > E_{\text{down}}$. The result of the simulation is shown in figures 2.4(b) and (c), where we can see that the time evolution of the terrace width is clearly reproduced by our simulation with an appropriate set of parameters.

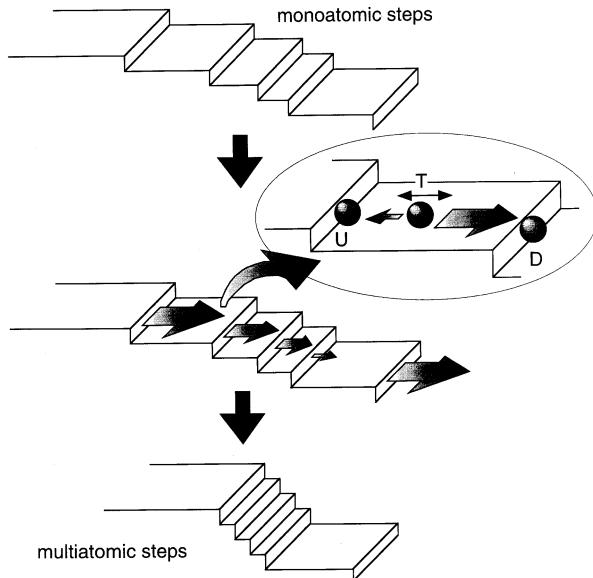


Figure 2.3. Schematic of the Step Bunching Mechanism on vicinal GaAs (001) surfaces based on the Schwoebel effect.

One may ask why E_{up} is larger than E_{down} on GaAs surfaces grown by MOVPE. This point is not clarified at present. Our recent investigation using ultra-high vacuum scanning tunnelling microscopy (UHV-STM) suggests that lack of As dimer rows at terrace regions just below the multiatomic step regions might be one origin of such differences [92]. In typical growth conditions where the multiatomic step are formed, the GaAs surface is reconstructed with $c(4 \times 4)$ pattern. Although the terrace regions of the vicinal surfaces is terminated nearly perfectly with $c(4 \times 4)$ reconstruction, (4×3) -like reconstruction is dominant at the terrace regions just below the multiatomic step edges. In addition, there are some disordering of As dimer rows on GaAs surfaces. On the other hand, AlAs surfaces had similar surface reconstruction both at terrace and multiatomic step regions, but disordering of the dimer rows at the terraces below the multiatomic step edges was not observed, and multiatomic steps started to debunch quickly as AlAs is grown on GaAs multiatomic steps.

The quantum wires of figure 2.1 can be formed on surfaces with multiatomic steps as long as the growth of the well layer preferentially takes place from the step edges, as it results in the thicker well width at the multiatomic step regions than at the terrace regions, i.e. lateral thickness modulation. This, in fact, can take place in GaAs/AlGaAs systems [11] and InGaAs/GaAs systems [80] as described below. As we can see in figure 2.2, the multiatomic steps are fairly periodic, straight and regular over $1 \mu\text{m} \times 1 \mu\text{m}$ regions. Thus, we can obtain high-

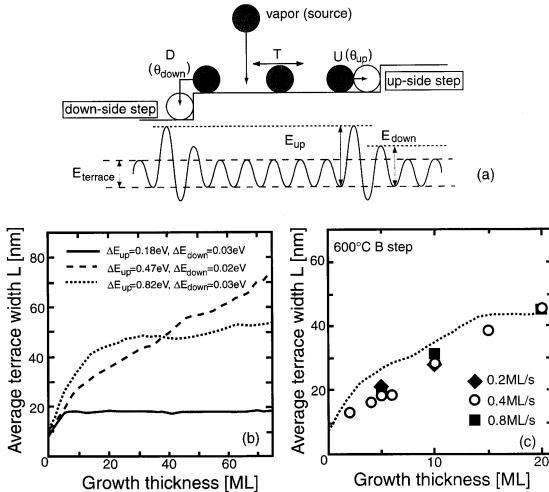


Figure 2.4. (a) The Monte Carlo model for step bunching. (b) Typical time evolution of step bunching obtained by Monte Carlo simulation. (c) Comparison of experimental results and simulation. For the experiment, the growth temperature was 600°C, and $E_{up} - E_{terrace} = 0.47$ eV and $E_{down} - E_{terrace} = 0.02$ eV were used in the simulation.

density uniform arrays of quantum wires. Furthermore, the average period L of multiaatomic step can be controlled by the growth conditions, but are insensitive to the substrate misorientation angles. Therefore, the size of QWRs can be controlled by controlling the substrate misorientation angle α as well as growth conditions and the amount of growth.

In the following, we describe the results on InGaAs QWRs and their optical properties and application to QWR lasers [65, 80]. Lasing operation using our QWRs in the active region have also been demonstrated [65, 93], which implies the anisotropic nature of dipole moment as a consequence of lateral confinement.

2.2.3 Fabrication of InGaAs quantum wire structures using multiaatomic steps

2.2.3.1 Growth modes of strained InGaAs on GaAs multiaatomic steps

We have reported on the formation of QWRs using GaAs/Al(Ga)As [11, 77] and InGaAs/GaAs [80] layers using multiaatomic steps. For the fabrication of QWRs, it is important to realize uniform and straight multiaatomic steps, and for GaAs/AlGaAs QWRs, the uniformity of barrier AlGaAs layers is thus considered to be one of the critical issues. However, when AlGaAs was grown on GaAs surfaces with multiaatomic steps, we found that the surface morphology, for example, the period of multiaatomic steps, was changed, and that uniformity of

the AlGaAs multiaatomic steps tends to be degraded as compared to the initial GaAs surface. We think this is due to the difference of sticking coefficient or migration length between Ga and Al atoms. If AlAs was grown instead of AlGaAs, the degradation of the step uniformity is not so critical as that of AlGaAs and improvement in the PL linewidth was demonstrated [77], although there seems to be some limitation in the thickness of AlAs layer. Therefore, it is advantageous to grow InGaAs layers directly on coherent GaAs multiaatomic steps. In addition, introduction of In into the well layers are also attractive for the application of lasers operating at infrared region at around 980 nm. On the other hand, there are difficulties in the growth of strained films. For these reasons, we investigated the growth mode of the strained InGaAs layer on GaAs multiaatomic steps to realize InGaAs/GaAs QWRs.

For MOVPE growth, a horizontal low-pressure MOVPE system was used with triethylgallium (TEGa), trimethylindium (TMIn) and AsH₃ as source materials. Vicinal (001) GaAs substrates misoriented by 2.0° and 5.0° towards the [110] direction were used. As a reference, a singular (001) substrate was also used. These three types of substrates were also used in the same growth run. Typical AsH₃ partial pressures were 4.2×10^{-4} atm for the growth of coherent GaAs multiaatomic steps and 6.7×10^{-5} atm for the formation of the InGaAs QWR structures. We grew 3 nm thick InGaAs on GaAs multiaatomic steps to investigate the growth mode of the strained layer and to form QWRs. The growth rate of GaAs for the formation of multiaatomic steps was fixed at 0.11 nm s⁻¹, and that of InGaAs for the formation of QWRs was about 0.050 nm s⁻¹. The content of InGaAs was between 0.1 and 0.3. Growth temperature was 600 °C.

Figures 2.5(a) and (b) show AFM images of GaAs layer surfaces prior to the growth of InGaAs layers on the 2.0°- and 5.0°-misoriented substrates, respectively. Here, the growth temperature was 600 °C. Coherent multiaatomic steps were observed over several micrometre areas especially on 5.0°-misoriented substrates. Multiaatomic steps on the 5.0°-misoriented surfaces were straighter than those on the 2.0°-misoriented ones. Moreover, the uniformity of the spacing of the multiaatomic steps on the 5.0°-misoriented surface was much better than that on the 2.0°-misoriented one. Average heights and spacing of the multiaatomic steps were 2.8 and 80 nm on the 2.0°-misoriented surface, and 6.1 and 70 nm on the 5.0°-misoriented surface, respectively. Figures 2.5(c) and (d) show AFM images of thin In_{0.1}Ga_{0.9}As layer surfaces on coherent GaAs multiaatomic steps. The thickness of the InGaAs layer was 3 nm on a singular (001) GaAs substrate. Spacings of multiaatomic steps were almost the same as those of underlying GaAs multiaatomic steps, and the surfaces of terrace regions were oriented to the exact (001) direction. Three-dimensional growth mode was not observed for InGaAs growth even on a singular (001) GaAs surface. These results suggest that In_{0.1}Ga_{0.9}As layers were grown in step flow growth mode rather than in Stranski-Krastanow growth mode, and the thickness modulation which is required to form QWRs can be realized.

Next, the variation of surface morphology of thin InGaAs layers with In

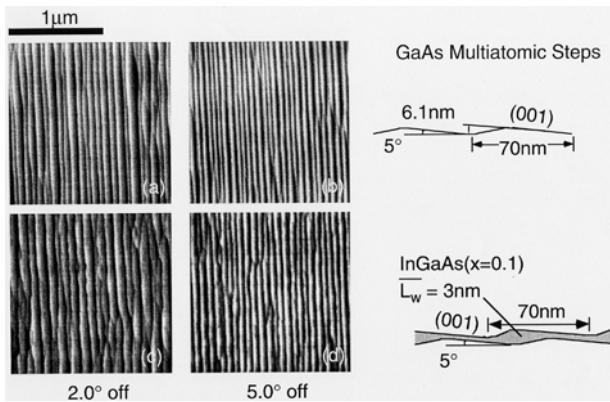


Figure 2.5. AFM images of GaAs and InGaAs grown on GaAs vicinal (001) surfaces at 600 °C.

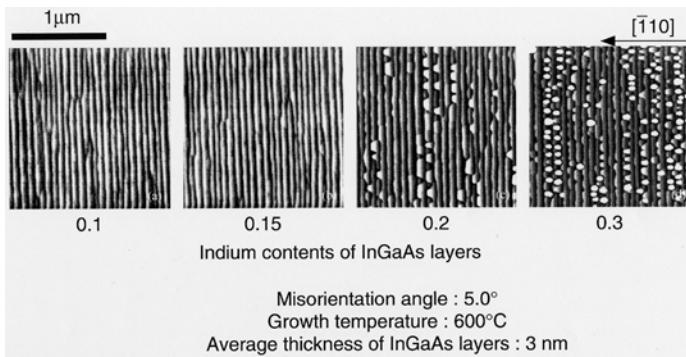


Figure 2.6. The indium content dependence of surface morphology of a 3 nm thick InGaAs layer grown on GaAs multiatomic steps.

content is shown in figure 2.6. The substrate misorientation angles were 5.0°. When the In content is smaller than 0.15, InGaAs layers were grown in step flow growth mode. For $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$ growth, prominent undulation of multiatomic step front was partly observed due to the increase of the strain effect. For $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ growth, many three-dimensional InGaAs island structures were formed at the edge of GaAs multiatomic steps in order to reduce the stress by the difference of lattice constant between InGaAs and GaAs.

In order to further confirm the growth mode of InGaAs at the edge of GaAs multiatomic steps, we measured the average inclination angles of multiatomic step areas to (001) terraces as a function of In content of InGaAs layers. Figure 2.7 shows the average inclination angles measured from line scan profiles of AFM images. The inset shows the definition of the average inclination angle, and

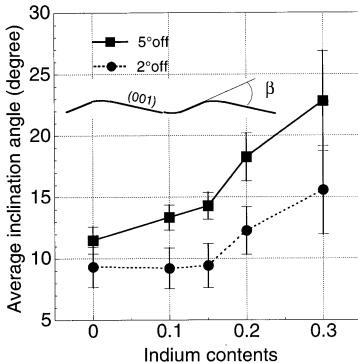


Figure 2.7. Indium content dependence of average inclination angle at the edge of InGaAs multiatomic step areas. See illustration for the definition of inclination angle.

error bars indicate standard deviations of inclination angles. These data exclude the three-dimensional island-like area for $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$ and $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$. As the In content of the InGaAs layer is increased, the average inclination angle of the multiatomic step areas to the (001) terrace increases for both 2.0° - and 5.0° -misoriented substrates. These results suggest that In atoms were more likely to stick to the multiatomic step regions than to the terrace regions, thus the preferential growth of InGaAs layer from step edges can take place. Therefore, InGaAs QWR structures are expected to be formed at the edge of GaAs multiatomic steps when the In content is smaller than 0.15.

For an In content larger than 0.2, three-dimensional island structures connected with wire-like structures are formed. These are likely to result from the Stranski-Krastanow growth mode. In addition, some kind of ordering of island structures can be seen along multiatomic step edges, particularly for $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ surfaces. These results suggest that these three-dimensional InGaAs islands on multiatomic steps can be used as position-controlled QD arrays [94]. We do not go into the detail of this QD formation as it is not the subject of our present work.

2.2.3.2 Formation of InGaAs QWRs

As we described in the previous section, the lateral growth of InGaAs takes place from multiatomic step edges. Thus, the InGaAs layer is expected to be thicker at the multiatomic step edges than on the terrace regions. This thickness modulation can be used to introduce lateral quantum confinement both for electrons and holes, and high-density arrays of InGaAs QWRs shown in figure 2.1 can be realized. We grew this structure by MOVPE on GaAs vicinal substrates misoriented by 2° or 5° toward $[\bar{1}10]$ direction. The thickness of the InGaAs layer is 3 nm on singular GaAs substrates, and In content was 0.1 or 0.15. We simultaneously grew InGaAs single quantum well structures as reference samples.

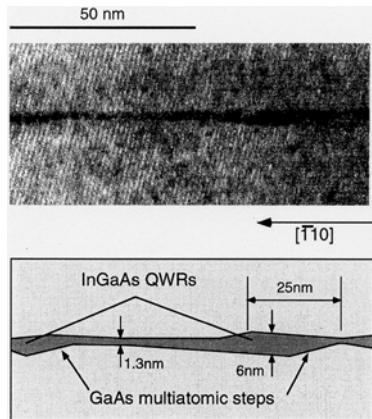


Figure 2.8. TEM image of InGaAs QWRs grown on a 5°-misoriented substrate. The In content of QWR is 0.1.

Figure 2.8 shows a cross-sectional transmission electron microscopy (TEM) image and a schematic illustration of the InGaAs QWR structure. Here the indium content of the InGaAs layer was about 0.1, and the sample was grown on a 5.0°-misoriented substrate. It was found that the thin InGaAs layer (dark area) was thicker at the edge of GaAs (bright areas) multiautomic steps than on the (001) terrace regions as expected, whereas InGaAs/GaAs QW grown on the singular (001) substrate had flat heterointerfaces. However, the amplitude of the corrugation resulting from the multiautomic steps surface seems to be decreased on the InGaAs top as compared to those of the underlying GaAs multiautomic steps. This is probably due to a slight segregation of indium atoms during the upper GaAs barrier layer growth at 600 °C. We have found that such ‘flattening’ can be eliminated by reducing the growth temperature of InGaAs and top GaAs layer down to 480 °C, but most of the results of optical characterization are obtained with somewhat flattened samples. From this TEM image, it was estimated that the lateral width and the vertical thickness of InGaAs QWR structures were about 25 and 6 nm, respectively, and that the QW thickness on the (001) terrace was about 1.3 nm.

2.2.4 Optical characterization of InGaAs quantum wires

2.2.4.1 Photoluminescence

We have carried out systematic characterization of our InGaAs quantum wires, and we think all of our results described below show one dimensionality of QWRs. We first measured the photoluminescence (PL) of $\text{In}_{0.1}\text{Ga}_{0.9}\text{As}$ QWRs on the vicinal substrates and the reference QW on the singular (001) substrate at 77 K, as shown in figure 2.9. The PL peak position of the reference QW was

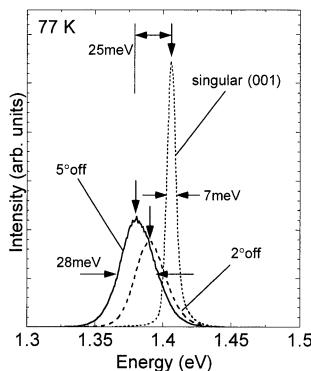


Figure 2.9. The PL spectra of InGaAs QWRs and the reference QW.

1.406 eV, which corresponds to the emission from the $\text{In}_{0.1}\text{Ga}_{0.9}\text{As}$ strained QW with a well thickness of about 3 nm. PL spectra of QWRs shift to a lower energy by 15 meV for the 2.0° -misoriented sample and 25 meV for the 5.0° -misoriented sample with respect to that of the QW. Full width at half maximum (FWHM) of the PL spectra were about 29, 28 and 7 meV for 2.0° -, 5.0° -misoriented and singular (001) samples, respectively. For the $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$ QWRs, we observed similar red-shift from a reference sample, being 23 and 31 meV for 2.0° - and 5.0° -misoriented samples, respectively.

In GaAs/AlGaAs QWRs formed at the edge of the multiatomic steps on the 5.0° -misoriented substrate, we reported a red-shift by 23 meV with respect to that of a reference QW, which agreed well with the numerical calculations for the QWR size obtained from the cross-sectional TEM images [11]. Furthermore, it was reported that the average indium content of a MOVPE-grown InGaAs layer on a vicinal GaAs surface was smaller than that on a singular (001) surface, as investigated by an energy dispersive x-ray (EDX) analysis, and PL spectra of InGaAs/GaAs QWs on the misoriented surfaces shifted to the higher energy side with respect to that on the singular (001) surface, that is, PL spectra showed blue-shift [95]. Therefore, lower energy shifts (red-shift) for the misoriented samples with respect to the singular (001) samples are mainly caused by the formation of thick InGaAs layers at the edge of GaAs multiatomic steps. We, however, note that our PL results are not direct evidence for lateral confinement, as we observed in a red-shift of PL and not the blue-shift expected for the quantum confinement. In addition, due to the lateral variations of the indium content and associated complicated strain effect, qualitative discussion in the PL spectra seems to be extremely difficult. Therefore, we must admit that the one-dimensional nature of QWRs is not demonstrated only with the present PL results.

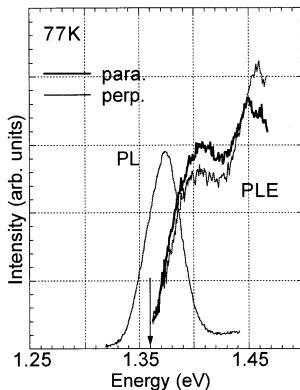


Figure 2.10. PLE spectra of InGaAs QWRs.

2.2.4.2 *Polarization anisotropy in photoluminescence excitation spectra*

We measured photoluminescence excitation (PLE) spectra of QWRs at 77 K and the results are shown in figure 2.10. In the PLE measurements, the detection energy was 1.359 eV, which corresponded to the lower energy side of the PL spectrum of the QWRs as indicated by an arrow, and the electric field of the excitation is either parallel (thick line) or perpendicular (thin line) to the QWRs. Two peaks can be seen in PLE spectra at 1.408 eV and 1.455 eV in either cases of polarization. The peak at 1.408 eV can be ascribed to heavy hole (HH)-electron excitonic transitions in the first subband of QWRs. On the other hand, the identification of the other peak is somewhat difficult because it is possible that it involves both the transition of light hole (LH)-electron excitons of the QWRs and that of the excitons in the excited states of the QWRs. The latter is more likely, because electroluminescence spectra at room temperature showed similar features in the energy difference between the ground and the excited states, as will be described in the next section.

We can see polarization anisotropy in the heavy hole exciton transition. Its peak intensity for perpendicular polarization was about 85% of that for parallel polarization. Although the quantitative argument is difficult because of the complicated valence band mixing and the strain distribution with complicated cross-sectional shape of the sample, this polarization anisotropy in PLE spectra is thought to reflect the one-dimensional nature of our QWRs. We also observed a large Stokes shift in the PLE spectra of about 30 meV at 77 K. Although a clear and detailed explanation for the origin of this Stokes shift cannot be given at present, it may arise from potential fluctuations along the wire axis, caused by size fluctuations and indium content variations of the InGaAs QWRs. Some fluctuations at the edge of InGaAs multatomic steps, which can be shown in some AFM images in the previous section, may also imply the existence of indium content variations along the wire axis.

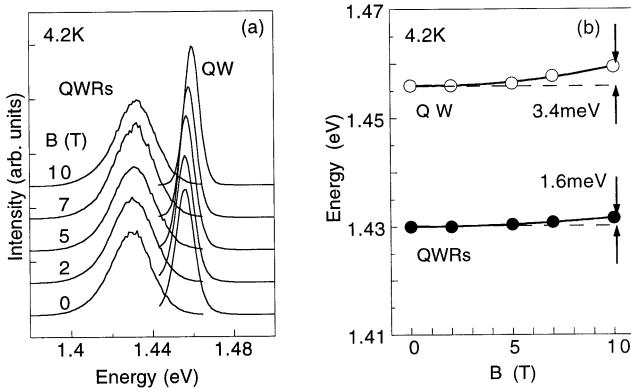


Figure 2.11. (a) The PL spectra of InGaAs QWRs grown on a 5° -misoriented substrate in magnetic field B . (b) The PL peak energy of InGaAs QWR, plotted as a function of magnetic field B . In each figure, the results of a reference QW are also shown.

2.2.4.3 Photoluminescence characterization in magnetic fields

Figure 2.11(a) shows PL spectra of InGaAs QWRs on the 5.0° -misoriented substrate and QW on the singular (001) substrate measured at 4.2 K in magnetic fields B up to 10 T. The direction of applied magnetic fields was parallel to the growth direction. The PL peak energies show clear blue-shifts for both samples as the magnetic field increases. The amount of the shift at 10 T is 3.4 meV for the QW sample, while it is 1.6 meV for the QWR sample, as shown in figure 2.11(b), where the peak energies are plotted as a function of magnitude of magnetic fields B . In addition, the peak shift, ΔE , is well fitted by the parabolic relation to B of diamagnetic shift of excitons, that is, $\Delta E = \beta B^2$, where β is the diamagnetic coefficient and is $16 \mu\text{eV T}^{-2}$ for QWRs and $45 \mu\text{eV T}^{-2}$ for QW, respectively.

As shown in the cross-sectional TEM image of figure 2.8, the InGaAs layer on multiatomic steps has in-plane thickness variation, and the thickest portion can be observed at the edge of GaAs multiatomic steps. Since its local thickness at the step edges is larger than the width of InGaAs QW on the singular substrate, the red-shift of PL peak energy of the QWRs relative to the QW is caused in the first order by such thickness difference as described above. It is known, however, that the diamagnetic coefficient, β , decreases as the well thickness of the QW becomes narrower [96]. In addition, the diamagnetic coefficient of unstrained InGaAs QW is calculated to be $19 \mu\text{eV T}^{-2}$ in the two-dimensional limit (we used reduced effective mass of excitons of $0.053m_0$ [97]). This value is somewhat larger than the experimental value for QWRs. Therefore, these magnetoluminescence results clearly demonstrate the existence of lateral quantum confinement and one-dimensional state in the strained InGaAs QWRs at the edge of multiatomic steps.

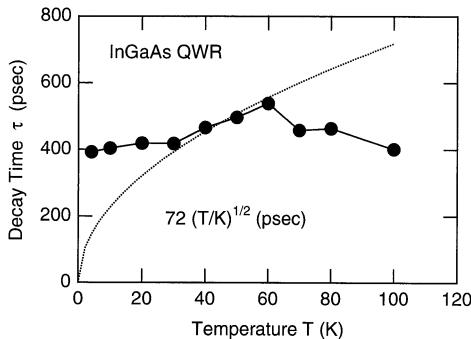


Figure 2.12. The temperature dependence of the radiative recombination lifetime τ .

2.2.4.4 Time-resolved photoluminescence

We also measured the decay time of PL from InGaAs QWRs by using the time-correlated single photon counting method. The measured PL decay time τ are plotted in the full circles as a function of temperature in figure 2.12. The dotted curve in the figure shows a result of fitting by $\tau \propto T^{1/2}$, which is predicted and reported for the radiative recombination lifetime of one-dimensional excitons in QWRs [98–101]. A reasonable agreement with the experiment and theory is obtained in the temperature range of $30 \text{ K} \leq T \leq 60 \text{ K}$, suggesting the one dimensionality of excitons. The deviation from the theory below 20 K is attributed to the localization, while that above 60 K is due to the contribution of the non-radiative component. The localization of carriers originates from the non-uniformity of QWRs. This is partly due to the fluctuation in the width and thickness of QWRs. In addition, In distribution is less uniform as compared to the singular surface, as we discussed from the Stokes shift in the previous section. The intrinsic radiative lifetime is estimated by following [100] and to be 160 psec assuming the centre-of-mass of excitons of InGaAs lattice matched to InP and $0.2m_0$ [102].

2.2.5 InGaAs quantum wire lasers

2.2.5.1 Device application of QWRs utilizing atomic steps

So far, there have been some attempts towards device application of quantum nanostructures formed using atomic steps. For electronic device application, QWR transistors or electron wave interference transistors [85] were proposed. In particular, as mentioned in the previous section, the period of multiaatomic steps is controlled in the range of 30–100 nm, which is comparable to the typical value of the Fermi wavelength of two-dimensional electron gas (2DEG). Therefore, by introducing the potential modulation having such periodicity, it is possible to realize the lateral surface superlattice (LSSL) type of electron wave interference

devices (EWIDs) [78, 79]. Coherent electron wave interference effects in LSSL structures were observed in the transconductance characteristics of the field effect transistor devices. The multiaatomic steps are also applied to realize single electron transistors (SETs), where the introduction of periodical modulation and squeezing the channel width below 1 μm by lithography and gate electric field is combined to form multiple quantum dots and multiple tunnel junctions [103].

On the other hand, applications of QWR structures using monoatomic steps on vicinal substrates to semiconductor lasers has also been reported [56, 104, 105]. In QWR lasers, the direction of the cavity is set to be parallel or perpendicular to the QWRs in a conventional Fabry–Perot cavity. In this case, the gain spectra are expected to be anisotropic between these two cavity directions [106]. When QWRs or QDs are formed by using *in situ* growth techniques, the quantum wires can be embedded in the distributed-feedback reflectors (DBRs). Therefore, QWRs and QDs can be compatible with QW structures in vertical cavity surface emitting lasers (VCSEL), and in fact, VCSELs using the triangular shaped InGaAs QWRs grown by selective area MOVPE [62] and self-assembled InGaAs QDs grown by Stranski–Krastanow growth mode [63] were attempted. In this section, we describe the first successful laser operation of InGaAs/GaAs QWRs on GaAs multiaatomic steps by current injection [65, 93].

2.2.5.2 Fabrication of InGaAs QWR laser diodes

We fabricated laser diodes with InGaAs QWRs in its active region as follows. We grew the separate confinement heterostructure (SCH) of figure 2.13 on a vicinal n-GaAs (001) substrate misoriented by 5.0° towards the $[\bar{1}10]$ direction, and a singular n-GaAs (001) substrate for a reference. Carrier concentration of the substrates was more than $1.0 \times 10^{18} \text{ cm}^{-3}$. Indium content of InGaAs layers was 0.15, and the growth temperature was 600°C . SiH₄ gas was used as n-type dopant for the growth of n-AlGaAs cladding layers, whereas p-type doping of p-AlGaAs cladding layers was achieved by an intentional incorporation of carbon (C) impurities under the growth condition of low V/III ratio (V/III = 29, using TMGa, TMA1 and AsH₃). Carrier concentrations of n- and p-type AlGaAs cladding layers were between $2.0 \times 10^{17} \text{ cm}^{-3}$ and $8.0 \times 10^{17} \text{ cm}^{-3}$, and aluminium content of AlGaAs layers was 0.32. After the growth of a thick AlGaAs layer, its surface morphology tends to be rough, that is, it is difficult to form coherent and straight AlGaAs multiaatomic steps. Therefore, we grew (GaAs)₂/(AlAs)₁ superlattice buffer layers to form the surfaces with monoatomic steps before the growth of non-doped GaAs layers which is to form uniform multiaatomic steps. It was confirmed by AFM observations that GaAs multiaatomic steps and InGaAs QWR structures can be formed after the thick Si-doped AlGaAs cladding layers, followed by the growth of Si-doped (GaAs)₂/(AlAs)₁ superlattice buffer layers. Furthermore, in order to eliminate the ‘flattening’ of InGaAs layers on multiaatomic steps described in the previous section, a 40 nm thick non-doped GaAs upper layer was grown at 480°C on the InGaAs QWR

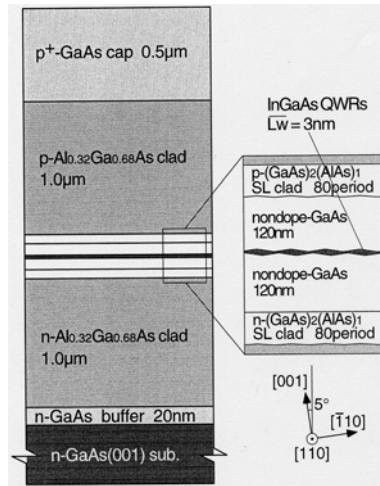


Figure 2.13. The sample structure of separate confinement heterostructure lasers.

layer, followed by a GaAs layer grown at 600 °C.

After the MOVPE growth of SCH structures, standard Fabry–Perot-type laser diodes (LDs) with as-cleaved 110 facet mirrors were fabricated. Cavity length was about 600 μm and its direction was in either the [110] or the [1̄10] direction, where the [110] direction is parallel to the QWRs and multiaatomic steps and the [1̄10] direction is parallel to the misorientation direction of the vicinal substrates, that is, perpendicular to the QWRs. The width of the electrodes was about 300 μm. Electroluminescence (EL) and lasing operation were investigated by current injection under the pulsed bias conditions at room temperature and 77 K. The pulsed bias conditions were as follows: repetition rate of 12 kHz and pulse width of 1.0 μs.

2.2.5.3 Characterization of laser diodes

Figure 2.14 shows EL spectra of a QWR-LD structure and its dependence on injection currents measured at room temperature and plotted in logarithmic scale. The cavity direction of the sample was parallel to the QWRs, that is, in the [110] direction. The EL spectra have a peak at 1.312 eV and are also asymmetric even at the smallest injection current investigated here. This asymmetry is also caused by the feature around 1.36 eV, and it becomes more prominent at higher injection currents. The asymmetry is probably due to the thermal distribution in the ground subband of QWRs and to the population in the excited states, because the energy difference between 1.312 eV and around 1.36 eV is comparable to that between two peaks in the PLE spectra shown in the previous section. The weak feature at 1.422 eV can be ascribed to the EL from GaAs. At 77 K, this SCH-LD showed

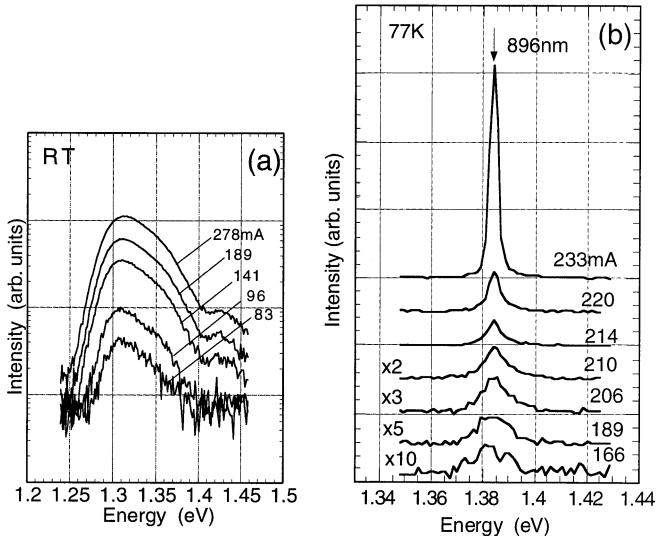


Figure 2.14. Electroluminescence and lasing spectra at (a) room temperature and (b) 77 K of QWRs on 5° -misoriented GaAs (001) substrates and its dependence on pulsed injection currents. The direction of cavity is parallel to QWRs, and the vertical axis of (a) is shown in the logarithmic scale.

clear lasing operation by pulsed current injection. Figure 2.14(b) shows EL and lasing spectra at 77 K. The spectra are fairly symmetric and become sharper and stronger as the current increases, and finally LD start lasing above 214 A cm^{-2} . The FWHM of the lasing spectra was less than 3 meV, and the wavelength at lasing operation was around 896 nm (1.384 eV), which was almost the same as the peak position of PL spectra at 77 K described in the previous section. Figure 2.15(a) shows current-light output ($I-L$) characteristics at 77 K under the pulsed bias conditions. It was roughly estimated the threshold current to be 212 A cm^{-2} , which corresponds to the threshold current density of 118 A cm^{-2} . These results are the first demonstration of a laser diode using self-organized InGaAs QWRs on GaAs multiatomic steps in its active layer.

Figure 2.15 also shows $I-L$ characteristics at 77 K with pulsed current injection of a QWR-LD (b) which have the cavity in the $[\bar{1}10]$ directions and a reference QW-LD (c). The threshold current I_{th} of QW-LD was estimated to be 150 mA, and that of QWR-LDs with cavity parallel and perpendicular to the QWRs array direction was 212 and 105 mA, respectively. These values give threshold current density of J_{th} of 83, 118 and 58 A cm^{-2} , respectively. The threshold current is the smallest for the QWR-LD with the cavity direction perpendicular to the QWR direction.

We ascribe this anisotropy in the threshold current to the anisotropy in the

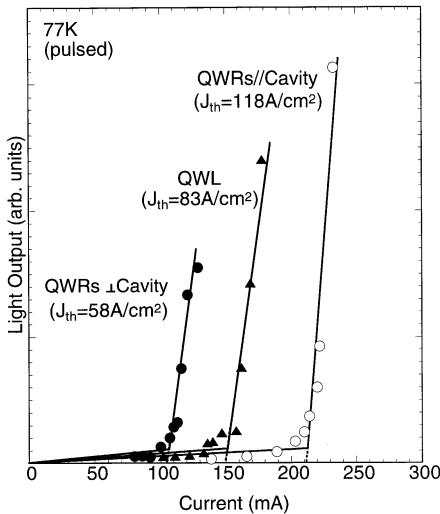


Figure 2.15. $I-L$ characteristics of QRS and QW lasers measured at 77 K. QWR lasers with a cavity parallel (a) and normal (b) to QWR, and (c) QW lasers.

gain in QWRs. In QWRs, lateral quantum confinement leads to the in-plane anisotropy of the electronic dipole moment, which leads to the dependence of gain on the direction of QWRs against the light propagation. Figure 2.16 shows a schematic of the relationship between the QWRs array and the cavity direction, that is, the direction of the light propagation for QWR-LDs investigated here. It is predicted that maximum gain can be obtained when the electric field vector \mathbf{E} of the light propagation becomes parallel to the direction of the QWRs array [106]. This indicates that gain is maximum when QWRs are perpendicular to the direction of light propagation, as shown in figure 2.16. On the other hand, if the light propagates parallel to the wires, the gain is expected to be smallest because the electric field vector is normal to the quantum wires. Therefore, the threshold current is the smallest when the cavity is normal to the quantum wires as we saw in the experiment. Such phenomenon has also been experimentally confirmed, for example, as described in [56–58]. Also in our present study, the threshold current density of the QWR-LD was lower when the cavity direction was perpendicular to the QWRs array direction, as shown in figure 2.15. Therefore, the result obtained here can be attributed to one-dimensional characteristics of InGaAs QWRs, and shows the first successful demonstration of self-organized InGaAs QWRs laser formed on GaAs multatomic steps.

The self-organized InGaAs QWRs used here might have the fluctuation of their size uniformity of about 13%, and their continuity is not perfect. We have not yet made any effort to reduce the threshold current density, and therefore, the threshold current density is not small enough to discuss the effect of one-

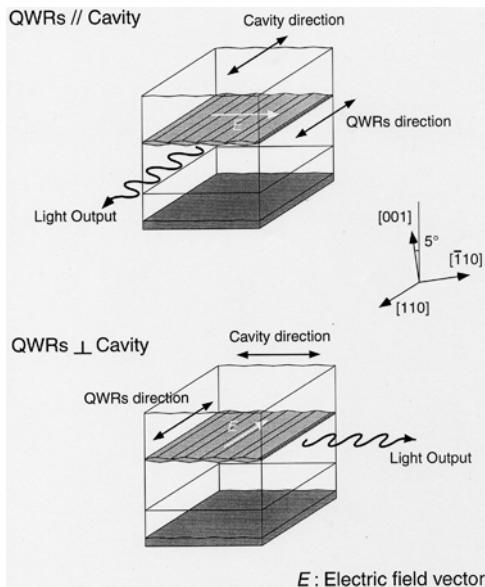


Figure 2.16. The schematics of the relationship between directions of QWRs and cavity. When the cavity direction is perpendicular to QWRs, the electric field vector E of the light is parallel to QWRs.

dimensional density of states in our QWRs. A lot of improvements can be made to reduce its threshold current density, for example, by introducing the current blocking layers and high reflection (HR) coated {110} facet mirrors. Further improvement of lasing characteristics is also expected by the optimization of epitaxial growth conditions, doping conditions of cladding layers, ohmic contact of p- or n-type electrodes, and so on.

2.3 GaAs and InAs quantum dots and related nanostructures by selective area growth

2.3.1 MOVPE selective area growth for quantum nanostructures

Selective area (SA) MOVPE on masked substrates is one of the promising methods for the fabrication of quantum nanostructures and related devices with the following advantages.

- (1) Formation of defect-free quantum nanostructures without process-induced damages and contamination.
- (2) Realization of extremely small nanometre-scale structures beyond the limit of conventional process techniques.

- (3) Realization of atomically smooth interfaces due to the appearance of crystalline facets.
- (4) Controllability of the position of quantum structures by patterning with standard lithography and etching techniques.
- (5) Versatility for the realization of various kinds of structures with the design of mask pattern of the substrates.
- (6) Realization of excellent size uniformity not affected by lithographic size fluctuation due to '*self-limited growth*'.

Application of SA-MOVPE for the formation of quantum nanostructures was first reported by Asai *et al* [14], where a very narrow channel of two-dimensional electron gas (2DEG) was formed near the top of the (100) surface on micrometre-scale mesa stripes, and also on {111} A facets. Then QWRs were successfully demonstrated for the first time at the edge of facet sidewalls of {111} B [15], and various kinds of QWR structures were reported using SA-MOVPE [16, 19–21]. Quantum dots have also been demonstrated using SA-MOVPE, such as tetrahedra QDs [22] grown on (111)B GaAs with triangular mask openings, or QDs near the top of pyramids with four-fold symmetry formed on (001) GaAs [23–25] and InP [26]. In addition, self-limited growth, in which the shape and size of the structures was maintained even if the growth proceeded, was found at the ridge of GaAs wire structures in SA-MOVPE using diethylgalliumchloride [107] and at the top of truncated tetrahedra on (111)B GaAs substrates [108]. Recent researches on the formation of QWRs and QDs by SA-MOVPE and their application of the various devices have been actively reported, for example, hexagonal-facet GaAs/AlGaAs lasers [17] with optical waveguides [18] and triangular facet laser with rectangular optical waveguides [67].

In this section, we describe the fabrication of quantum dots and single electron transistors by SA-MOVPE. We will show GaAs pyramidal structures with four {110} facet sidewalls are grown on SiNx masked (001) GaAs substrates with square openings. GaAs or InAs QD structures are then formed on the top of the pyramids after the growth of appropriate barrier layers [109, 110]. PL measurement of a single InAs QD showed that its linewidth was as narrow as 0.57 meV. These techniques are utilized to form a ‘quantum dot network’ in which each QD is connected with QWRs [111]. By further designing the pattern of the masked substrates, we have also succeeded in the fabrication of single-electron transistors (SETs) [69, 112], and resistance-load inverter by integrating a SET and a gated QWR structure [113].

2.3.2 Selective area MOVPE of GaAs and InAs quantum dots on pyramidal structures

2.3.2.1 Formation of GaAs pyramidal structures

First, MOVPE selective area growth was carried out on a masked substrate shown in figure 2.17(a). The masked substrate was prepared for semi-insulating (001)

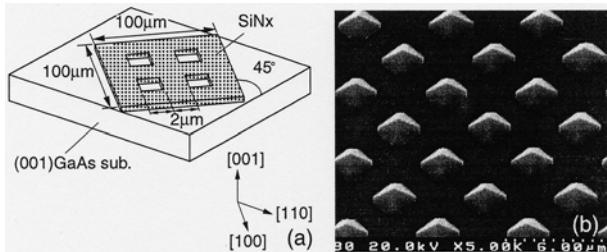


Figure 2.17. (a) A schematic illustration of the SiNx mask pattern for SA-MOVPE for $2\text{ }\mu\text{m} \times 2\text{ }\mu\text{m}$ size GaAs pyramids. (b) The SEM image of GaAs pyramidal structures after growth.

GaAs coated with a 40 nm thick SiNx layer, and photolithography and wet chemical etching were used to define the mask pattern. In the mask pattern, square openings of $2\text{ }\mu\text{m}$ by $2\text{ }\mu\text{m}$ were periodically aligned with a pitch of $4\text{ }\mu\text{m}$ in the $[100]$ and $[010]$ directions, and they were defined only in the inside of island-like patterns, whose size was $100\text{ }\mu\text{m}$ by $100\text{ }\mu\text{m}$ and also aligned periodically in the $[100]$ and $[010]$ directions with $200\text{ }\mu\text{m}$ pitch. At the outside of these islands, the SiNx mask was removed. It was shown that the use of such island-like pattern was essential to control the growth rate [114] and to achieve self-limited growth mode described later. Such island-like patterning are utilized for every SA-MOVPE growth, except for device structures described in section 2.3.4. The growth was carried out in a low-pressure, horizontal, RF-heated, quartz MOVPE reactor system with the working pressure of 76 Torr. Trimethylgallium (TMGa), trimethylaluminum (TMA1) (or triethylaluminum TEA1), trimethylindium (TMIn), and 20% AsH₃ diluted in H₂ were used as source materials. Typical partial pressures are 1.9×10^{-6} atm, 6.7×10^{-7} atm, for TMGa and TMIn, respectively.

Figure 2.17(b) shows the SEM image of pyramidal structures selectively grown on the masked substrates. We can see pyramidal structures having four {011} facets were formed, because the growth rate for {011} facets is slow compared to other low-index surfaces. Therefore, as growth proceeds, the top area of the truncated pyramids becomes narrower. GaAs/AlGaAs quantum dots structures can be overgrown on the top position of the pyramids.

2.3.2.2 Self-limited growth

We measured the size of the top portion of the pyramids by AFM and SEM. Figure 2.18 shows the time evolution of the size of the top portion of the pyramids. Growth temperature was 730 °C. The size was determined from the line profile of the AFM measurement (see inset of figure 2.23), and they were measured from either the central or edge part of the island-like patterning. In the initial

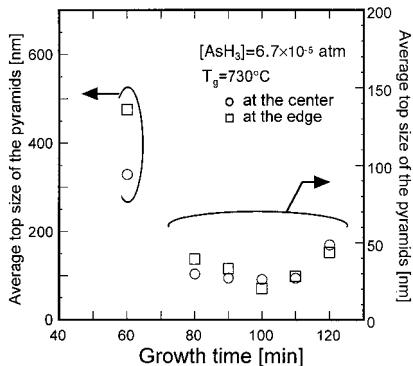


Figure 2.18. The time-evolution of the top size of GaAs pyramidal structures.

stage of growth, differences in the top size within the pattern were very large. However, as growth proceeded, the top size of all pyramids became almost the same, and saturated at about 50 nm under these growth conditions, leading to a highly uniform pyramid array. This result suggests that no growth occurred on the top or sidewalls of the pyramids after the pyramids were completely formed, thus the growth was carried out in *self-limited* mode.

The growth process of the pyramidal structures in self-limited growth mode is as follows: in the early stage of growth, part of the reactants diffuse laterally in the vapour phase outside the $100 \mu\text{m} \times 100 \mu\text{m}$ island pattern, because the outside of the island works as a large sink for the reactants. Then the amount of diffused reactants at the edge of the pattern are larger than at the centre. Therefore, non-uniform growth occurs within the island, as schematically shown in figure 2.19(a). However, as the growth proceeds, pyramidal formation is completed from the centre to the edge of the island pattern. Once the pyramidal formation is completed, no growth occurs on the top region of the pyramids or {011} facet sidewalls, because the growth rate is much slower than that on any other surface. In addition, the desorption rate of Ga atoms at the top of pyramids increases rapidly because of the weak bonding energy, so that growth saturates and most Ga atoms start to diffuse out of island-like patterns, as shown in figure 2.19(b).

Figure 2.20 shows the growth temperature dependence of the size of the top portion of the pyramids. The growth time was 80 min and AsH_3 partial pressure was 6.7×10^{-5} atm. Under these conditions, pyramidal formations were already completed, and uniform pyramids were obtained. The top size decreased as the growth temperature decreased. It was reported that there was a critical width on the ridge structures in selective area MOVPE using diethylgalliumchloride [107] and MBE [36], and also the top size of truncated tetrahedra on (111)B GaAs substrates [108], and the sizes were mainly determined by growth temperature. In the present case, we can also conclude that the shape and top size of the

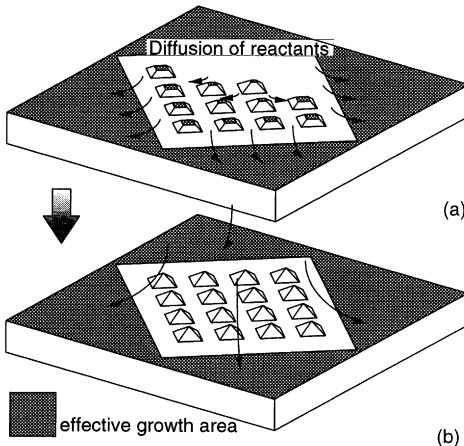


Figure 2.19. Schematic illustrations of the lateral diffusion of the reactants in vapour phase.

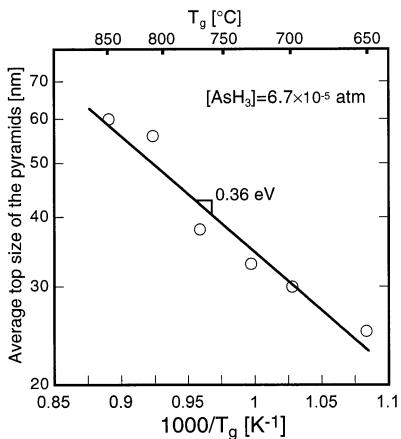


Figure 2.20. The growth temperature dependence of the average size of the top portion of the GaAs pyramids.

pyramids are determined by the temperature and by the thermal equilibrium between adsorption and desorption processes of Ga atoms near the top portion of the pyramid. From this temperature dependence, the activation energy was estimated to be 0.36 eV. This value is thought to correspond to the difference $E_d - E_a$ between desorption energy from the steps B_d and adsorption energy to the steps E_a .

The top size was also found to depend on AsH₃ partial pressure, and was

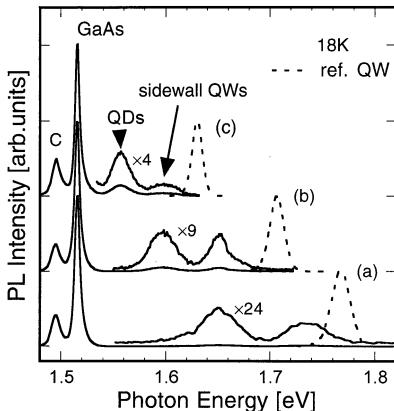


Figure 2.21. The PL spectra of GaAs QDs with various vertical well thickness.

smaller for larger partial pressure. By controlling the growth temperature and AsH₃ partial pressure, the top size can be reduced to about 25 nm in self-limited growth mode. Therefore, quantum dots with lateral dimension of around 20 nm can be formed on the pyramids with optimized growth conditions.

2.3.2.3 Formation and photoluminescence study of GaAs quantum dots

The self-limited growth is also expected to be effective to regularize the top width of the pyramidal structures, and thus to fabricate uniform arrays of GaAs QDs on the pyramids. To form QDs, GaAs was selectively grown by self-limited growth at the high temperature of 730 °C and low AsH₃ pressure of 6.7×10^{-5} atm. Under these growth conditions, the top of the pyramid was saturated at about 40 nm. Next, AlGaAs/GaAs single quantum wells were grown on these pyramids at low temperature of 650 °C and high AsH₃ pressure of 5.0×10^{-4} atm. AlGaAs and GaAs layers were grown both at the top of pyramids and on {011} facet sidewalls under the overgrowth condition. In this case, the growth rate of the pyramidal top area measured by SEM was about 1.8 times higher compared to that of {011} sidewalls, that is, the quantum dot structures connected to the sidewall quantum wells were formed on the top of GaAs pyramids. The schematic view of a cleaved cross section of the quantum dot structure is also shown in the inset of figure 2.26. The lateral width of QDs is estimated to be 30 nm, and the effective lateral confinement dimension is the same as or somewhat larger than 30 nm since the QDs are connected with QWs on {110} facet sidewalls.

Figure 2.21 shows the PL spectra of QDs with various thickness of the GaAs layer measured at 18 K. The thickness of the GaAs well layer on reference QW on planar substrates are (a) 2 nm, (b) 3 nm, and (c) 4.5 nm, respectively. Each spectra have two peaks at the higher energy side of the PL peak from the GaAs buffer layer (1.51 eV). By comparing with the results of SEM, the lower and

higher side of the peak correspond to the emission from the top QDs and sidewall QWs, respectively, and the peak positions agree well with the estimated well width of QWs and QDs. For example, in spectra (c), the peak at 1.633 eV for the reference QW (broken curve) agrees with the calculated value for QW with 4.5 nm well width. The peaks of 1.546 eV and 1.596 eV correspond to the emissions from the top quantum dots and sidewall quantum wells, and the calculated well widths from PL peak energies are 12.0 nm and 6.0 nm, respectively. From the SEM observation, the layer thicknesses were 10.4 nm for the top quantum dot and 5.8 nm for the sidewall quantum wells, and these values agree with those estimated from PL peaks. The width of PL from QD is 22 meV. This relatively small value suggests that uniform quantum dot arrays were formed using self-limited growth mode, but size fluctuation is not still negligible in the present QDs.

2.3.3 Fabrication and characterization of high-density quantum dot arrays and quantum dot network

2.3.3.1 High-density array of QDs

In order to apply QDs to quantum devices, it is important to fabricate precisely controlled high-density QDs with high-uniformity. For this purpose, we prepared masked substrates with circular openings whose diameters were about 180 nm and were periodically aligned in 230 nm pitch by electron beam lithography and wet etching. [Figure 2.22\(a\)](#) shows SEM images of GaAs pyramidal structures. The structure consists of four {001} facets similar to the GaAs growth shown in [figure 2.17\(b\)](#). Similar pyramidal structures were also formed for growth of AlGaAs or GaAs/AlGaAs multilayers. GaAs were grown on the above AlGaAs pyramids to form QDs. [Figure 2.22\(b\)](#) shows an SEM image of cleaved cross section of GaAs/AlGaAs multilayer structures grown on the masked substrates. The thicknesses of GaAs and AlGaAs were 4 nm and 12 nm, respectively, on planar substrates. It was found that GaAs QD on the top portion of the AlGaAs pyramid and sidewall QWs on {011} facets can be formed. Therefore, GaAs QDs can be formed at the top portion of AlGaAs pyramids and embedded by AlGaAs. [Figure 2.23](#) shows the time evolution of the top size of AlGaAs pyramidal structures. It is noted that the AlGaAs pyramids were completed at the growth time of about 3 min. Due to the self-limited growth mode, the average top size of pyramids more or less stays constant over a growth time of 3 min. It was also found that the average top width of AlGaAs pyramids was less than 20 nm, with standard deviation of 2.55 nm at the growth time of 3 min, indicating that the uniformity is better than $\pm 20\%$.

The growth sequence for formation of QD was as follows. First a GaAs buffer layer was grown. Next, AlGaAs was grown at the growth temperature of 720 °C and under lower AsH₃ partial pressure of 6.7×10^{-5} atm for the formation of uniform AlGaAs pyramids. Next at the lower growth temperature of 650 °C and under higher AsH₃ (pressure of 5.0×10^{-4} atm) a GaAs well was overgrown

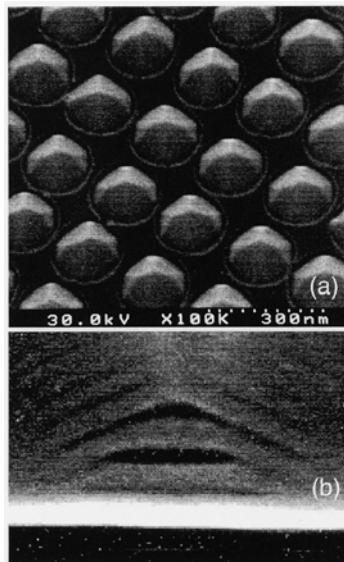


Figure 2.22. (a) SEM images of high-density GaAs pyramidal structures and (b) the cleaved cross-sectional SEM image of GaAs/AlGaAs multilayer structures.

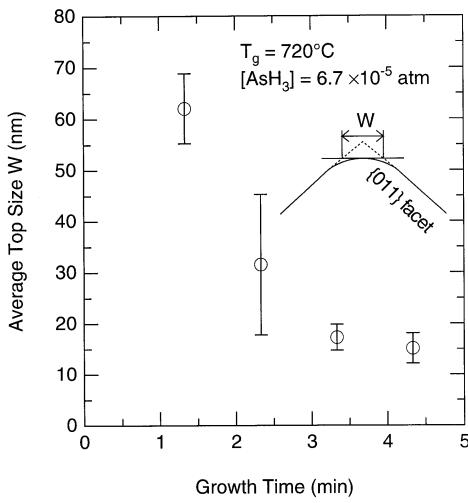


Figure 2.23. Time evolution of average size at the top portion of the AlGaAs pyramids. The inset shows the definition of the top size of pyramids.

on the AlGaAs pyramids, where the growth occurred on the {011} facet sidewall under these growth conditions [16]. Finally, an upper barrier layer of AlGaAs

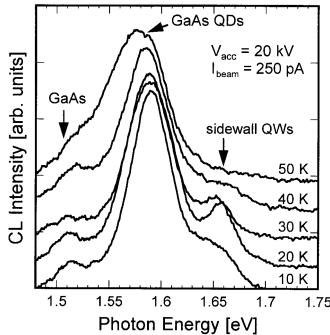


Figure 2.24. The CL spectra of QDs at various measurement temperatures.

and a GaAs capping layer were overgrown on the whole surface. Here, the GaAs well thickness was about 3 nm for planar substrate. The lateral size of QD was estimated to be about 20 nm from AFM measurements. To confirm the realization of GaAs QDs and estimate their size, cathodoluminescence (CL) measurement was done and the results are shown in figure 2.24. Emission from QDs are clearly observed at 1.59 eV at 10 K. The vertical size of the QDs was estimated to be 7 nm from the calculation taking account of the lateral confinement of 10 meV for the thickness of the single QW on the planar substrate. In the same way, the thickness of sidewall QWs was estimated to be about 4 nm from the peak at 1.65 eV and 10 K. These size estimations are consistent with SEM observation. The emission from the sidewall QW disappeared on increasing temperature to 50 K. This is because the carrier capture efficiency of the QDs was higher than that of sidewall QWs, suggesting that high quality GaAs QDs were formed on AlGaAs pyramids. Somewhat broad peaks of the emission from the QDs were mainly caused by size distribution of the QDs with the $100 \mu\text{m} \times 100 \mu\text{m}$ island-like pattern. In addition, the peak broadening might be also due to large carrier density in CL measurement, as the excitation power strongly affects the full width at half maximum (FWHM) of the emission peak from QDs with sharp and discrete density of states. Nagamune *et al* [115] reported on the excitation power dependence of the FWHM of the photoluminescence peak of the dot and that of 12.5 nm thick QW measured at 15 K. In their experiment, the FWHM of the dot decreased with decreasing laser power, while that of the QW was almost constant. In addition, the FWHM of the dot was about 3 times broader than that of QW at the excitation laser power of $10 \mu\text{W}$, however, it began to decrease further and was close to the experimental resolution below the laser power of $0.1 \mu\text{W}$, and finally reached the experimental resolution of 0.9 meV at the laser power of 15 nW. Therefore, in our data, broader FWHM (51 meV) of the GaAs QDs was caused by the large carrier density excited by electron beam as well as the size distribution.

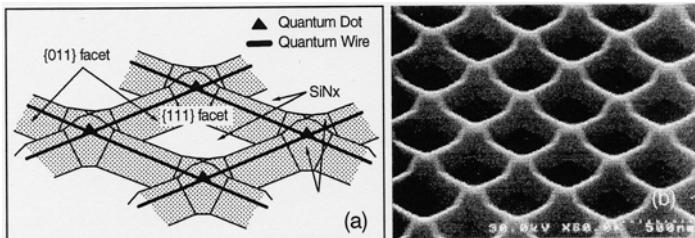


Figure 2.25. (a) A schematic illustration of ideal quantum wire-dot coupled structures and (b) the SEM image of AlGaAs structure consisting of pyramidal dots connected to a wire grown on masked substrate.

2.3.3.2 Quantum dot network: a quantum wire-dot coupled structure

The QD arrays we have shown so far are, in principle, isolated with each other, and they can be applied to optical devices, such as QD lasers. The application of QDs to electron devices, especially single electron devices, is also a very important research field for future electron devices. For fabrication of such electron devices, well-arranged QDs have to be connected to each other. For example, in single electron transistors (SETs) QWRs and QDs are connected through tunnelling barriers. Lateral surface superlattice (LSSL)-type of electron wave interference devices (EWIDs) is also possible if potential energy is periodically modulated by the coupling of QDs and QWRs. In fact, it was attempted to fabricate quantum dot-tunnel barrier systems by MBE on patterned substrates [116] to realize SETs. In SA-MOVPE, such coupled or connected structures can be realized simply by modifying and designing the pattern of the masked substrate. As such an example, we proposed ‘quantum dot network’ in which QDs and QWRs are coupled and connected to each other and form a network as schematically shown in figure 2.25(a).

To realize such quantum wire-dot coupled structures, we carried out selective area growth on masked substrates in which $200\text{ nm} \times 200\text{ nm}$ square SiNx masks were periodically formed both in the [100] and [010] directions with 400 nm pitch. The SEM image of figure 2.25(b) shows results of AlGaAs selective area growth. Pyramidal structures with high uniformity are formed, and they are connected to each other with wire structures. The pyramidal structures consist of {114} A and B facets near the top region, and {111} A and B facets at the bottom, and the wires connecting QDs consist of {011} facets.

Figures 2.26(a) and (b) show cleaved cross-sectional SEM images of the AlGaAs/GaAs multilayered structure at the pyramid and wire part, respectively (see figure 2.26(c)), together with their schematic illustrations. The thicknesses of GaAs and AlGaAs layers were 8.4 nm and 25 nm on the planar substrate, respectively. For the pyramid parts, {111} A and {111} B facet sidewalls preferentially appear at the bottom of the pyramids for both GaAs (dark regions)

and AlGaAs (bright regions), while {113} and {114} facets appear at the top portion of the pyramids for GaAs and AlGaAs, respectively. For the wire parts connected to the pyramids, {011} facets appear as the sidewalls of the trapezoids for both GaAs and AlGaAs growth. After the formation of the pyramid and triangular shaped wires are completed, the shape of the top portion of the pyramids and wires remains unchanged, although the growth proceeds on each sidewall. Therefore, uniform GaAs QDs surrounded by {113} and {114} facets can be formed at the top portion of AlGaAs pyramids. GaAs ridge QWRs connected to GaAs QDs can also be formed at the ridge of AlGaAs triangular wire structures. This result indicates that coupled arrays of QD and QWR structures can be realized using this technique. From cleaved cross-sectional SEM observation, the thickness of GaAs on the pyramids is 1.8–2.1 times thicker than that on {111} planes. On the other hand, the growth thickness of GaAs on {011} sidewalls of QWRs is much thinner under this growth condition. The thickness of the QWRs would be enhanced under lower growth temperature and/or higher AsH₃ partial pressure growth conditions because of higher growth rate on {011} plane with these conditions [16].

In order to fabricate coupled QD (CQD) structures, a thin GaAs layer was grown first on a SiNx-masked substrate as a buffer layer. Next, AlGaAs was grown to form the pyramid and wire structures. Then, AlGaAs/GaAs QWs were grown on the AlGaAs pyramid and wire structures. The thickness of the GaAs QWs was 2.8 nm for the planar substrate. Finally, AlGaAs upper barrier layers and GaAs cap layers were grown under higher AsH₃ partial pressure condition of 5.0×10^{-4} atm to bury the QDs and QWRs. [Figure 2.27\(a\)](#) shows cathodoluminescence (CL) spectra of CQDs at 4 K. In the measurement, about 25 CQDs at $2 \mu\text{m} \times 2 \mu\text{m}$ area were excited by a 20 kV electron beam with a 300 pA current. A peak and a shoulder are seen at 1.61 eV and 1.67 eV, respectively. To identify the origins of these CL peak and shoulder, spatially resolved CL images were observed, and a result at photon energy of 1.612 eV is shown in figure 2.27(b). Clear bright spots with the periodicity of 400 nm were observed. Therefore, this result suggests that the origin of the emission at 1.61 eV comes from QDs on the top portion of the pyramids surrounded by {113} and {114} facets. The lateral size of QDs are estimated to be 100 nm. Thus, the lateral confinement is not very large with the present sample, and their effective thickness is estimated to be 5 nm. In contrast, emission at 1.67 eV was observed from all parts of the structures. This peak energy corresponds to an emission from QWs with 3.5 nm well width, and thought to be originated either from QWs on {111} facets or on {110} facets.

The direction of ridge QWRs connecting QDs are defined by the initial patterning of the mask and in the <100> direction in the above samples. For this case, clear emission from QWRs was not observed as shown in the CL spectrum of figure 2.27(a). We also attempted to fabricate QD networks with QWR directions of <110>. For the case of QD networks with <110>-oriented wires, the emission from both QDs and ridge QWRs were resolved from spatially resolved

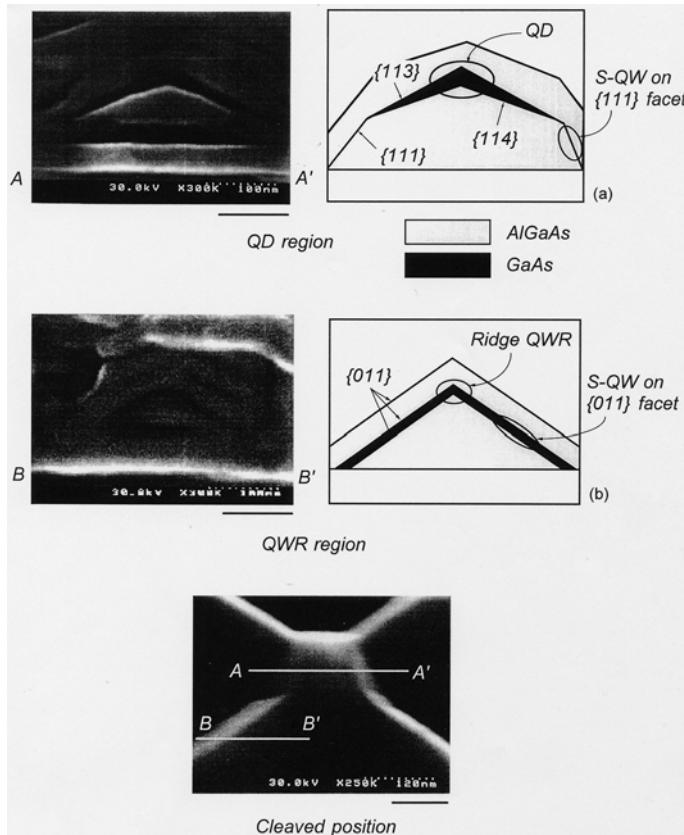


Figure 2.26. Cleaved cross-sectional SEM images and their schematic illustration of CQDs.

CL images. Therefore, quantum wire-dot coupled structures are expected to be formed more easily with $\langle 110 \rangle$ -oriented QWRs than with $\langle 100 \rangle$ -oriented QWRs.

2.3.4 Position controlled InAs quantum dots on GaAs pyramidal structures

2.3.4.1 Growth of InAs QDs on GaAs pyramidal structures

A number of studies are reported on the In(Ga)As grown on planar GaAs substrates by using the self-assembled islanding method in the Stranski-Krastanow growth mode [41–51]. High density of In(Ga)As QDs with excellent optical quality are demonstrated in those self-assembled QDs, and QD lasers were reported [59, 60]. However, it is not easy to control their size uniformity and, especially, dot positions on the substrates, which is thought to be a critical problem for the application to the electron devices.

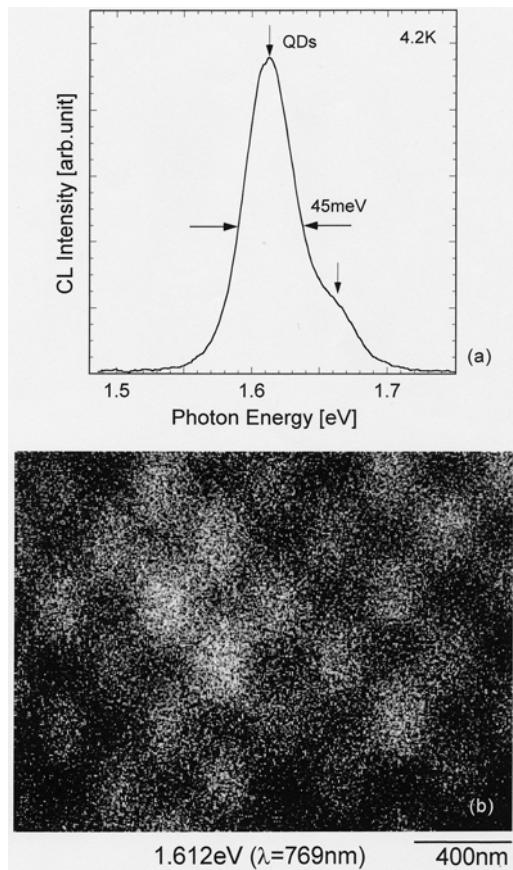


Figure 2.27. (a) The CL spectrum of GaAs CQDs and (b) the spatially resolved CL image at the photon energy of 1.612 eV.

In the following, we demonstrate a novel formation method of position controlled GaAs and InAs QD structures by the combination of SA-MOVPE and self-assembling growth of InAs on GaAs. As compared to the conventional self-assembled InAs quantum dots, our approach has the advantage that the position of the quantum dots are controlled by designing the pattern of the masked substrate. In addition, the number of quantum dots formed are determined by the patterning for GaAs pyramidal structures, as described below.

The selective area growth was carried out on GaAs (001) substrates having ~ 150 nm diameter circle windows on a SiNx mask. The period of the pattern has some variations, ranging from 160–500 nm, which corresponds to the density of the open circular windows to be $4 \times 10^9 \text{ cm}^{-2}$ and $4.4 \times 10^8 \text{ cm}^{-2}$, respectively. These circle patterns were defined only at $100 \mu\text{m}^2$ square area, and SiNx

mask was removed outside this square region as similar to [figure 2.17\(a\)](#). After GaAs micropyramidal structures having four-fold symmetry {110} facets were grown at 700 °C, InAs was grown at 500 °C with average thicknesses of InAs at 1.6×2.5 ML. Growth rate of GaAs is 0.13 nm s^{-1} in the present experiment. The partial pressures of AsH₃ were 5.0×10^{-4} atm for underlying GaAs pyramidal growth, and 6.7×10^{-5} atm for the formation of InAs dots, respectively.

First, we investigated the growth mode of high-density GaAs pyramidal structures formed by SA-MOVPE. [Figure 2.28](#) shows plan-view SEM images of GaAs pyramids which are grown for different GaAs growth times. The schematic illustrations of GaAs pyramids are shown on their right-hand side. The growth times were 1'40'', 2'40'' and 4'40'' for the samples in figures 2.28(a), (b) and (c), respectively. We can see uniform GaAs pyramids having four-fold symmetric {001} facet sidewalls. However, the shape near the top of the pyramidal structure looks different for different growth time. At the initial stage of the growth, GaAs pyramids were mainly consisting of {124} facet sidewalls at the top portion of the pyramids, as shown in figure 2.28(a). As the growth proceeded, the area of bottom {011} facets increased toward {124} facet sidewalls at the top portion of the pyramids, as shown in figure 2.28(b). Finally, the pyramids were completely formed with {011} facets whose growth rate was slowest of all the low-index surfaces (figure 2.28(c)). At the bottom corners of {110} facet sidewalls, {111} B and {111} A facets were also observed.

We then grew InAs on these pyramidal structures which have different shape near the top. SEM images are the growth are summarized in [figure 2.29](#) together with the schematic illustration of InAs quantum dots. The growth amount of InAs was to 2.0 monolayers (MLs) for planar substrates. Sample (a) has twin InAs dots on the top of GaAs pyramids in the direction of [̄10], and sample (b) has single InAs dots. However, for sample (c), no InAs dots were observed.

The difference in the formation of InAs quantum dots is explained by the difference of the facets near the top of the pyramids. In the case of sample (a), the top portion of GaAs pyramids consist of wide {124} surface from SiNx masked area and also the bottom {111} facet area, and start to nucleate near the top. As a result, twin InAs dots are formed. If the area of the {124} facets becomes smaller, the number of steps and kink decreases, and InAs dots are formed only at the top of the GaAs pyramids, as shown in sample (b). In the case of sample (c), the area of {124} facets is further reduced by increasing the amount of GaAs growth, and GaAs pyramids are completed with {011} facets. In this case, formation of QDs at the top becomes difficult, because the effective area that contributes to the growth of InAs becomes too small. These results suggest that we can control the position and number of InAs dots on GaAs pyramids by changing the growth time of the underlying GaAs pyramids.

Next, we observed the shape and the density of InAs dots on the top of GaAs pyramids and on the reference planar substrate for various InAs growth amounts. The growth time of underlying GaAs pyramids were fixed at 2'40'' ([figure 2.29\(b\)](#)). The diameter and the periodicity of the circle patterns were

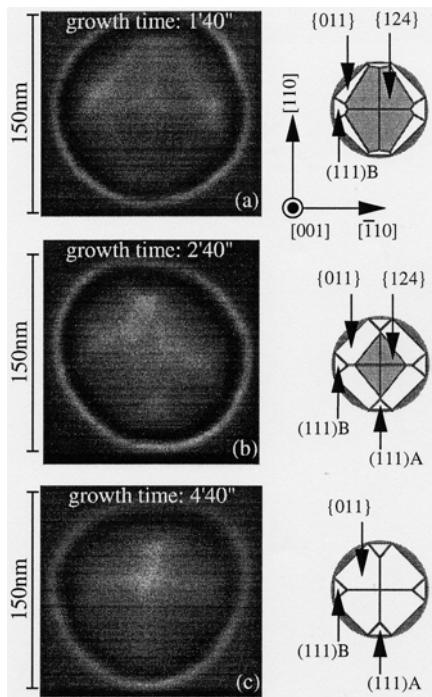


Figure 2.28. A plan view of SEM images and schematic illustrations of underlying GaAs pyramids for different growth times.

150 nm and 300 nm, respectively. Figures 2.30(a)–(c) show SEM images of InAs dots on GaAs pyramids, and figures 2.30(d)–(f) show AFM images of InAs dots formed by S–K growth mode on planar substrate. The average growth thicknesses were 1.6 ML for samples (a) and (d), 2.0 ML for sample (b) and (e) and 2.5 ML for samples (c) and (f), respectively. As increasing the growth thickness, the size of InAs dots on GaAs pyramids became slightly bigger. However, we observe very little change in the InAs dot size on pyramids. On the other hand, the density of InAs dots on the planar substrate increased by about two orders of magnitude by increasing the growth time. For 1.6 ML InAs growth, small InAs dots were formed on the top portion of GaAs pyramids (figure 2.30(a)), while they were not formed yet on the planar substrate (figure 2.30(d)). Nucleation and growth are much easier on GaAs pyramids, because many step and kink sites were on the top portion. On the other hand, in the case of sample (c), many large InAs dots were formed on the planar substrate and the density of small dot decreased as shown in figure 2.30(c), but we did not observe such big dots on the pyramids (figure 2.30(f)), probably because excess In atoms diffused outside the patterned area of 100 μm square. Almost the same results were obtained for the

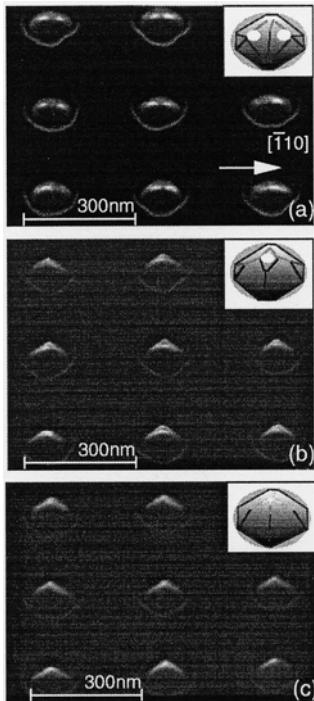


Figure 2.29. SEM images and schematic illustrations of InAs dots on GaAs pyramids, for different growth times of underlying GaAs pyramids and fixed growth thickness of InAs (2.0 ML).

density of the circle patterns on the substrate from $4.4 \times 10^8 \text{ cm}^2$ – $4 \times 10^9 \text{ cm}^2$. Therefore, these results indicate that we can control the number and positions of InAs dots by changing the position, numbers, and the growth time of underlying GaAs pyramids.

2.3.4.2 Optical characterization of InAs quantum dots on GaAs pyramids

Figure 2.31(a) shows the schematic illustration and layer structure of the position controlled InAs quantum dots used for optical characterization. The growth time for underlying GaAs pyramids were 2'40'' and the growth thickness of InAs was 2.0 ML. Figure 2.31(b) shows the PL spectra of InAs dots on GaAs pyramids with the density of $4.4 \times 10^9 \text{ cm}^{-2}$ on the reference planar substrate measured at 4 K. We observed an emission peak at 1.34 eV from InAs quantum dots formed on GaAs pyramids. For InAs quantum dots on a planar substrate, an emission peak from InAs quantum well of wetting layer at 1.44 eV and a broad peak from the InAs quantum dots on GaAs pyramids at 1.31 eV were observed,

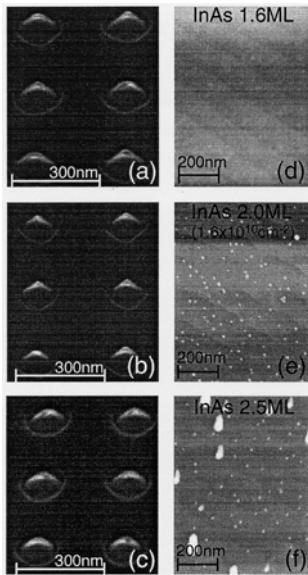


Figure 2.30. The relation between the size and density of InAs dots on pyramids and the growth thickness of InAs. (a)–(c) show SEM images of InAs dots on GaAs pyramids. (d)–(f) show AFM images of InAs dots on the planar substrate.

implying that these quantum dots were formed in S–K mode. The linewidth of the emission peak from InAs dots on pyramids was little broader than that on the planar substrate, probably because of the size non-uniformity of the underlying GaAs pyramids. In addition, the PL peak energy of InAs dots on pyramids shifts to higher energy by about 40 meV, as compared with InAs dots on the planar substrate.

In order to further confirm the zero-dimensional properties of our InAs QDs, we carried out micro-PL measurement of QDs on GaAs pyramids, and limit the number of quantum dots to be measured by focusing the excitation laser spot to a few micrometre ($3\text{--}5 \mu\text{m}$) regions. The measurement was carried out with a $\times 20$ microscope objective, and the PL was excited and collected through the same microscope objective. The collected signal was dispersed into a spectrometer with a CCD multichannel detection system. With micro-PL measurement, the main broad peak of the InAs quantum dots on GaAs pyramids was found to be composed of a number of sharp spike-like emission lines as shown in figure 2.32(a). When the excitation intensity was further reduced, an extremely sharp single emission line was observed, as shown in figure 2.32(b). If we estimate the width of this emission line by the Gaussian lineshape, the full width at the half maximum is measured to be 0.57 meV. We think this linewidth is limited by the spectral resolution of our measurement system, and expect much

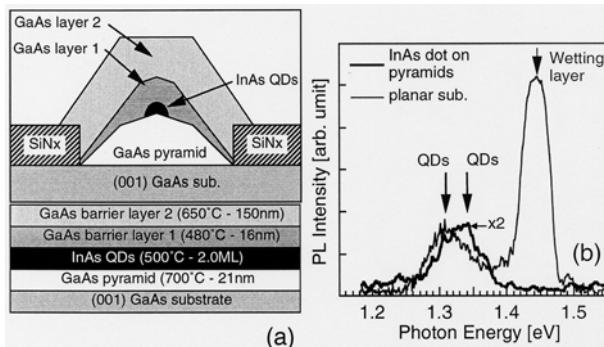


Figure 2.31. (a) A schematic illustration and layer structure of InAs quantum dots. (b) The PL emission spectra of InAs quantum dots on GaAs pyramids and S-K mode InAs quantum dots on planar substrate measured at 4 K.

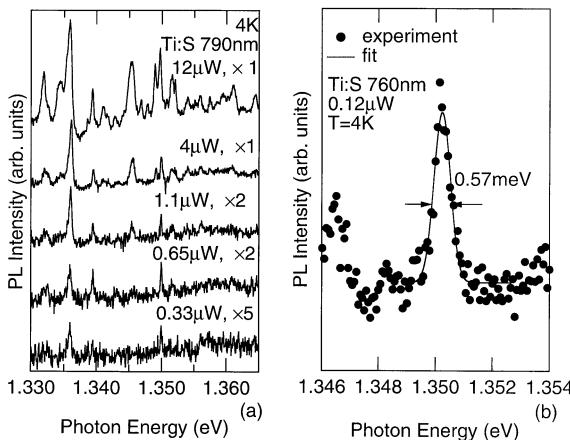


Figure 2.32. (a) The micro-PL spectra of InAs QDs on GaAs pyramidal structures measured at various excitation intensities. (b) A single sharp PL spectrum from an InAs QD on a GaAs pyramid (full circles). The results of fitting with Gaussian are also shown with a full curve.

narrower intrinsic width. These kind of sharp PL lines are already reported in various kinds of QDs [117–121], including those in InGaAs quantum dots formed by using Stranski–Krastanow growth mode on planar substrates [122–124], and is thought to result from the delta-function-like density of states in a zero-dimensional system. These results clearly show that the position controlled InAs QDs are obtained by SA-MOVPE.

In the present experiment, the number of QDs we measure is estimated to be about 80, considering the density of pyramids and size of the laser spot. It is

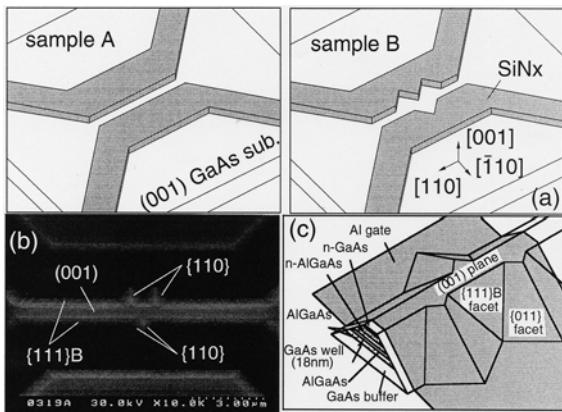


Figure 2.33. (a) A schematic illustration of the SiNx masked substrate for selective area MOVPE, (b) the SEM image after the selective growth on the masked substrate and (c) the schematic illustration of the device structure.

possible and easy for selective area MOVPE to control and reduce this number simply by making sparse patterning on the masked substrates. In fact, we have already confirmed that formation of similar InAs QDs at the top of the pyramids as well as similar GaAs pyramidal structures by the control of the growth amount of GaAs and InAs, even when the distance of the opening window is set to be $10\ \mu\text{m}$. Realization of such sparse quantum dots is expected to be effective for the spectroscopy of a single quantum dot and unveil intrinsic properties of QDs.

2.3.5 Fabrication of single electron transistors and their transport properties

2.3.5.1 Fabrication process and device structures

In this section, we describe selective area growth and fabrication of transistor structures using QWRs and QDs. Firstly, a selective doped double heterostructure was grown by selective area MOVPE on masked substrates of figure 2.33(a) at $700\ ^\circ\text{C}$. The layer sequence was as follows: a 400 nm GaAs buffer layer, a 50 nm $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$, a 15 nm GaAs well layer, a 10 nm undoped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ layer, a 50 nm n-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ layer, and a 10 nm n-doped GaAs capping layer. Typical mobility and sheet electron concentration of two-dimensional electron gas (2DEG) grown on a planar substrate were $6.2 \times 10^4\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ and $1.2 \times 10^{12}\ \text{cm}^{-2}$ at 77 K, respectively.

A schematic illustration of the mask pattern for quantum wire transistor (sample A) and an SET (sample B) structure is shown in figure 2.33(a). Large openings at both sides are for source and drain electrodes and are connected with a line opening aligned in the $[\bar{1}10]$ direction, which corresponds to a channel

of the device and quasi-one-dimensional electron gas (Q-1DEG) formed after selective area growth. The initial width of this line is about $1\text{ }\mu\text{m}$. For the patterning for SET, three prominent regions are attached to the line as illustrated in the figure. Due to these prominences, the channel has periodic variation in its width and thickness. This leads to the formation of a quantum dot near the central prominence and two tunnelling barriers beside a dot, which are connected to quantum wires.

[Figure 2.33\(b\)](#) shows the SEM image after the selective growth and (c) the schematic illustration of the device structure. A (001) plane appears on the top portion of the whole structure for both samples. The wire region consists of {111} B facet sidewalls, and {110} facet sidewalls are formed near the prominences. The thickness of the channel GaAs layer is about 18 nm which is estimated from the cathodoluminescence peak energy, as described later. The width of the channel region is measured to be 400 nm.

After the growth, Ge/Au/Ni ohmic electrodes were formed on the wide 2DEG regions for source and drain by a lift-off process followed by alloying at $350\text{ }^\circ\text{C}$ for 5 min in N_2 atmosphere. An Al Schottky electrode for the gate was formed by a lift-off process on the channel and the prominent region. Since no growth occurs on {111} B facets, both edges of the Q-1DEG channel are exposed to air. Therefore, the Schottky gate has direct contact at the edges of the Q-1DEG, and controls the electric field perpendicular to the edges. The gate length is 2 or $5\text{ }\mu\text{m}$.

2.3.5.2 Quantum wire transistors

[Figure 2.34](#) shows the drain current I_{D} -gate voltage V_{G} characteristics of sample A measured at constant source-drain bias V_{DS} of 0.2 mV at 3.4 K. Conductance oscillations are observed near the pinch-off voltage. This conductance oscillation is thought to result from a random interference due to impurity scattering or the fluctuations in the width of the channel of the quasi-one-dimensional electron gas. We also measured its magnetoresistance (MR) at 3.4 K at several gate voltage bias conditions. As shown in [figure 2.35\(a\)](#), the device showed clear MR oscillations both at $V_{\text{G}} = 0\text{ V}$ and $V_{\text{G}} = -1.25\text{ V}$. From these MR, the Landau index N_{L} was plotted as a function of reciprocal magnetic field $1/B$ which gives MR maxima, and the results are shown in [figure 2.35\(b\)](#). A deviation from the straight line was observed when $V_{\text{G}} = -1.25\text{ V}$ and -1.30 V . By fitting the data to a theory based on the parabolic potential approximation [125], the carrier density and the effective channel width can be estimated as summarized in [table 2.1](#). These results indicate that the channel width as well as carrier density can be changed by controlling the gate voltage via expansion of the depletion layer, and Q-1DEG are formed in the channel of our device.

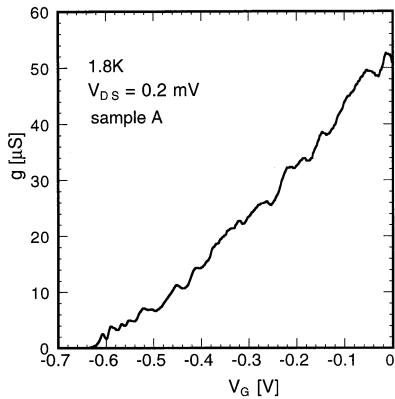


Figure 2.34. I_D-V_G characteristics of a QWR transistor (sample A).

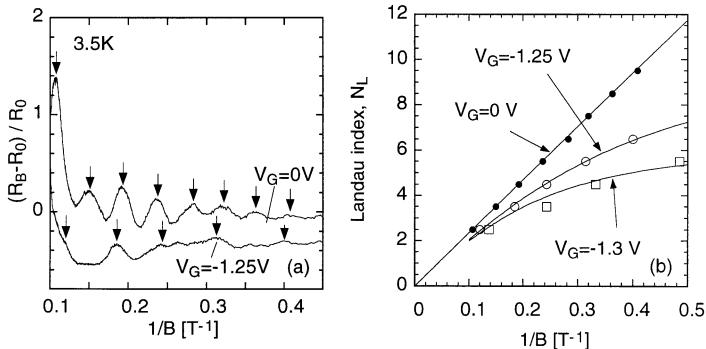


Figure 2.35. (a) Magnetoresistance oscillations of a QWR transistor measured for $V_G = 0$ V and $V_G = -1.25$ V. (b) A Landau plot of MR oscillation of (a).

Table 2.1. Calculated carrier densities, effective width of the Q-1DEG channel, and the effective confinement potential $\hbar\omega_0$ in the parabolic potential approximation.

Gate voltage	Carrier density	Effective width	$\hbar\omega_0$
0 V	$1.12 \times 10^{12} \text{ cm}^{-2} (N_e^{2D})$	—	—
-1.25 V	$1.07 \times 10^7 \text{ cm}^{-1} (N_e^D)$	105 nm	3.6 meV
-1.35 V	$6.72 \times 10^6 \text{ cm}^{-1} (N_e^{1D})$	66 nm	5.8 meV

2.3.5.3 Single electron transistor

Figure 2.36 shows I_D-V_G characteristics of sample B at $V_{DS} = 0.2$ mV for various temperatures. Clear conductance oscillations were observed near the

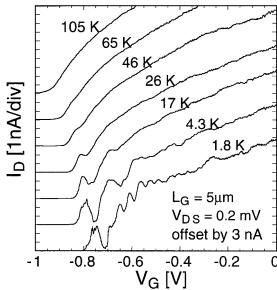


Figure 2.36. Temperature dependence of I_D – V_G characteristics.

pinch-off, and oscillations were observed up to 65 K. If we compare these results with those of sample A, the amplitude of oscillation is much larger and the width of conductance peak is much narrower. In addition, clear and very similar conductance oscillations were observed in devices we fabricated with the same patterning, although the pinch-off gate voltages were slightly different in each device due to run-to-run fluctuation of the doping carrier density for the MOVPE growth. These results suggest that three prominences in the channel region play an important role for the present conductance oscillations.

To investigate the origin of conductance oscillation near the threshold, we measured I_D – V_{DS} characteristics at 2.1 K for various gate voltage conditions, and the typical results are shown in figure 2.37(a). We can see that the conductance near the zero bias voltage first decreases as V_G , but then decreases for further increases of V_G . In addition, nonlinearity in the I_D – V_{DS} characteristics is enhanced when the conductance around zero bias is small. These are thought to be the Coulomb gaps and their modulation by gate voltage, as plotted in figure 2.37(b). Thus this result together with the oscillation of the conductance as a function of V_G , indicates that sample B operates as a single electron transistor. This also indicates that quantum dot is formed beneath the gate, and that it is connected with quantum wires with tunnelling barriers. The maximum Coulomb gap Δ is about 12 mV, which corresponds to a thermal energy $k_B T$ of 140 K. Following a simple model of Coulomb blockade, the total capacitance C_Σ of the dot can be estimated to be about 13 aF from $\Delta = e/C_\Sigma^2$. The radius R of the dot is estimated to be about 14 nm, using the relation of $C_\Sigma = 8\epsilon_r\epsilon_0 R$.

We would like to discuss here a possible mechanism for the formation of small dots and tunnelling barriers. In order to investigate where the quantum dots and tunnel barriers are formed, we made a similar sample with a much thinner GaAs channel layer, and investigated the thickness of the GaAs layer by cathodoluminescence measurement. The result of CL measurement is shown in figure 2.38. In the channel region, we can see the cathodoluminescence at 1.72 eV, which corresponds to the GaAs layer thickness of 2.5 nm. At the prominent regions, on the other hand, we can see the emission at 1.68 eV, which corresponds

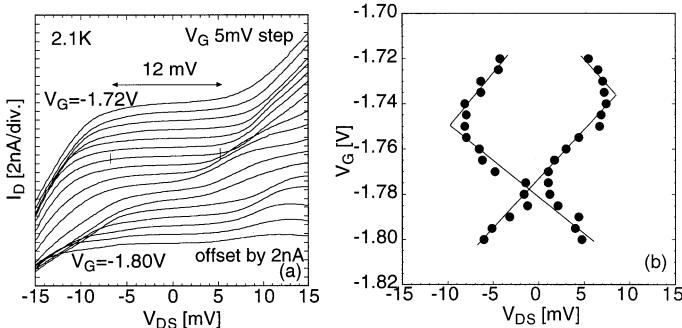


Figure 2.37. (a) I_D - V_G characteristics of sample B measured at 2.1 K. (b) The variation of Coulomb gaps by gate voltage V_G (Coulomb diamond).

to the thickness of 3.5 nm. In between the two emissions in the channel and the prominent regions there seems to be higher energy regions on both sides of the prominent regions. In addition, the energy seems to be higher in the middle of the prominent regions. From the results of cathodoluminescence, we model the formation of quantum dots and barriers as follows. At sufficiently small gate bias voltage, there is a small energy variation along the channel due to thickness modulation of the channel GaAs layer (figure 2.39(a)). However, the gate voltage is negatively biased, the depletion layer from the {111} B facet and surface expand to the channel of Q-DEG, and finally, the channel is partly depleted at the regions where the channel is thin. As a result, two quantum dots together with three tunnelling barriers are formed as schematically shown in figure 2.39(b). Therefore, the electron transport near the threshold is dominated by the single electron charging effect in two quantum dots.

2.3.6 Resistance-load SET inverter circuit

Finally, we fabricated a logic circuit using an SET by selective area MOVPE [113]. The growth was carried on a masked substrate of figure 2.40(a), which was designed to operate a resistance-load SET inverter proposed by Likharev [126]. Figure 2.40(b) shows an SEM image of the device structure grown on the masked substrate. In the structure, a channel of narrow wire is formed in the [110] direction from source to drain. The prominent part in the channel region corresponds to a SET as described in the previous section. An additional channel to measure the output voltage V_{out} of the inverter is formed and attached to the channel inbetween SET and drain. The channel between drain and output electrode works as gate bias (V_R) controlled load resistance R_L of the inverter. An equivalent circuit of the device is shown in figure 2.40(c). The length between source and drain electrodes was 10 μm , and the widths of the Al Schottky contact for the SET- and R_L -control gates were 2 μm and 1 μm , respectively. It is

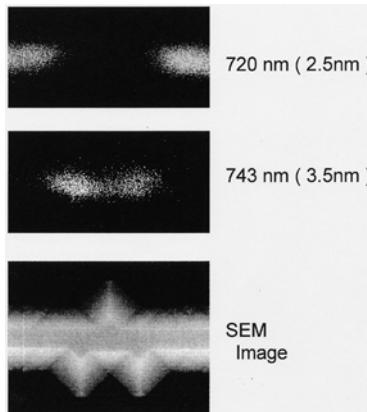


Figure 2.38. Spatially resolved cathodoluminescence images of sample B at photon energies of (a) 1.72 eV and (b) 1.68 eV.

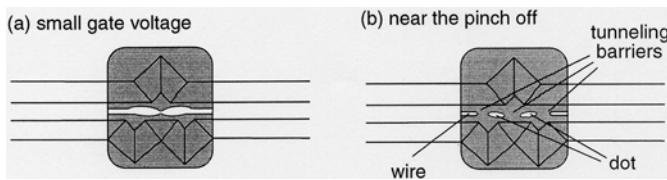


Figure 2.39. Formation model for quantum dots and tunnelling barriers.

possible to set the optimized value of R_L in the inverter circuit by only applying the gate voltage V_R . The initial resistance value of R_L without applying the V_R was roughly estimated to be several kilo-ohms, then, this value can be controlled, for example, about $60\text{ k}\Omega$ by applying the V_R of -3.6 V in this experiment.

The drain current, I_D versus SET-control gate voltage, V_G characteristics of this device were measured at 1.9 K with V_R of 0 V and -3.6 V for the constant source-drain voltage V_{DS} of 0.2 mV , and the results are shown in figure 2.41. Clear and similar conductance oscillations near the pinch-off gate voltage were observed for both V_R of 0 V and -3.6 V . In particular, the periods of oscillation peaks were the same for both measurements, indicating that V_R did not affect the conductance oscillations at all. Figure 2.42(a) and (b) show I_D-V_{DS} characteristics of this device with V_R of 0 V and -3.6 V , and with different V_G values indicated in figure 2.41(a) and (b) by arrows, respectively. Clear Coulomb gaps and their modulations were observed for both cases. The maximum Coulomb gap Δ in this sample is about 6 mV , and total capacitance, C_Σ of the dot can be estimated to be about 27 aF , which corresponds to the dot radius R of 28 nm . These values are very similar to those described in the previous section with different devices, indicating the reproducibility of our SET

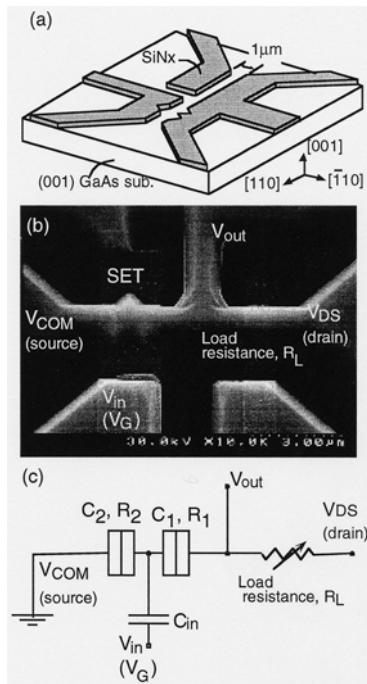


Figure 2.40. (a) A mask pattern for a resistance-load SET inverter circuit proposed by Likharev [126]. (b) A SEM image of the device structure grown on this masked substrate. (c) The equivalent circuit of the device.

fabrication method. They also indicate that V_G and V_R do not interfere with each other, and can be controlled independently.

Next, I_D – V_{in} (input voltage) and V_{out} (output voltage)– V_{in} characteristics were measured at 1.9 K under the constant V_{DS} (V_{dd}) condition. Figure 2.43 shows I_D – V_{in} and V_{out} – V_{in} characteristics of this inverter circuit. V_{DS} was 0.5 mV. V_{in} corresponds to the gate voltage, V_G for the SET. Open and full circles correspond to I_D versus V_{in} and V_{out} versus V_{in} in the Coulomb blockade region near the pinch-off gate voltage. V_{peak} in the horizontal-axis means one of the conductance peaks in the Coulomb oscillations. As V_{in} was swept from V_{peak} (i.e. V_{in} – V_{peak} of zero) to the negative gate voltage, V_{out} was smoothly increased. This result clearly shows that this structure works as an inverter. From I_D – V_G and I_D – V_{DS} measurements with applying the gate voltage V_R to R_L , conductance oscillation was confirmed to be originated from Coulomb blockade effects.

At present, the logic swing, ΔV_{out} relative to ΔV_{in} ($\Delta V_{out}/\Delta V_{in}$) is small. We can fabricate higher gain inverter circuits by further optimization of device parameters, such as, the value of load resistance, tunnelling resistance, dot size and so on.

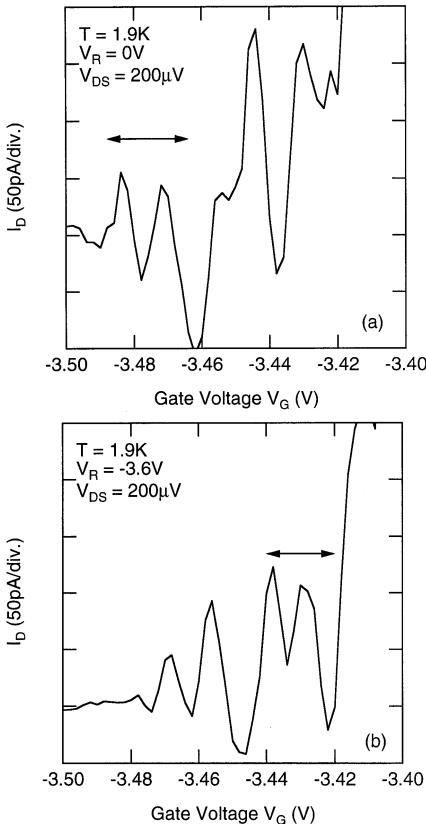


Figure 2.41. Drain current I_D versus SET-control gate voltage, V_G characteristics of the device with V_R of (a) 0 V and (b) -3.6 V for the constant source-drain voltage, V_{DS} of 0.2 mV measured at 1.9 K.

2.4 Summary and outlook

We have described the formation of quantum wire (QWR) and quantum dot (QD) structures by metal-organic vapour phase epitaxial (MOVPE) growth, and their characterization and application to the quantum devices, such as electron interference devices, QWR lasers, single electron transistors, and single electron circuits. Our results demonstrate that MOVPE growth is versatile and a powerful technique for the formation and variety of quantum nanostructures and devices.

First, multatomic steps are used to introduce thickness modulation of the well layer of GaAs/AlGaAs and InGaAs/GaAs heterostructures, and thus to realize lateral quantum confinement. Spontaneous formation of multatomic steps during MOVPE growth on vicinal (001) GaAs surfaces were investigated experimentally and also theoretically by Monte Carlo simulation. The one-

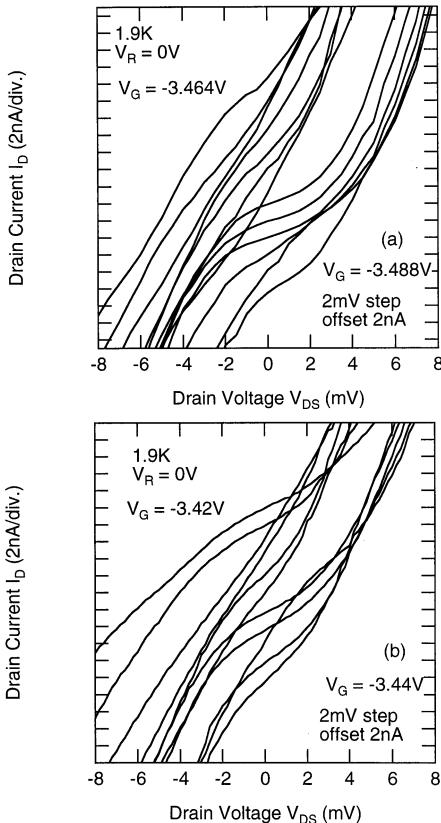


Figure 2.42. I_D – V_{DS} characteristics of the device with V_R of (a) 0 V and (b) −3.6 V, and with different V_G values indicated by the arrows in figure 2.41 (a) and (b).

dimensional nature of our QWRs were verified by the systematic study of their optical properties. Quantum wires and QWR lasers were successfully formed on vicinal surfaces. The threshold current in QWR lasers was found to be smaller when the cavity direction was perpendicular to the quantum wires as compared to the case when the cavity and quantum wires are parallel to each other, which can be ascribed to a result from the anisotropic nature of dipole moment in QWRs originating from lateral quantum confinement.

Next, selective-area MOVPE growth was used to form GaAs and InAs quantum dots on GaAs pyramidal structures. A variety of quantum structures including QWRs, QDs and quantum dot network, has been demonstrated by SA-MOVPE. The formation of zero-dimensional electronic states was demonstrated by the sharp photoluminescence from InAs QD formed at the top of GaAs pyramidal structures. A new type of quantum structures in which GaAs quantum

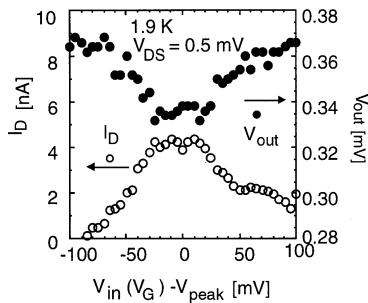


Figure 2.43. $V_{\text{out}}-V_{\text{in}}$ and $I_{\text{D}}-V_{\text{in}}$ characteristics of the inverter circuit.

dots were connected with GaAs quantum wires were also applied to GaAs single electron transistors (SETs) and resistance-load single electron inverter circuits. In SETs, Coulomb blockade type conductance oscillations were observed at low temperatures near the pinch-off voltage, and the inverter operation using this SET with load resistance was confirmed.

With all the above achievements, a lot of improvements are required in the presently fabricated quantum nanostructures for their size, size fluctuation, density, and so on. These points are particularly important when we think of practical device application. For this purpose, more detailed understanding of crystal growth, for example, during the formation of multiautomatic steps, in selective area growth, or in Stranski-Krastanow mode, in microscopic and macroscopic scale are necessary. However, this in turn suggests that there are still a lot of possibilities to be explored for the formation of quantum nanostructures by crystal growth. We also need further understanding of the physics of semiconductor nanostructures through the characterization of the optical and electrical properties of unique structures ranging from single isolated QWR and QD, arrays of QWRs and QDs, to coupled QWRs and QDs, some of which are not realized without crystal growth techniques. For these reasons, efforts should be continued to seek some breakthrough in the formation of quantum nanostructures not only for the understanding of their physics but also for their device application.

Acknowledgments

The authors acknowledge Drs J Ishizaki, S Hara, and K Kumakura and T Umeda, and F Nakajima, for their generous contribution to the present work, and, K Ohkuri, K Nakakoshi, M Akabori, and Y Oda for experimental support and fruitful discussions. They also wish to thank Professor H Hasegawa for fruitful discussions and constant encouragement. This work is partly supported by a grant-in-aid for specially promoted research on ‘Formation and Characterization of High-Density Semiconductor Quantum Dots’ from the Ministry of Education, Science, Sports and Culture, Japan.

References

- [1] Bastard G 1988 *Wave Mechanics Applied to Semiconductor Heterostructures* (New York: Halsted)
- [2] Sakaki H 1980 *Japan. J. Appl. Phys.* **19** L735
- [3] Arakawa Y and Sakaki H 1982 *Appl. Phys. Lett.* **40** 939
- [4] Beenakker C J W and van Houten H 1991 *Solid State Physics* vol 44, ed H Ehrenreich and D Turnbull (New York: Academic) p 1
- [5] Petroff P M, Gossard A C and Wiegmann W 1984 *Appl. Phys. Lett.* **45** 620
- [6] Fukui T and Saito H 1987 *Appl. Phys. Lett.* **50** 824
- [7] Nakata Y, Ueda O, Takeuchi A, Nakamura S and Muto S 1995 *J. Cryst. Growth* **150** 341
- [8] Fukui T and Saito H 1990 *Japan. J. Appl. Phys.* **29** L483
- [9] Inoue K, Kimura K, Maehashi K, Hasegawa S, Nakashima H, Iwane M, Matsuda O and Murase K 1993 *J. Cryst. Growth* **127** 1041
- [10] Takeuchi M, Shiba K, Sato K, Huang H K, Inoue K and Nakashima H 1995 *Japan. J. Appl. Phys.* **34** 4411
- [11] Hara S, Ishizaki J, Motohisa J, Fukui T and Hasegawa H 1994 *J. Cryst. Growth* **145** 692
- [12] Brasil M J S P, Bernussi A A, Cotta M A, Marquezini M V, Brum J A, Hamm R A, Chu S N G, Harriott L R and Temkin H 1994 *Appl. Phys. Lett.* **65** 857
- [13] Nakamura Y, Koshiba S and Sakaki H 1993 *Appl. Phys. Lett.* **69** 4093
- [14] Asai H, Yamada S and Fukui T 1987 *Appl. Phys. Lett.* **51** 1518
- [15] Fukui T and Ando S 1989 *Electron. Lett.* **25** 410
- [16] Fukui T, Ando S and Fukai Y K 1990 *Appl. Phys. Lett.* **57** 1209
- [17] Ando S, Kobayashi N and Ando H 1993 *Japan. J. Appl. Phys.* **32** L1293
- [18] Ando S, Kobayashi N and Ando H 1995 *Japan. J. Appl. Phys.* **34** L4
- [19] Nishioka M, Tsukamoto S, Nagamune Y, Tanaka T and Arakawa Y 1992 *J. Cryst. Growth* **124** 502
- [20] Tsukamoto T, Nagamune Y and Arakawa Y 1992 *J. Appl. Phys.* **71** 533
- [21] Tsukamoto S, Nagamune Y, Nishioka M and Arakawa Y 1993 *Appl. Phys. Lett.* **62** 49
- [22] Fukui T, Ando S, Tokura Y and Toriyama T 1991 *Appl. Phys. Lett.* **58** 2018
- [23] Lebans J A, Tsai C S, Vahala K J and Kuech T F 1990 *Appl. Phys. Lett.* **56** 2642
- [24] Nagamune Y, Tsukamoto S, Nishioka M and Arakawa Y 1993 *J. Cryst. Growth* **126** 707
- [25] Nagamune Y, Nishioka M, Tsukamoto S and Arakawa Y 1994 *Appl. Phys. Lett.* **64** 2495
- [26] Galeuchet Y D, Rothuizen H and Roentgen P 1991 *Appl. Phys. Lett.* **58** 2423
- [27] Akasaka T, Kobayashi Y, Ando S and Kobayashi N 1997 *Appl. Phys. Lett.* **71** 2196
- [28] Kapon E, Hwang D M and Bhat R 1989 *Phys. Rev. Lett.* **63** 430
- [29] Colas E, Simphony S, Kapon E, Bhat R, Hwang D M and Lin P S D 1990 *Appl. Phys. Lett.* **57** 914
- [30] Kojima K, Mitsunaga K and Kyuma K 1990 *Appl. Phys. Lett.* **56** 154
- [31] Isshiki H, Aoyagi Y, Sugano T, Iwai S and Meguro T 1993 *Appl. Phys. Lett.* **63** 1528
- [32] Wang X L, Ogura M and Matsuhata H 1995 *Appl. Phys. Lett.* **66** 1506
- [33] Sogawa T, Ando S and Kanbe H 1994 *Appl. Phys. Lett.* **64** 472
- [34] Sugiyama Y, Sakuma Y, Muto S and Yokoyama N 1995 *Appl. Phys. Lett.* **67** 256

- [35] Nakamura Y, Koshiba S, Tsuchiya M, Kano H and Sakaki H 1991 *Appl. Phys. Lett.* **59** 700
- [36] Koshiba S, Noge H, Akiyama H, Inoshita T, Nakamura Y, Shimizu A, Nagamune Y, Tsuchiya M, Kano H, Sakaki H and Wada K 1994 *Appl. Phys. Lett.* **63** 363
- [37] Fujikura H and Hasegawa H 1995 *J. Cryst. Growth* **150** 327
- [38] Tanaka M, Tomita N, Higuchi T, Shimomura S, Adachi A, Sano N and Hiyamizu S 1995 *J. Cryst. Growth* **150** 388
- [39] Motohisa J, Kumakura K, Kishida M, Yamazaki T, Fukui T, Hasegawa H and Wada K 1995 *Japan. J. Appl. Phys.* **34** 1098
- [40] Chikyow T and Koguchi N 1992 *Appl. Phys. Lett.* **61** 2431
- [41] Houzay F, Guille C, Moison J M, Henoc P and Barthe F 1987 *J. Cryst. Growth* **81** 67
- [42] Fujita Sz, Nakaoka Y, Uemura T, Tabuchi M, Noda S, Takeda Y and Sasaki A 1989 *J. Cryst. Growth* **95** 224
- [43] Leonard D, Krishnamurthy M, Reaves C M, DenBaars S P and Petroff P M 1993 *Appl. Phys. Lett.* **63** 3203
- [44] Moison J M, Houzay F, Barthe F, Leprince L, Andre E and Vatel O 1994 *Appl. Phys. Lett.* **64** 196
- [45] Ahopelto J, Yamaguchi A A, Nishi K, Usui A and Sakaki H 1993 *Japan. J. Appl. Phys.* **32** L32
- [46] Oshinowo J, Nishioka M, Ishida S and Arakawa Y 1994 *Appl. Phys. Lett.* **65** 1421
- [47] Nishi K, Mirin R, Leonard D, Medeiros-Ribeiro G, Petroff P M and Gossard A C 1995 *Proc. 7th Int. Conf. IPRM (IEEE Catalog #95CH35720)* p 759
- [48] Kawabe M, Chun Y J, Nakajima S and Akahane K 1997 *Japan. J. Appl. Phys.* **36** 4078
- [49] Carlsson N, Georgsson K, Montelius L, Samuelson L, Seifert W and Wallenberg R 1995 *J. Cryst. Growth* **156** 23
- [50] Xie Q, Madhukar A, Chen P and Kobayashi N P 1995 *Phys. Rev. Lett.* **75** 2542
- [51] Tanaka S, Iwai S and Aoyagi Y 1996 *Appl. Phys. Lett.* **69** 4096
- [52] Nötzel R, Temmyo J and Tamamura T 1994 *Nature* **369** 131
- [53] Pearah P J, Chen A C, Moy A M, Hsieh K C and Cheng K Y 1993 *Appl. Phys. Lett.* **62** 729
- [54] Kim S J, Asahi H, Takemoto M, Asami K, Takeuchi M and Gonda S 1996 *Japan. J. Appl. Phys.* **35** 4225
- [55] Mukai K, Ohtsuka N, Sugawara M and Yamazaki S 1994 *Japan. J. Appl. Phys.* **33** L1710
- [56] Saito H, Uwai K and Kobayashi N 1993 *Japan. J. Appl. Phys.* **32** 4440
- [57] Yoshida J, Kikuchi A, Nomura I and Kishino K 1995 *Proc. 7th Int. Conf. IPRM (IEEE Catalog #95CH35720)* p 29
- [58] Pearah P J, Chen A C, Moy A M, Hsieh K C and Cheng K Y 1994 *IEEE J. Quantum Elec.* **JQE-30** 608
- [59] Kirstaedter N *et al* 1994 *Electron. Lett.* **30** 1416
- [60] Shoji H, Mukai K, Ohtsuka N, Sugawara M, Uchida T and Ishikawa H 1995 *IEEE Photon. Technol. Lett.* **12** 1385
- [61] Temmyo J, Kuramochi E, Sugo M, Nishiya T, Nötzel R and Tamamura T 1995 *Electron. Lett.* **31** 209
- [62] Arakawa T, Nishioka M, Nagamune Y and Arakawa Y 1994 *Appl. Phys. Lett.* **64** 2200

- [63] Arakawa Y, Nishioka M, Nakayama H and Kitamura M 1996 *IEICE Trans. Electron.* **E79-C** 1487
- [64] Saito H, Nishi K, Ogura I, Sugou S and Sugimoto Y 1996 *Appl. Phys. Lett.* **69** 3140
- [65] Hara S, Motohisa J and Fukui T 1998 *Solid State Electron.* **42** 1233
- [66] Sakaki H, Yusa G, Someya T, Ohno Y, Noda T, Akiyama H, Kadoya Y and Noge H 1995 *Appl. Phys. Lett.* **67** 3444
- [67] Ando S, Kobayashi N and Ando H 1996 *Japan. J. Appl. Phys.* **35** L411
- [68] Okada H, Fujikura H, Hashizume T and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 1672
- [69] Kumakura K, Motohisa J and Fukui T 1998 *Physica E* **2** 809
- [70] Nötzel R, Temmyo J, Kozen A, Tamamura T, Fukui T and Hasegawa H 1996 *Festkörperprobleme* vol 35 (Advances in Solid State Physics) p 103
- [71] Motohisa J, Tanaka M and Sakaki H 1989 *Appl. Phys. Lett.* **55** 1214
- [72] Sfaxi L, Lelarge F, Petit F, Cavanna A and Etienne B 1996 *Solid State Electron.* **40** 271
- [73] Chavez-Pirson A, Ando H, Saito H and Kanbe H 1993 *Appl. Phys. Lett.* **62** 3082
- [74] Ando H, Saito H, Chavez-Pirson A, Gotoh H and Kobayashi N 1996 *Appl. Phys. Lett.* **69** 1512
- [75] Tsubaki K, Tokura Y, Fukui T, Saito H and Susa N 1989 *Electron. Lett.* **25** 728
- [76] Ishizaki J, Goto S, Kishida M, Fukui T and Hasegawa H 1994 *Japan. J. Appl. Phys.* **33** 721
- [77] Hara S, Motohisa J, Fukui T and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** 4401
- [78] Motohisa J, Akabori M, Hara S, Ishizaki J, Ohkuri K and Fukui T 1996 *Physica B* **227** 295
- [79] Akabori M, Motohisa J, Irisawa T, Hara S, Ishizaki J and Fukui T 1997 *Japan. J. Appl. Phys.* **36** 1966
- [80] Hara S, Motohisa J and Fukui T 1997 *J. Cryst. Growth* **170** 579
- [81] Nötzel R, Ledentsov N N, Däweritz L, Ploog K and Hohenstein M 1992 *Phys. Rev. B* **45** 3507
- [82] Schönherr H P, Fricke J, Niu Z C, Friedland K J, Nötzel R and Ploog K H 1998 *Appl. Phys. Lett.* **72** 566
- [83] Higashiwaki M, Yamamoto M, Higuchi T, Shimomura S, Adachi A, Okamoto Y, Sano N and Hiyamizu S 1996 *Japan. J. Appl. Phys.* **35** L606
- [84] Skala S L, Chou S T, Cheng K-Y, Tucker J R and Lyding J W 1994 *Appl. Phys. Lett.* **65** 722
- [85] Hata K, Kawazu A, Okano T, Ueda T and Akiyama M 1993 *Appl. Phys. Lett.* **63** 1625
- [86] Ohkuri K, Ishizaki J, Hara S and Fukui T 1996 *J. Cryst. Growth* **160** 235
- [87] Schwoebel R L and Shipsey E J 1996 *J. Appl. Phys.* **37** 3682
- [88] Schwoebel R L 1969 *J. Appl. Phys.* **40** 614
- [89] Ishizaki J, Ohkuri K and Fukui T 1996 *Japan. J. Appl. Phys.* **35** 1280
- [90] Fukui T and Saito H 1988 *J. Vac. Sci. Technol. B* **6** 1373
- [91] Tokura Y, Saito H and Fukui T 1989 *J. Cryst. Growth* **94** 46
- [92] Ishizaki J, Ishikawa Y and Fukui T 1997 *Mater. Res. Soc. Symp. Proc.* **448** 95
- [93] Hara S, Motohisa J and Fukui T 1998 *Electron. Lett.* **34** 894
- [94] Kitamura M, Nishioka M, Oshinowo J and Arakawa Y 1995 *Appl. Phys. Lett.* **66** 3663
- [95] Hiramoto K, Tsuchiya T, Sasagawa M and Uomi K 1994 *J. Cryst. Growth* **145** 133
- [96] Tarucha S, Okamoto H, Iwasa Y and Miura N 1984 *Solid State Commun.* **52** 815

- [97] Arent D J, Deneffe K, Van Hoof C, De Boeck J and Borghs G 1989 *J. Appl. Phys.* **66** 1739
- [98] Citrin D S 1992 *Phys. Rev. Lett.* **69** 3393
- [99] Akiyama H, Koshiba S, Someya T, Wada K, Noge H, Nakamura Y, Inoshita T, Shimizu A and Sakaki H 1994 *Phys. Rev. Lett.* **72** 924
- [100] Gershoni D, Katz M, Wegscheider W, Pfeiffer L N, Rogan R A and West K 1994 *Phys. Rev. B* **50** 8930
- [101] Higashiwaki M, Kuroyanagi K, Fujita K, Egami N, Shimojmura S and Hiyamizu S 1998 *Solid State Electron.* **42** 1581
- [102] Sugawara M, Okazaki N, Fujii T and Yamazaki S 1993 *Phys. Rev. B* **48** 8102
- [103] Akabori M, Yamatani K, Motohisa J and Fukui T 1999 *J. Cryst. Growth* **195** 579
- [104] Hu S-Y, Miller M S, Young D B, Yi J C, Leonard D, Gossard A C, Petroff P M, Coldren L A and Dagli N 1993 *Appl. Phys. Lett.* **63** 2015
- [105] Hu S-Y, Yi J C, Miller M S, Leonard D, Young D B, Gossard A C, Dagli N, Petroff P M and Coldren L A 1995 *IEEE J. Quantum Electron.* **JQE-31** 1380
- [106] Asada M, Miyamoto Y and Suematsu Y 1985 *Japan. J. Appl. Phys.* **24** L95
- [107] Yamaguchi K and Okamoto K 1991 *Appl. Phys. Lett.* **59** 3580
- [108] Ando S, Honda T and Kobayashi N 1993 *Japan. J. Appl. Phys.* **32** L104
- [109] Kumakura K, Nakakoshi K, Motohisa J, Fukui T and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** 4387
- [110] Umeda T, Kumakura K, Motohisa J and Fukui T 1998 *Physica E* **2** 714
- [111] Kumakura K, Motohisa J and Fukui T 1997 *J. Cryst. Growth* **170** 700
- [112] Kumakura K, Motohisa J and Fukui T 1998 *Solid State Electron.* **42** 1227
- [113] Nakajima F, Kumakura K, Motohisa J and Fukui T 1999 *Japan. J. Appl. Phys.* to be published
- [114] Kumakura K, Nakakoshi K, Kishida M, Motohisa J, Fukui T and Hasegawa H 1994 *J. Cryst. Growth* **145** 308
- [115] Nagamune Y, Watabe H, Nishioka M and Arakara Y 1995 *Appl. Phys. Lett.* **67** 3257
- [116] Dilger M, Haug R J, Eberl K, Kurtenbach A, Kershaw Y and Klitzing K v 1996 *Appl. Phys. Lett.* **68** 3132
- [117] Hess H F, Betzig E, Harris T D, Pfeiffer L N and West K W 1994 *Science* **264** 1740
- [118] Brunner K, Abstreiter G, Böhm G, Tränkle G and Weimann G 1994 *Appl. Phys. Lett.* **64** 3320
- [119] Gammon D, Snow E S and Katzer D S 1995 *Appl. Phys. Lett.* **67** 2391
- [120] Empedocles S A, Norris D J and Bawendi M G 1996 *Phys. Rev. Lett.* **77** 3873
- [121] Kamada H, Temmyo J, Notomi M, Furuta T and Tamamura T 1997 *Japan. J. Appl. Phys.* **36** 4194
- [122] Marzin J-Y, Gérard J-M, Izraël A, Barrier D and Bastard G 1994 *Phys. Rev. Lett.* **73** 716
- [123] Grundmann M *et al* 1995 *Phys. Rev. Lett.* **74** 4043
- [124] Hessman D, Castrillo P, Pistol M-E, Pryor C and Samuelson L 1996 *Appl. Phys. Lett.* **69** 749
- [125] Berggren K-F, Roos G and van Houten H 1998 *Phys. Rev. B* **37** 10 118
- [126] Likharev K K 1987 *IEEE Trans. Magn.* **MAG-23** 1142

Chapter 3

New quantum wire and quantum dot structures by selective MBE on patterned high-index substrates

Klaus H Ploog and Richard Nötzel

*Paul Drude Institute for Solid State Electronics, D-10117 Berlin,
Germany*

The clever combination of lithographic patterning of high-index substrates with self-organized growth and atomic hydrogen assistance allows fabrication of GaAs quantum wires and quantum dots as well as coupled wire-dot arrays with superior electronic properties by molecular beam epitaxy. The selectivity of patterned growth on high-index GaAs substrates differs qualitatively from that on low-index ones. In this paper we briefly review our recent results in this field.

3.1 Introduction

Epitaxial growth on patterned substrates has become an important technique for the monolithic integration of optoelectronic devices and for the fabrication of quantum wires (QWRs) and quantum dots (QDs). During the last decade the fabrication of QWRs and QDs in patterned growth was mainly concentrated on the formation of V-groove [1, 2] and ridge-type [3–7] structures on low-index GaAs(100) and (111) substrates using metalorganic vapour phase epitaxy (MOVPE) and molecular beam epitaxy (MBE). In the growth of GaAs/Al_xGa_{1-x}As heterostructures the plane-specific migration length of group III adatoms, being larger on the sidewalls, leads to the evolution of slow-growing side facets due to the migration of adatoms to the top of narrow ridges or to the bottom of V-grooves, which produces the nanostructures with additional lateral confinement. The aim of all these activities has been to optimize the growth conditions and pattern configuration that could lead to high shape uniformity

of the QWRs and QDs, high-density arrays, strong lateral confinements, and compatibility to subsequent processing steps. The growth of dense arrays of V-groove or ridge-type structures suffers, however, from enhanced size fluctuations and reduced growth selectivity, i.e. lateral carrier confinement, due to fast planarization of the growth front [8], while isolated deep V-grooves or ridges are difficult to process by subsequent lithographic steps. Moreover, the formation of extended side-facets of high uniformity is very sensitive to small misalignments of the pattern which easily generates strong size fluctuations giving rise to irregular dot-like optical emission patterns [9]. To overcome these problems we have studied the MBE growth of $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ heterostructures (without and with atomic hydrogen) on patterned $\text{GaAs}(311)\text{A}$ substrates. In the following we will show that on patterned high-index substrates novel growth mechanisms for selective growth result in QWR and QD structures and arrays with improved structural and electronic properties.

3.2 Results

3.2.1 Quantum wires on stripe-patterned $\text{GaAs}(311)\text{A}$

The selectivity of growth on patterned $\text{GaAs}(311)\text{A}$ substrates differs qualitatively from that on low-index ones. Mesa stripes along $[01\bar{1}]$ develop a fast growing sidewall on one side (i.e. in the sector towards the next (100) plane) during MBE of GaAs with a very smooth convex-curved surface profile without any faceting [10]. As shown in figure 3.1 the evolution of this profile is due to preferential migration of Ga adatoms from the mesa top and bottom to the sidewall, which is opposite in direction to that on patterned low-index substrates. Using mesa heights in the quantum-size regime (10–20 nm) this new growth mechanism, which is unique for $[311]\text{A}$ oriented substrates, allows the fabrication of quasi-planar lateral single and multiple $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ QWRs of several 10 nm width (figure 3.2). It is important that these sidewall QWRs exhibit narrow luminescence lines, high luminescence efficiency and strong confinement of photogenerated carriers (figure 3.3) up to room temperature [11]. Strained $\text{Ga}_x\text{In}_{1-x}\text{As}$ sidewall QWRs show strongly enhanced optical nonlinearity due to internal piezoelectric fields [12].

The shape and hence the electronic properties of these sidewall QWRs can be tuned by adjusting the mesa height through the etching process [13]. Starting from a 3 nm GaAs quantum well (QW) and $x = 0.5$ in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ barrier, the lateral QWR confinement energy reaches 100 meV for shallow mesas with step heights of 10–30 nm. This value reproduces the 2:1 growth selectivity for the thickness of the QWR to that of the QW. When increasing the mesa height to 50 nm, the confinement energy increases by a factor of two, corresponding to a QWR thickness of 11 nm. This unique flexibility in tuning the electronic properties has no equivalent counterpart in V-groove or ridge-type structures on low-index substrates.

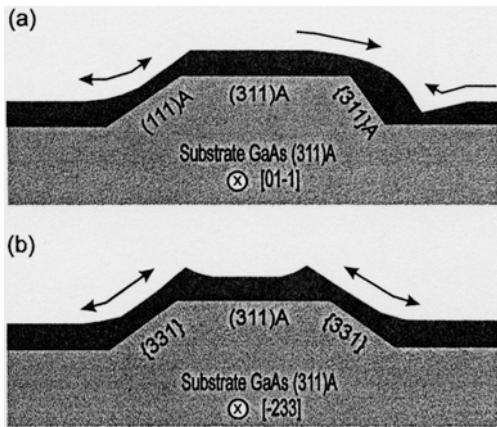


Figure 3.1. Schematic illustration of the growth mode on patterned GaAs(311)A substrates with mesa stripes oriented (a) along [01-1] and (b) along the perpendicular [-233] direction. The arrows indicate the preferential migration of Ga adatoms during MBE growth.

Dense 3D arrays of stacked GaAs sidewall QWRs have been fabricated on 500 nm pitch holographic gratings [14]. The shape and size of the wires are perfectly reproduced in the stacked layers, indicating a self-limiting growth behaviour. It is important to note that, compared to single QWRs, the uniformity and growth selectivity is even enhanced, due to the small periodicity which allows Ga adatoms to migrate directly from the (opposite) slow growing to the fast growing sidewall. In addition, the small QWR periodicity allows very efficient capture of photogenerated carriers. The luminescence spectra taken at low temperature are dominated by the emission from the wires with only a weak contribution from the adjacent QWs due to carrier localization (figure 3.4). With increasing temperature the carriers in the QWs become delocalized. Only the strong QWR emission is observed up to room temperature. No thermal repopulation of the QWs occurs at elevated temperatures owing to the large lateral confinement energy of 220 meV in the present structure.

3.2.2 Dot-like structures on square- and triangular-patterned GaAs(311)A

When GaAs(311)A substrates are patterned with arrays of square-shaped holes aligned along [011] and [233], 3D confined dot-like structures evolve in the regions between the holes on the original substrate surface during MBE of GaAs/Al_xGa_{1-x}As heterostructures (figure 3.5) [15]. The preferential migration of Ga adatoms from the sidewalls of the holes to the top generates triangular-shaped tips along [233] while the opposite (111) sidewalls inside the holes show pronounced roughening. Using triangular-shaped holes on the starting (311)A

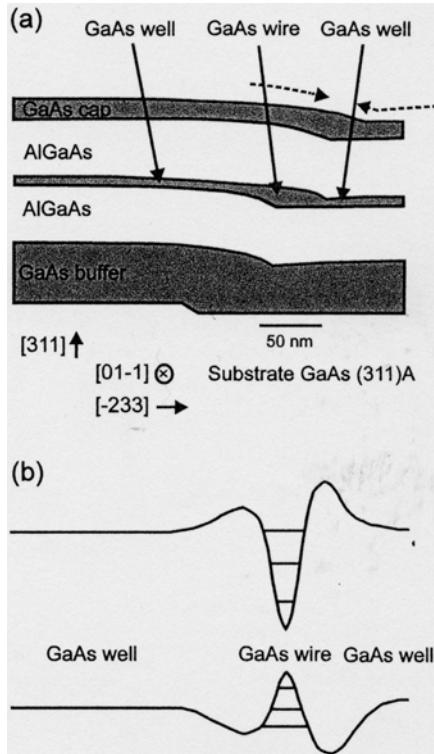


Figure 3.2. (a) Schematic cross section of the sidewall quantum wire formed along $[01\bar{1}]$ mesas on patterned GaAs(311)A substrates. (b) Real-space energy band diagram along the lateral direction. The 1D confinement originates from the strong variation of the quantum well width along $[-233]$.

surface with the acute angle along $[233]$ eliminates the rough (111) sidewalls, and similar structures having an overall smooth surface morphology can be produced. Cathodoluminescence imaging reveals distinct width variations of embedded GaAs QWs with a very high contrast definition for the triangular-hole starting pattern. This gives clear evidence for lateral confinement of carriers in the top portion of the tips, which are arranged in a lateral periodic array.

3.2.3 Coupled wire-dot arrays on stripe-patterned GaAs(311)A

The selectivity of growth, i.e. the formation of sidewall QWRs on patterned GaAs(311)A substrates, depends on the misalignment of the mesa stripes from the $[011]$ direction and on the sidewall slope [16]. For wet-chemically etched mesa stripes, which exhibit a 50° steep sidewall along $[011]$, the selectivity of growth is hardly affected for misalignments as large as 20° ([figure 3.6](#)). This makes the wire

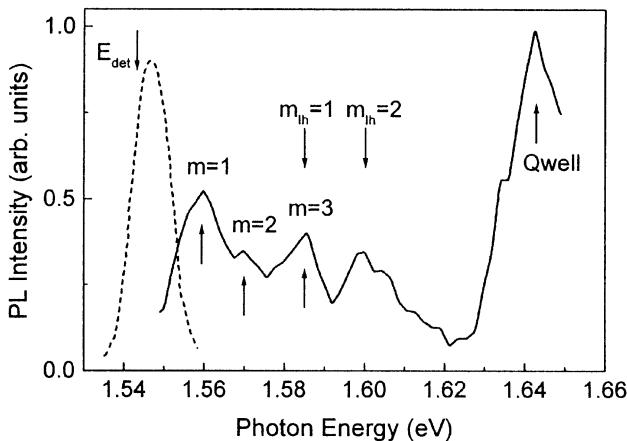


Figure 3.3. Low-temperature photoluminescence (dotted curve) and photoluminescence excitation (full curve) spectra of single GaAs sidewall quantum wire taken by near-field scanning optical microscopy [11].

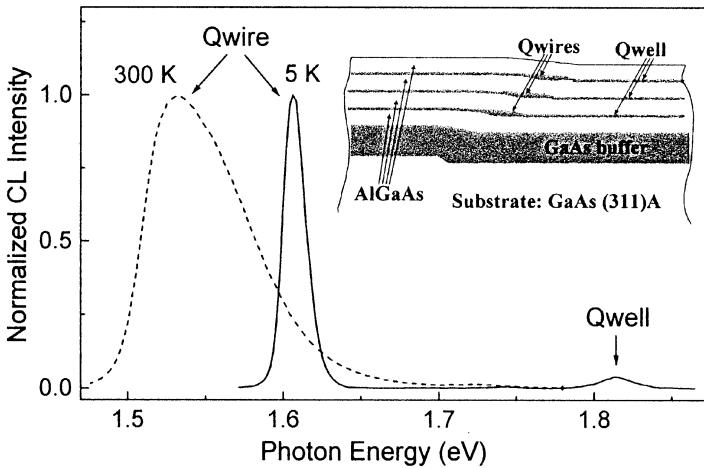


Figure 3.4. Schematic cross section of submicron arrays of vertically stacked GaAs sidewall QWRs (inset) and temperature-dependent cathodoluminescence spectra.

formation highly stable against small deviations of the mesa stripes from the [011] direction. Above 20° misalignment the selectivity of the fast growing sidewall monotonically decreases and the surface profile develops into that of the slow growing sidewall along the perpendicular [233] direction for misalignments larger than 50°. In contrast, on vertical sidewalls prepared by reactive ion etching almost no growth occurs for mesa stripes along [011]. The evolution of the fast growing

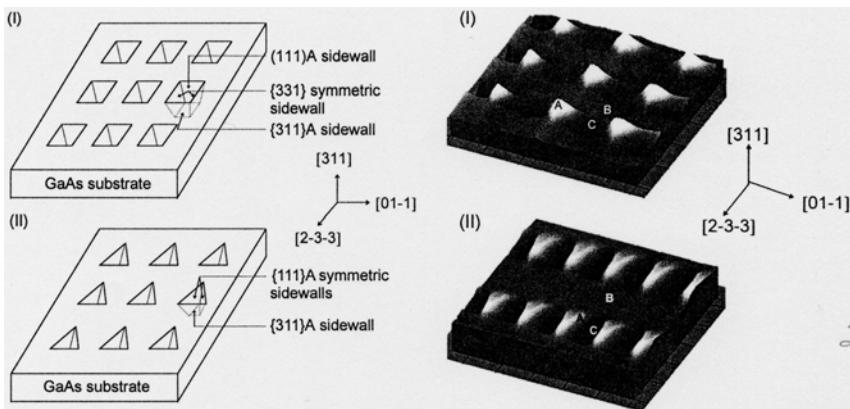


Figure 3.5. Square- and triangular-hole patterned GaAs(311)A (left) and topographic AFM images of triangular-shaped dot-like (TD) structures obtained by deposition of GaAs/Al_xGa_{1-x}As heterostructures (right).

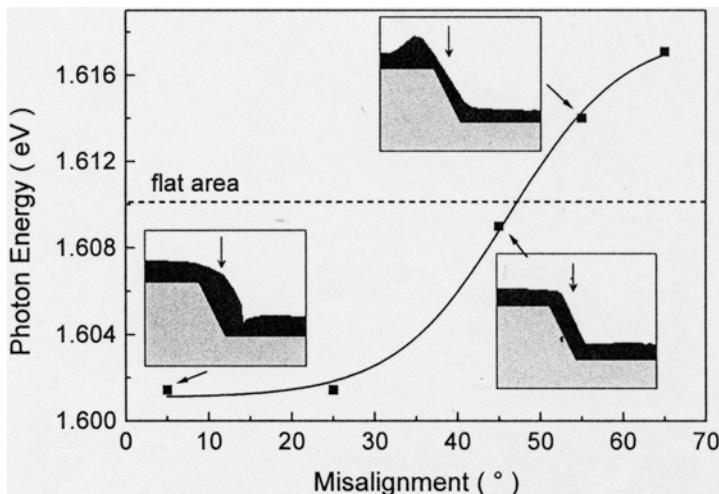


Figure 3.6. Cathodoluminescence peak energy of the mesa sidewall emission as a function of misalignment from the [01-1] direction.

sidewall is, however, recovered for misaligned mesa stripes. This complicated growth behaviour is related to the atomic configuration of the starting side facets after etching which is characterized in first approximation by the tilt of the Ga–As tetrahedron of the bulk structure with respect to the (100) and (111) planes.

The dependence of the growth selectivity on the misalignment of the mesa stripes can be directly applied to the formation of coupled wire-dot structures [16].

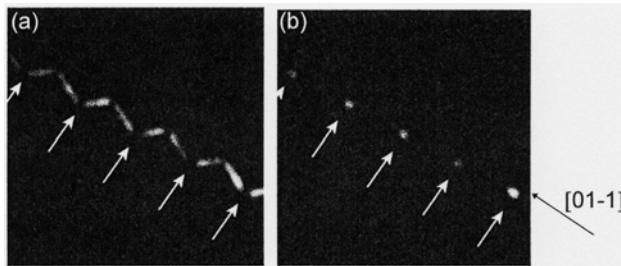


Figure 3.7. 5 K cathodoluminescence images detected at the peak energy of the wire emission at 1.668 eV along the sidewalls and at the peak energy of the dot emission at 1.644 eV in the corners.

The idea is to create a dot-like structure with locally increased layer thickness, i.e. smaller band-gap energy, at the corner of two intersecting mesas that are symmetrically inclined from the [011] azimuth to connect a wire structure with lower growth selectivity, i.e. larger band-gap energy. We have demonstrated this by fabricating a zigzag pattern with $4.5 \mu\text{m}$ long sidewalls, alternatingly misaligned by $\pm 30^\circ$ with respect to the [011] direction. Preparation of the sidewall with a depth of 15 nm was achieved by a combined CF_4 and Cl_2/N_2 RI etching process which produces a non-vertical slope. The existence of coupled wire-dot arrays, monolithically grown by MBE, was confirmed by spectrally and spatially resolved cathodoluminescence spectroscopy (figure 3.7). It is important to note that in these coupled wire-dot arrays there are no additional energy barriers between QWR and QD due to the unique lateral mass transport for their formation, i.e. they are ideal for applications as memory devices and shift registers.

3.2.4 Uniform quantum dot arrays by hydrogen-enhanced step bunching on stripe-patterned GaAs(311)A

Atomic hydrogen strongly promotes step bunching on high-index planes in well-ordered 1D step arrays with lateral periodicities on the mesoscopic scale. On GaAs(331)A substrates this has been utilized for the formation of doped $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ QWRs [17] which exhibit strong anisotropy of the electron conductivity. On GaAs(311)A substrates the hydrogen-induced periodic step arrays are oriented along [233] with a lateral periodicity of 40 nm, i.e. similar in size but perpendicular in direction to the QWRs on patterned substrates discussed before. In patterned growth, the step bunches thus generate a periodic thickness modulation of the $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ wire structure to produce a linear array of QDs along the sidewall (figure 3.8) [18]. The 3D confinement of excitons in the dots manifests itself in the splitting of the microluminescence spectra into sharp lines. For the corrugated GaAs QWs in the flat parts, the sharp lines are

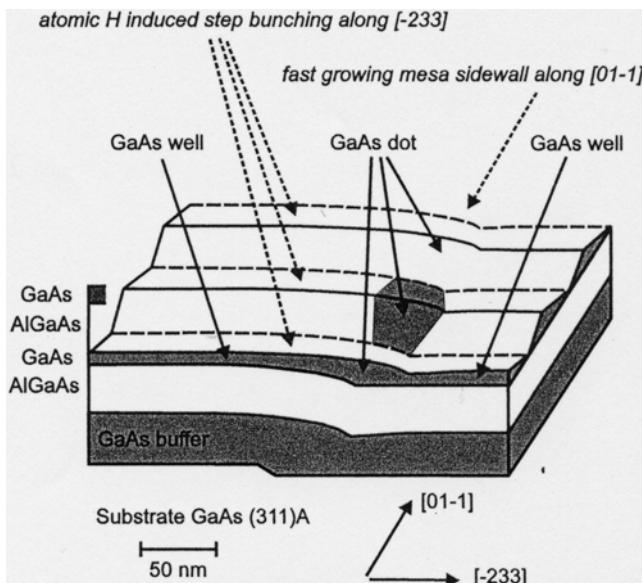


Figure 3.8. Schematic illustration of the quantum-dot array formed by the quasi-periodic step bunching along $[-\bar{2}3\bar{3}]$ in atomic hydrogen assisted MBE, perpendicular to the QWR formation in patterned growth at the fast-growing sidewall of shallow mesa stripes oriented along $[01-\bar{1}]$.

statistically distributed over the broad envelope due to exciton localization at random interface fluctuations along the step bunches. In contrast, the spectral range of the emission from the QDs at the sidewall is much narrower than the average spectra, directly evidencing the improved size uniformity. These spectra are dominated by one single sharp line without background emission over a range of 20–30 meV (figure 3.9). The observed linewidth of $60 \mu\text{eV}$ is limited by the spectrometer resolution. The spectra remain almost unchanged when the diameter of the optical probe area is increased from 2 to $16 \mu\text{m}$ corresponding to several ten to several hundred dots [18]. This area is large enough for further processing of devices.

3.3 Concluding remarks

Suggestions for QWR and QD devices (transistors, lasers, information storage, etc) have been around for quite some time, but there are still physics and related technology issues that will determine whether they assume any practical importance. Semiconductors have led the research, development and exploitation cycle of artificial multilayers and nanostructured solids. It is time now to think about 3D nanometre-scale materials engineering using much wider classes of

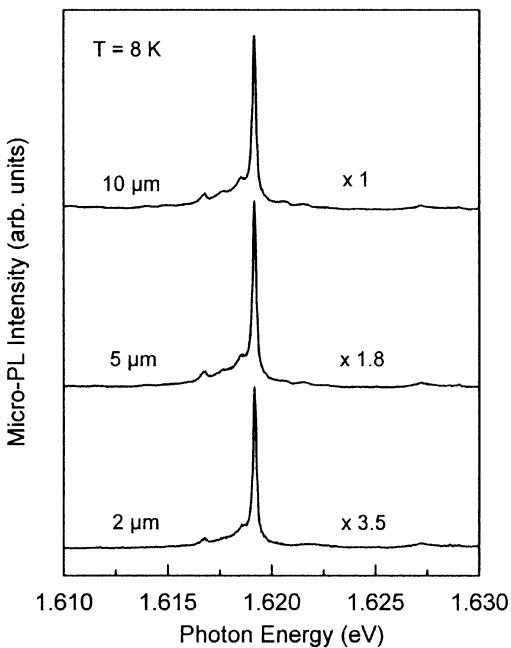


Figure 3.9. High-resolution μ -PL spectrum of GaAs quantum-dot array along the [01–1] sidewall for three different diameters of the probed area.

materials. The electrical, magnetic, optical, thermal or other properties of solids might all be tailored, restricted only by the elements in the periodic table, for application in sensing, computation, information storage and transmission. Almost every material now in use can be improved upon. This is not only with respect to its primary function but also, as exemplified by the case of semiconductor heterojunctions, the wide range of secondary attributes. One can envisage bringing together two materials properties that until now have been disparate and to exploit carefully tailored hybrid properties. The technology for this materials engineering is certainly not confined to the epitaxy/lithography route. Radical alternatives are needed to be able to put in place every atom in a solid, in order to achieve a particular set of materials properties.

Acknowledgments

The authors would like to thank their colleagues L Däweritz, J Fricke, U Jahn, Z C Niu, M Ramsteiner, H P Schönher and A Trampert for their active contributions to the work reported here. Part of this work was supported by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie of the Federal Republic of Germany.

References

- [1] Kapon E, Hwang D M and Bhat R 1989 *Phys. Rev. Lett.* **63** 430
- [2] Wang X L, Ogura M and Matsuhata H 1997 *J. Cryst. Growth* **171** 341
- [3] Walther M, Röhr T, Böhm G, Tränkle G and Weimann G 1993 *J. Cryst. Growth* **127** 1045
- [4] Rajkumar K C, Madhukar A, Rammohan K, Rich D H, Chen P and Chen L 1993 *Appl. Phys. Lett.* **63** 2905
- [5] Tsukamoto S, Nagamune Y, Nishioka M and Arakawa Y 1993 *Appl. Phys. Lett.* **63** 355
- [6] Koshiba S, Noge H, Akiyama H, Inoshita T, Nakamura Y, Shimizu A, Nagamune Y, Tsuchiya M, Sakaki H and Wada K 1994 *Appl. Phys. Lett.* **64** 363
- [7] Fujikura H and Hasegawa H 1995 *J. Cryst. Growth* **150** 327
- [8] Christen J, Grundmann M, Kapon E, Colas E, Hwang D M and Bimberg D 1992 *Appl. Phys. Lett.* **61** 67
- [9] Kihara M, Fujikura H and Hasegawa H 1997 *Appl. Surf. Sci.* **117/118** 695
- [10] Nötzel R, Menniger J, Ramsteiner M, Trampert A, Schönherr H P, Däweritz L and Ploog K H 1997 *J. Cryst. Growth* **175/176** 1114
- [11] Richter A, Behme G, Süptitz M, Lienau Ch, Elsässer T, Ramsteiner M, Nötzel R and Ploog K H 1997 *Phys. Rev. Lett.* **79** 2145
- [12] Nötzel R, Ramsteiner M, Niu Z C, Schönherr H P, Däweritz L and Ploog K H 1997 *Appl. Phys. Lett.* **70** 1578
- [13] Nötzel R, Ramsteiner M, Niu Z C, Däweritz L and Ploog K H 1997 *IEEE Int. Symp. Compound Semicond. (IEEE97TH8272)* ed M Melloch and M A Reed (Bristol: Institute of Physics Publishing) p 65
- [14] Nötzel R, Jahn U, Niu Z C, Trampert A, Fricke J, Schönherr H P, Kurth T, Heitmann D, Däweritz L and Ploog K H 1998 *Appl. Lett.* **72** 2002
- [15] Niu Z C, Nötzel R, Jahn U, Ramsteiner M, Schönherr H P, Fricke J, Xiao Z B, Däweritz L and Ploog K H 1998 *Appl. Phys. A* **67** 135
- [16] Fricke J, Nötzel R, Jahn U, Schönherr H P, Däweritz L and Ploog K H *J. Appl. Phys.* submitted
- [17] Schönherr H P, Fricke J, Niu Z C, Friedland K J, Nötzel R and Ploog K H 1998 *Appl. Phys. Lett.* **72** 566
- [18] Nötzel R, Niu Z C, Ramsteiner M, Schönherr H P, Trampert A, Däweritz L and Ploog K H 1998 *Nature* **392** 56

Chapter 4

New fabrication techniques and optical properties of GaN and Si quantum dots

Yoshinobu Aoyagi¹, Satoru Tanaka² and Xinwei Zhao¹

*¹ The Institute of Physical and Chemical Research (RIKEN),
Wako-shi, Saitama 351-0198, Japan*

*² Research Institute of Electronic Science, Hokkaido University,
12-6, Kita-ku, Sapporo 060, Japan*

A small amount of external impurity on the surface or in the bulk sometimes plays an important role in quantum dots fabrication. Quantum dots fabrication of nitride semiconductors on a quasi-lattice matched substrate of AlGaN using small amounts of external impurity, Si antisurfactant, and fabrication of nanocrystalline Si using Er impurity are reported.

4.1 Introduction

It is quite difficult to control the formation process of quantum dots using conventional methods. For example, it is not possible to fabricate GaN quantum dots of quasi-lattice matched GaAIN using Stranski-Krastanov crystal growth mode conventionally used for fabrication of self-assembled quantum dots of InGaAs and related materials. In a system of GaN/AlGaN the strain energy on the surface is not enough to form the dots. However, if a small amount of external atoms of Si are introduced on a surface of quasi-lattice matched GaAIN, the surface energy of the substrate is reduced and GaN quantum dots are formed on the surface [1–5]. In the case of self-formed nanocrystalline Si obtained by annealing amorphous Si film, it is difficult to control the size of the nano-Si, because the nucleation and enlargement process of the nanocrystal cannot be independently controlled during annealing. However, by introducing a small amount of Er in the amorphous Si film and annealing the film, the nucleation of

dots is triggered by large strain energy near the Er atom and dot size controllability becomes high.

The optical properties of dots formed using the technique is excellent and bright UV luminescence for GaN and blue for nanocrystalline Si [7–10] are shown. This bright luminescence suggests to us the possibility of applying this technology to UV quantum dots laser and all Si OEIC.

This paper introduces a new technique for the formation of quantum dots using small amounts of external atoms on the surface or in the bulk.

4.2 Formation of self-assembled GaN quantum dots using Si antisurfactant and optical properties of the dots

There are many reports on the fabrication of quantum dots of III–V compounds on a semiconductor substrate using the selective epitaxy method combined with conventional electron beam lithography and etching, or using Stranski–Krastanov (S–K) growth of self-assembled dots. However, in the former method there is a limitation in reducing the dot size. A structural defect introduced during lithography and etching procedure influences electrical and optical properties. In the latter case, there is a limitation of materials in which S–K dots formation is available, since a strain energy at the heterointerface of the lattice mismatched system is the driving force for dot formation and enough strain energy is required for dot formation. As a typical example, the dot formation of GaAs on lattice matched AlGaAs substrate is impossible using the S–K mode. In the case of nitride compounds, in general, the lattice mismatch is not so large except InN or In-rich N or GaN on AlN. The dot formation of GaN on $\text{Ga}_{1-x}\text{Al}_x\text{N}$ or $\text{Ga}_{1-y}\text{Al}_y\text{N}$ on $\text{Ga}_{1-x}\text{Al}_x\text{N}$, which are very important for future UV lasers is very difficult. In this system two-dimensional crystal growth mode becomes dominant and dot formation is basically impossible in a conventional growth technique. Recently a new technique to form dots of GaN on $\text{Ga}_{1-x}\text{Al}_x\text{N}$ by distributing a small amount of Si atom on a $\text{Ga}_{1-x}\text{Al}_x\text{N}$ surface which acts as an antisurfactant is developed [1, 2]. By introducing this antisurfactant, surface free energy can be artificially controlled and the growth mode can be changed from two-dimensional mode to dots formation mode [1, 2]. This technique is quite general and the dot formation of InGaN [4] and $\text{Ga}_{1-y}\text{Al}_y\text{N}$ [5] on $\text{Ga}_{1-x}\text{Al}_x\text{N}$ are also successfully realized. In this section, the dot formation technique and the formation mechanism using the antisurfactant are reported and optical properties of the dots are discussed.

4.2.1 Experimental procedure

GaAlN thin film and GaN dots are grown using a horizontal MOCVD reactor operated at 76 Torr. Tetraethyl Si (TESi) of 0–220 nmol is fed on a $\text{Ga}_{0.88}\text{Al}_{0.12}\text{N}$ barrier layer grown on $\text{Ga}_{0.80}\text{Al}_{0.20}\text{N}$ buffer layer on 6H-SiC(0001)-Si surface.

GaN is grown after feeding of the TESi using trimethyl Ga (TMG)/NH₃. For measurement of surface morphology using the atomic force microscope (AFM), the growth is interrupted just after the TMG/NH₃ feeding and temperature was decreased to room temperature under an NH₃ environment. For optical measurement, the capping layer of GaAIN or AlN is grown on the GaN layer. The real amount of Si reaching the surface and reacting to AlGaN cannot be clearly defined at this moment, but it can be estimated roughly to be less than 0.01 monolayer for 3.4 nmol feeding of TESi from the calculation of reactor structure and sticking coefficient of TESi on the GaAIN. The samples grown by this method are evaluated using the high-resolution transmission electron microscope (HRTEM) and photoluminescence (PL).

4.2.2 Experimental results

[Figure 4.1](#) shows the GaN surface morphology observed by AFM as a function of the feeding rate of TESi. As shown in figure 4.1(a) surface morphology of GaN grown on GaAIN without Si on the surface shows a clear step pattern (two-dimensional morphology). It has been confirmed from AFM measurement that the GaAIN cladding layer is also grown under step flow growth mode. Therefore, it is clear that the GaN on GaAIN is grown under two-dimensional growth (step flow growth mode) without Si atoms on the surface of GaAIN. The height of the step is about 0.5–0.7 nm corresponding to 2–3 monolayers of GaN. By feeding TESi, the surface morphology clearly changes. As shown in figure 4.1(b), under the feeding rate of 22 nmol of TESi, the GaN surface shows clearly three-dimensional growth and the dot formation is started. The morphology of step flow growth partially remains on the surface but the island nucleation along the step is observed in part of the surface. As shown in figure 4.1(c), under the feeding rate of 220 nmol of TESi, large islands are formed on the surface and the density and size are controlled by controlling the growth temperature and feeding rate of TESi.

[Figure 4.2](#) shows the HRTEM image of the GaN dots. The dot has a disk-like structure with a dimension of about 30 nm (width) and 5 nm (height) and the dots are grown epitaxially on the GaAIN cladding layer. All of the grown dots are observed to be wurtzite (2H) structure within RHTEM resolution and other crystal structures like Si crystal cubic are not observed. Therefore, there is no possibility of Si clustering on the surface. However, the site of the absorption of Si is not clear at this moment.

This technique using small amounts of external atoms on the surface (which acts as an antisurfactant and controls the surface energy of the substrate as discussed later) for dot formation has been applied to form Ga_xIn_{1-x}N dots on Ga_{1-x}Al_xN substrate or Ga_{1-y}Al_yN dots on Ga_{1-x}Al_xN substrate. [Figure 4.3](#) shows the AFM image of GaAIN dots formed using this technique. For the GaAIN dot formation, the growth temperature of the dots should be decreased to suppress the surface migration of Al.

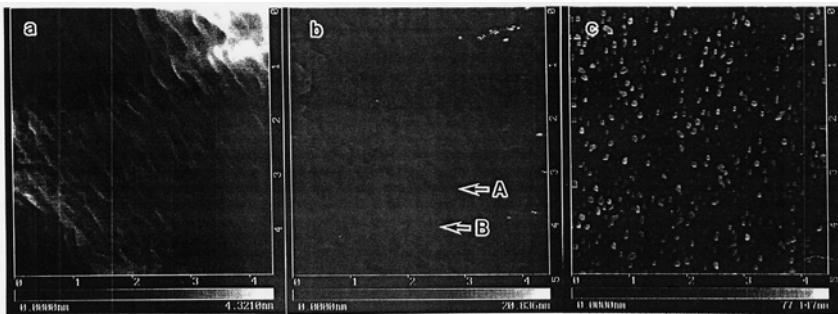


Figure 4.1. The AFM image of the GaN surface depending on TESi feeding rate. (a) Without TESi feeding. Step flowing from left-up to right-down is clearly seen. (b) With 22 nm TESi feeding. Small amounts of TESi suppress the step flow growth mode. The island appears (B). (A) is the GaAlN surface. (c) With 220 nm TESi feeding. Dots clearly appear.

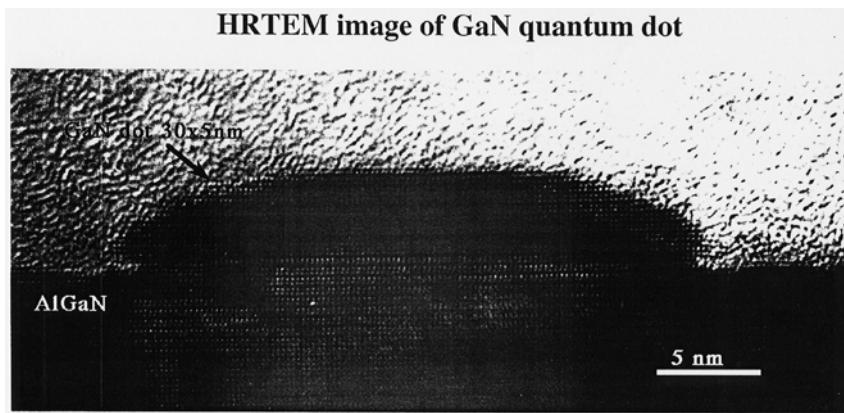


Figure 4.2. An HRTEM image of GaN dots. The dot has a size of 5 nm in height and 30 nm in diameter.

Figure 4.4 shows a luminescence spectra of GaN dots. Excitation lasers were a He-Cd laser (325 nm CW) for weak excitation conditions and a N₂ gas laser (377 nm pulse) for an experiment of laser action. Under the weak excitation condition, luminescence from dots is observed at 3.525 eV. This peak shifts toward the low-energy side for the dots with larger size, which is controlled by changing the growth time of GaN. This result suggests to us a quantum size effect of GaN dots. In this experiment a AlN capping layer is used.

Figure 4.5 shows the GaN dots laser structure. The barrier layer is Ga_{0.88}Al_{0.12}N and the cladding layer is Ga_{0.80}Al_{0.20}N. The cavity used for laser

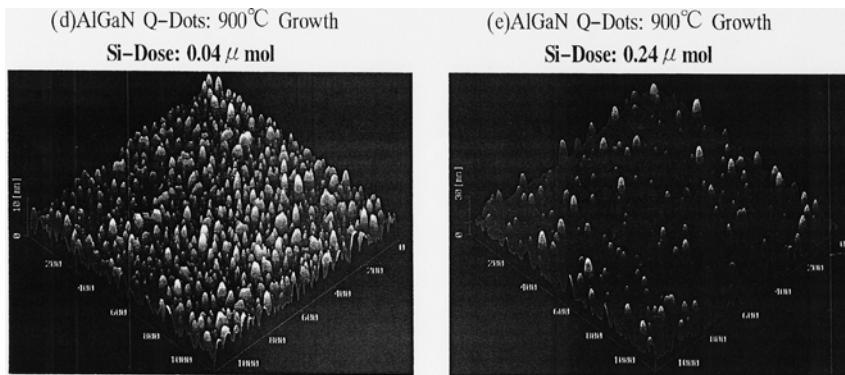


Figure 4.3. The AFM image of $\text{Ga}_{1-y}\text{Al}_y\text{N}$ dots on $\text{Ga}_{1-x}\text{Al}_x\text{N}$ substrate formed using Si antisurfactant. GaAIN dots can be formed by controlling the migration of Al and Ga by decreasing the growth temperature under TESi feeding.

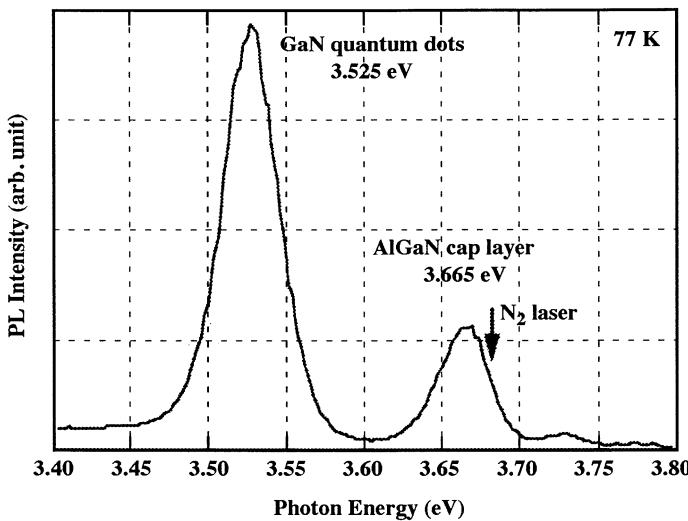


Figure 4.4. The photoluminescence spectrum of GaN dots. Strong emission from the dots at 3.525 eV is observed. The peak at the low-energy side is luminescence from the GaAIN substrate.

action was a cleaved (1100) SiC facet. A clear threshold at 0.75 MW cm^{-2} is observed above this. Narrowing of the spectrum was also observed above this threshold. These results suggest that the GaN UV dot laser is realized [3].

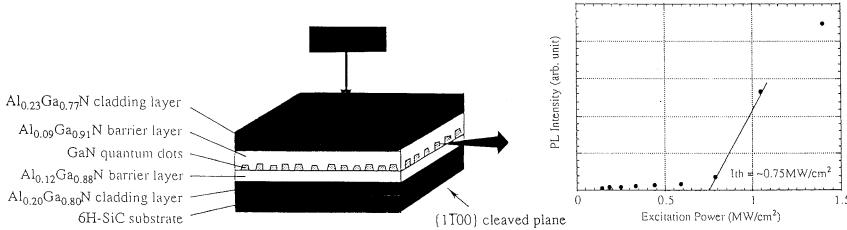


Figure 4.5. Structure of the GaN dot laser and the lasing characteristic under N₂ gas laser excitation. The clear threshold of lasing at 0.75 MW cm⁻² is observed.

4.2.3 Discussion

We discuss here the mechanism of GaN dots formation on a quasi-lattice matched Ga_{1-x}Al_xN substrate from the point of view of the kinetics and energy balance at the surface. It is an important issue for understanding the kinetics of dot formation that the step flow growth is suppressed under TESi feeding. As shown in figure 4.1(b) and (c), the transition of growth mode from step flow to island growth is related to Si coverage on the surface. According to the atomic model of the Ga_{1-x}Al_xN surface, there are N-dangling bonds at the step site. Meanwhile, Ga- or Al-dangling bonds are mainly at the terrace site and N-dangling bonds exist only at the defect site like dislocation on the terrace. In this model the polarity of the top surface of Ga_{1-x}Al_xN is the Al, Ga surface. This may be true because the Si surface of 6H-SiC(0001) is used for our experiment. We can consider that the TESi is adsorbed at the step site because there are many dangling bonds of N at the step and they easily bond to Si to form strong Si-N bonding. Though surface free energy at step seems to be very large (which can be easily considered from the result that step flow mode is dominant in GaN), this absorbed Si may reduce the free energy of step by, for example, producing reconstruction of step structure and suppressing absorption of TMG/NH₃ at the step at the moment of GaN (TMG/NH₃) growth. If the absorption of GaN at the step is not dominant, nucleation at the terrace becomes dominant and nucleation of dots are expected. In addition to this effect of Si at the step, Si can also produce Si-N bonds on the terrace and this Si-N bond can also suppress the GaN growth at the terrace.

However, even if nucleation occurs at the terrace, energy balance should be satisfied at the surface for the dot formation. In general, for the dot formation it follows that the energy balance of surface free energy should be satisfied:

$$S_{\text{substrate}} < S_{\text{film}} + S_{\text{interface}}. \quad (4.1)$$

If we neglect the interface between substrate and dots, the condition of dots formation is given by:

$$S_{\text{substrate}} < S_{\text{film}}. \quad (4.2)$$

This means that the surface free energy of the GaAlN substrate should be smaller

than that of GaN. However, considering the experimental results without feeding of TESi, bonding energy of Al–N > Ga–N and surface defect, the actual energy balance seems to be

$$S_{\text{substrate}} > S_{\text{film}}. \quad (4.3)$$

By introducing the Si atom on the surface, equation (4.3) is changed to equation (4.1).

The reduction of surface free energy D is given by

$$\begin{aligned} D(\text{reduction of surface free energy by TESi feeding}) \\ = S_{\text{substrate}} - S_{\text{substrate}(\text{after Si})} \\ = D(\text{reduction of free energy at step}(S_{\text{step}})) \\ + D(\text{reduction of free energy at the terrace}(S_{\text{terrace}})). \end{aligned} \quad (4.4)$$

That is, the step energy is reduced and the defect on the terrace is inactivated by the TESi feeding. So,

$$S_{\text{substrate}(\text{after Si})} = S_{\text{substrate}} - D < S_{\text{GaN}}. \quad (4.5)$$

Equation (4.5) is a condition of three-dimensional dots formation.

As discussed above, the mechanism of the change of the crystal growth mode and the dot formation by introducing TESi is qualitatively explained.

This model is not perfect because many assumptions are included. However, more quantitative discussion will be available by accumulating further experimental results.

4.3 Formation of nanocrystalline Si by Er doping and optical properties of the nanocrystal

Many reports discussing the mechanism of visible light emission and the application of the light emitting porous Si have been published after the report on strong red light emission from porous Si in 1990 [6]. It is also reported that green light can be emitted from oxidized porous Si. Recently, nanocrystalline Si (Nc-Si) embedded in amorphous Si is reported to emit strong blue and violet light [7–17]. In the case of porous Si it is difficult to apply the material to real light emitting devices because of fast degradation of light emission. In the case of oxidized porous Si, it seems to be difficult to apply the material to real devices because the driving voltage for electroluminescence will become large because of the insulating behaviour of oxidized Si. The Nc-Si is very stable and the wavelength of light emission is short. So, it is an interesting material for all future Si-OEIC. However, the emission mechanism of blue and violet light from the Nc-Si embedded in amorphous Si has not yet been made clear and size control of the Nc-Si is difficult. Recently, it is reported that the size of the Nc-Si can be well controlled by incorporating small amounts of Er atoms in the

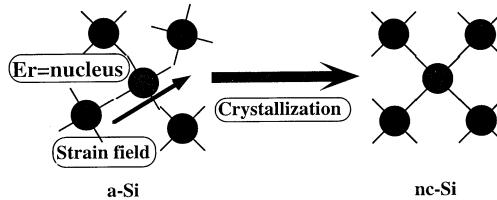


Figure 4.6. Schematic picture explaining the nanocrystalline Si formation. The strain field near Er external impurity induces nucleation of nanocrystalline Si.

film [18–22]. In dot formation, the nucleation and enlargement process of dots should be controlled independently for the control feature of dots. So far, the Nc-Si is produced by annealing the thin film of amorphous Si. However, in this case the size control of the Nc-Si was difficult because nucleation and enlargement processes are happening simultaneously during annealing. Er doped in Si reported here has large strain field near the Er atom because the atomic radius of Er is much larger than Si and this strain field can be a trigger of nucleation of nanocrystalline dots as shown schematically in figure 4.6. This means the nucleation process and enlargement process for the dot formation can be separated and controlled independently. In this section experimental results to form the Nc-Si using Er external atoms are reported and the emission mechanism of the light from the Nc-Si is discussed.

4.3.1 Experimental procedure

Er doped amorphous Si is deposited on an Si substrate by electron beam evaporation or laser ablation techniques under a background pressure of 5×10^{-9} Torr. The doping level of Er is 1×10^{16} – $1 \times 10^{21} \text{ cm}^{-3}$, and the thickness of the film was 10–500 nm. The deposited amorphous Si was annealed in vacuum or nitrogen atmosphere at a temperature of 700–1000 °C during 2 s–2 h and nanocrystalline Si was formed. The Nc-Si was evaluated using TEM, XRD and the optical properties of the Nc-Si were studied using PL, PR and absorption spectrum measurements.

4.3.2 Experimental results

Figure 4.7 shows the TEM image of the Nc-Si produced under the condition of 700 °C, 10 min annealing with Er doping of $1 \times 10^{17} \text{ cm}^{-3}$. The Nc-Si has random crystal orientation. So, only the nanocrystal having a crystal direction the same as the electron beam of TEM shows a clear lattice image. The size of the nanocrystal is about 3 nm and size distribution is not so large.

Figure 4.8 shows the average nanocrystal size grown at 700 °C as a function of annealing time. The thickness of amorphous film was 570 nm. As shown in

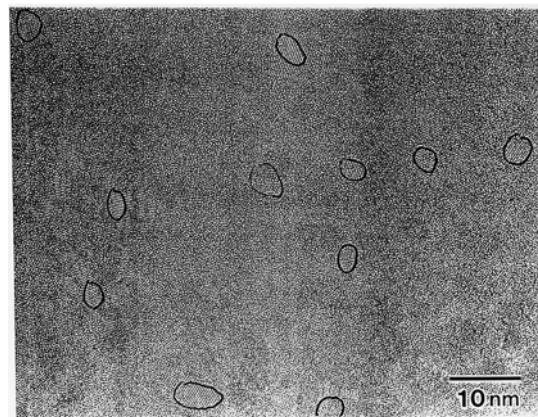


Figure 4.7. TEM image of nanocrystalline Si. The dose of Er was about $1 \times 10^{17} \text{ cm}^{-3}$. Annealing temperature was 700°C . The size of nanocrystal is about 3 nm.

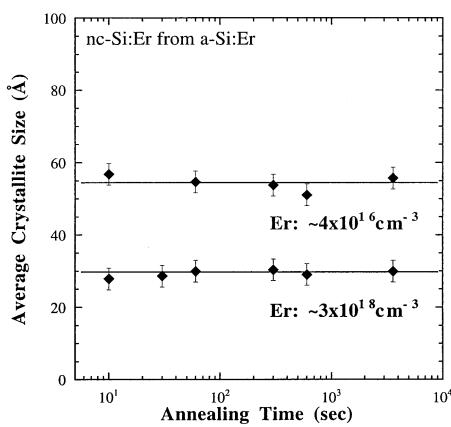


Figure 4.8. Er concentration dependence of averaged nanocrystalline Si as a function of annealing time. The average size does not depend on the annealing time and depends on the concentration of Er.

In this figure the average size of the nanocrystal does not depend on the annealing time and is well controlled. This means that the size of the nanocrystal has been determined at an initial stage of nucleation of the nanocrystal and enlargement of the nanocrystal does not occur during the annealing. If we increase annealing temperature up to 1000°C , the enlargement in size of the Nc-Si is observed as a function of annealing time. This result suggests that the enlargement of nanocrystal occurs during annealing time at this temperature. These results

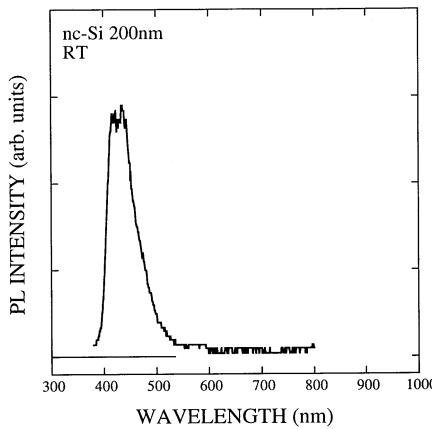


Figure 4.9. The photoluminescence spectrum of nanocrystalline Si with an average size of 3 nm. A strong blue emission at 420 nm is observed.

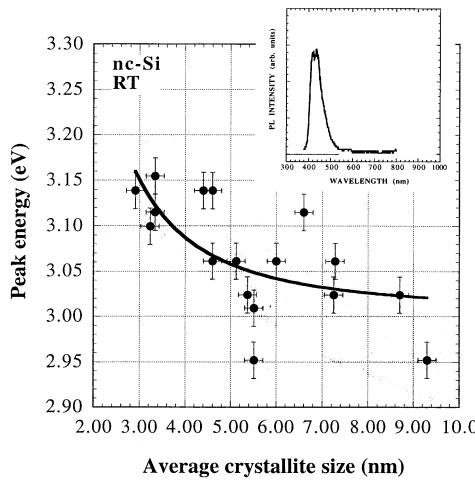


Figure 4.10. The size dependence of the wavelength of the luminescence from the nanocrystalline Si. The smaller the size is, the shorter the wavelength. This result shows clearly the size effect of nanodots.

suggest to us that nucleation of nanocrystal and enlargement can be independently controlled. The average size of the nanocrystal can be controlled by changing the annealing temperature.

Figure 4.9 shows the photoluminescence spectra of the Nc-Si at room temperature using the excitation of a He-Cd laser. A strong emission at 420 nm was observed and this wavelength is largely different from the spectrum for

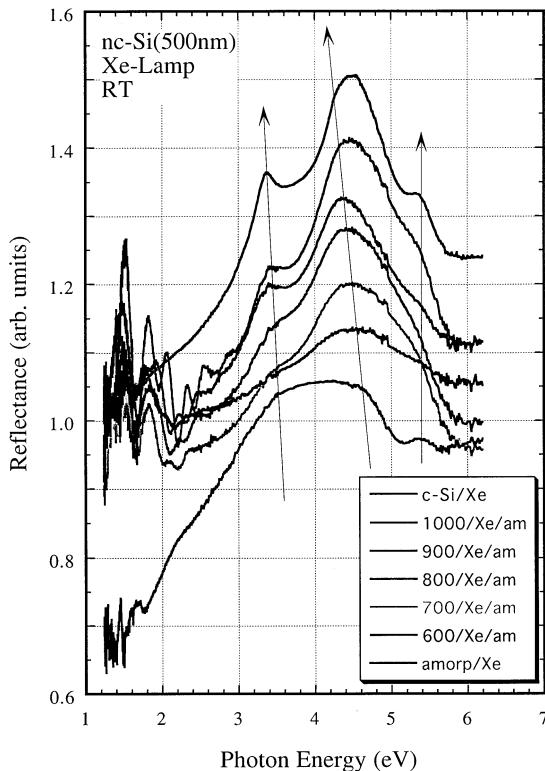


Figure 4.11. Optical absorption spectrum of nanocrystalline Si as a function of crystal size. The smaller the size is, the higher energy side of the absorption spectrum. A size effect appears also in the optical absorption spectrum.

porous Si and oxidized porous Si. Under some growth conditions violet emission of 390 nm is observed.

Figure 4.10 shows the size dependence of photoluminescence. The peak shifts to lower energy increasing the average size of the nanocrystal. Figure 4.11 shows the absorption spectrum of the Nc-Si with different average size. As shown by an arrow the absorption peak shifts to the lower energy side increasing the average size of nanocrystal. This size dependence of photoluminescence and absorption spectra show us that the origin of the luminescence is not related to the defect or oxidized interface but intrinsic phenomena from the Nc-Si. The Nc-Si produced by our technique is embedded in amorphous Si and has no potential barrier for carrier confinement. However, as shown in figures 4.10 and 4.11 we can see a size effect of optical properties and this size effect is different from the one observed in conventional potential confinement. The lifetime of the luminescence from Nc-Si embedded in amorphous Si was about 500 ps and is a

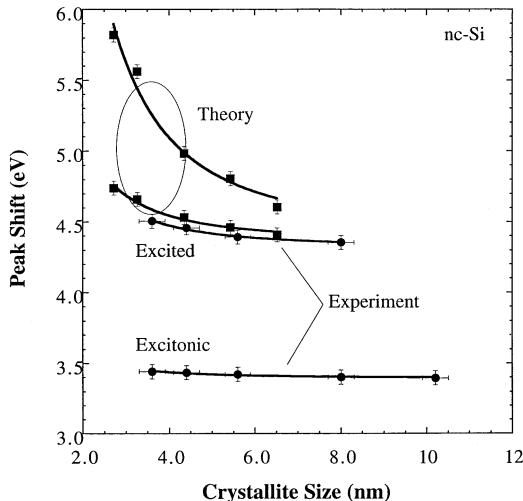


Figure 4.12. Experimental and theoretical results of size dependence of the absorption spectrum of nanocrystalline Si embedded in amorphous Si. A good agreement between experimental results and calculation is obtained.

value for direct transition of the semiconductor. We carried out a direct calculation of electronic state of nano-Si of about 10 000 atoms surrounded by amorphous Si and evaluated the size dependence of the absorption by changing the number of atoms of the nanocrystal. The experimental results are well explained by our calculation. These results suggest to us that the size effect can appear as a balance of the decoherent nature of an electron in amorphous phase and the coherent nature of an electron in a nanocrystal. Details will be reported elsewhere.

4.4 Conclusion

The crystal growth mode can be controlled by adding small amounts of external impurity on the surface of epitaxial substrate or in a film. We succeeded in forming dots of GaN on a quasi-lattice matched substrate and in controlling the dot size of the Nc-Si easily. In addition to these results, we also succeeded in improving the GaN crystal quality drastically by doping an In atom. This result suggests to us that the impurity of In inactivates the non-radiation centre for the luminescence. As discussed here, the role of small amounts of impurity for crystal growth is very important and very interesting for future nanostructure fabrication.

References

- [1] Tanaka S, Iwai S and Aoyagi Y 1996 *Appl. Phys. Lett.* **69** 4096

- [2] Tanaka S, Hirayama H, Nomura S, Iwai S and Aoyagi Y 1997 *Bull. Solid State Phys. Appl. (Japan. Soc. Appl. Phys.)* **12** 38
- [3] Tanaka S, Hirayama H, Aoyagi Y, Narukawa Y, Kawakami Y, Fujita Sz and Fujita Sg 1997 *Appl. Phys. Lett.* **71** 1299
- [4] Hirayama H, Tanaka S, Ramvall P and Aoyagi Y 1998 *Appl. Phys. Lett.* **72** 1736
- [5] Hirayama H, Tanaka S and Aoyagi Y 1999 *Microelectron. Eng.* to be published
- [6] Canaham L T 1990 *Appl. Phys. Lett.* **57** 1046
- [7] Zhao X, Schoenfeld O, Kusano J, Aoyagi Y and Sugano T 1994 *Japan. J. Appl. Phys. Lett.* **33** L649
- [8] Zhao X, Schoenfeld O, Aoyagi Y and Sugano T 1994 *Appl. Phys. Lett.* **65** 1290
- [9] Schoenfeld O, Zhao X, Hempel T, Blaesing J, Sugano T and Aoyagi Y 1994 *J. Cryst. Growth* **142** 268
- [10] Zhao X, Schoenfeld O, Kusano J, Aoyagi Y and Sugano T 1994 *Japan. J. Appl. Phys. Lett.* **33** L899
- [11] Nomura S, Zhao X, Schoenfeld O, Misawa K, Kobayashi T, Segawa Y, Aoyagi Y and Sugano T 1994 *Solid State Commun.* **92** 665
- [12] Schoenfeld O, Zhao X, Hempel T, Sugano T and Aoyagi Y 1994 *Japan. J. Appl. Phys.* **33** 6082
- [13] Zhao X, Schoenfeld O, Aoyagi Y and Sugano T 1994 *J. Phys. D: Appl. Phys.* **27** 1575
- [14] Zhao X, Schoenfeld O, Komuro S, Aoyagi Y and Sugano T 1994 *Phys. Rev. B* **50** 18654
- [15] Nomura S, Zhao X, Schoenfeld O, Misawa K, Kobayashi T, Segawa Y, Aoyagi Y and Sugano T 1995 *Japan. J. Appl. Phys.* **34** 125
- [16] Zhao X, Schoenfeld O, Nomura S, Komuro S, Aoyagi Y and Sugano T 1995 *Mater. Sci. Eng. B* **35** 467
- [17] Zhao X, Nomura S, Aoyagi Y and Sugano T 1996 *J. Non-cryst. Solids* **198/200** 847
- [18] Zhao X, Komuro S, Isshiki H, Maruyama S, Aoyagi Y and Sugano T 1997 *Appl. Surf. Sci.* **113/114** 121
- [19] Zhao X, Komuro S, Fujita S, Isshiki H, Aoyagi Y and Sugano T 1998 *Mater. Sci. Eng. B* **51** 154
- [20] Zhao X, Komuro S, Isshiki H, Aoyagi Y and Sugano T 1999 *Appl. Phys. Lett.* **74** 120
- [21] Komuro S, Katsumata T, Morikawa T, Zhao X, Isshiki H and Aoyagi Y 1999 *Appl. Phys. Lett.* **74** 377
- [22] Zhao X, Komuro S, Isshiki H, Aoyagi Y and Sugano T 1999 *J. Mater. Sci. Technol.* to be published

Chapter 5

Dislocation filtering techniques for MBE large mismatched heteroepitaxy

J M Zhou, Q Huang, H Chen and C S Peng

*Institute of Physics, Chinese Academy of Sciences, PO Box
603-(36), Beijing 100080, People's Republic of China*

5.1 Introduction

There exist few lattice matched semiconductor hetero systems, such as GaAs/AlGaAs in nature. Any epitaxy technique could not avoid producing lattice mismatched heterostructures. Lattice-mismatched growth has been widely investigated for a variety of material systems in the recent two decades, because of their attractive properties. Applications for lattice-mismatched films can be divided into two kinds: (i) utilization of the strained layer itself; (ii) using a common substrate as a template for the growth of relaxed layers. Typical of the first application is the GaAs/InGaAs/AlGaAs pseudomorphic high electron mobility transistor (PHEMT). In this case, a strained layer of InGaAs with low In-composition is induced as an electron channel for electron mobility enhancement, because InAs has a lower effective mass and much higher electron mobility, compared with GaAs. The layer thickness of InGaAs is kept below the critical thickness to avoid the formation of misfit dislocations which degrade the electrical and optical properties of carriers. Due to the limited thickness and strain required for coherently strained growth, the design options for electronics, optoelectronics and photonics applications are restricted.

When a substrate is used as a template for a highly mismatched epitaxy in the second application, the threading dislocation density in the epilayer can be very high (10^8 – 10^9 cm^{-2}) and techniques to reduce the threading dislocation density have been developed. However, there are still many problems to be solved for different hetero material systems. In this paper we will concentrate on the second application of large mismatched heteroepitaxy. In the first part of this paper,

several typical techniques for reducing threading dislocation densities will be reviewed. In the second part, dislocation filtering techniques of low-temperature growth of Si and SiGe and prelayer growth for cubic GaN on GaAs and InAs on GaAs, developed in our laboratory will be introduced.

5.2 Survey of several dislocation filtering methods

Within the range of elemental, compound and alloy semiconductors, few indeed are available as substrates, on which thick layers of these materials can be grown without dislocation. In most cases, heteroepitaxy faces large lattice mismatch problems, that restrict the selection and combination of the heteromaterials. In order to expand a variety of available semiconductor heteromaterials great efforts have been made to circumvent this restriction. All the reported methods were used to grow layers to block dislocations, generated during mismatched heteroepitaxy.

The majority of semiconductor heteroepitaxy takes place on the (001) surface. Most group IV, III–V and II–VI semiconductors have the same or similar fcc crystal structure. In heteroepitaxy, the crystal structure is replicated by the overlying crystal, where any defect intersects the original surface of the layer. The most common misfit dislocations in (001) semiconductor layers can be divided into two types. The first type of 60° dislocation with Burgers vector of $(a/2)\langle 011 \rangle$, lies at 60° to $\langle 110 \rangle$ line direction and propagates easily with threading segment lying on a $\{111\}$ glide plane. The second type is the pure edge $(a/2)\langle 110 \rangle$ dislocation, with Burgers vector lying in the interface at a right angle to the line direction.

No matter what method is used, the overlayer (so-called buffer) should fulfil to the three primary requirements: strain relief, flatness and low threading dislocation density on the surface. In mismatched heteroepitaxy, dislocations provide a means of relieving the internal stress which develops at the interface between the buffer and substrate. The efficacy of any method is able to block threading dislocation or suppress generation of threading dislocation instead of misfit dislocation at the interface. Among them the common methods used in the different material systems are constant composition layer, strained-layer superlattice (SLS) filter, compositional graded (linear or step), patterned substrate and post thermal or optical annealing. These methods have been used in certain cases with some restrictions.

5.2.1 Constant-composition layer

The simple method is to grow a constant composition layer on the substrate. According to Mathews' energy balance arguments [1] for defining the critical thickness for dislocation formation, above h_c , creation of misfit dislocations is energetically favourable and more dislocations will be introduced until the strain drops to the level where it is no longer energetically favourable for new dislocation creation. During the growth, islands easily form and 60° dislocations

at the edge of islands form an irregular array. This array, through dislocation interaction, will change the self-stress of the misfit dislocation segment and the glide force on the threading and block glissible threading dislocation. In any case, the annihilation of the dislocation will occur more rapidly when the density of threading dislocations is high and interactions are strong and gradually reduce as more dislocations are annihilated. Therefore, the constant composition layer with low dislocation density in most of the material systems should grow extremely thick (tens of micrometres) for large mismatched heteroepitaxy, and the density of threading dislocation could not reduce down to 10^7 cm^{-2} [2–6]. Though annealing high misfit layer during or after growth could lead to a notable reduction of density of threading dislocation, the constant-composition buffer layers are unlikely to be widely used.

5.2.2 Strained-layer superlattices (SLS)

There has been a large number of studies of the effect of SLS on the reduction of the threading dislocation density. SLS generally consist of alternating thin layers (from a few monolayers to several tens of monolayers) of different material, producing a structure with certain periodicity which is large in comparison with the atomic lattice, hence referred to as the term ‘superlattice’.

It has been proposed by several workers [7–9] that a difference in the elastic constants of materials can also reduce the threading dislocation density. The basis of this assertion is that a dislocation has a lower self-energy in a soft material than in a hard material; a dislocation would thus prefer to stay in a soft material rather than propagate into a hard material. The SLS can bend threading dislocations away from growth direction. The length of the bend segment of a dislocation increases with the strain level in the SLS. This is because the force acting on the threading dislocation from the coherent stress is proportion to both the strain and the layer thickness. Then SLSs with high misfit and thickness should be selected to sustain the maximum gliding force.

EI-Masny and Tarh [11] used GaAsP/InGaAs strained layer SLS as filters for growing GaAs on Si. They found that: (1) Almost all the threading dislocations are blocked and bend along the SLS interfacial places when the dislocation density is comparatively low. (2) Bend dislocations at the SLS interfaces interact with one another. The dislocation interactions are dependent on the type of dislocation and their strain fields. (3) The SLS layers do not have infinite capacity in bending the threading dislocations. The strain in the SLS layers is reduced if a high dislocation density impinges on these layers. (4) The bend dislocations in the low density regions can propagate along the SLS interface hundred of micrometres without disturbance.

In the case of high-defect density, the strain field resulting from a nearby dislocation can reduce the strain f in the SLS to $(f - S)$, where S is the dislocation induced strain relaxation. Thus, there might be a critical separation distance S_{crit} between dislocations below which the SLS will relax and will not be as effective

in bending the dislocations.

In practical applications, SLS filters consisting of different elastic constants between two materials for certain heteroepitaxy are difficult to find. Meanwhile the two materials could neither diffuse nor segregate to form a sharp interface.

In general this method would not be used in large lattice mismatched systems and its use is limited.

5.2.3 Composition graded layer

Compositional graded layers in linear or stepwise manner have been widely used in a variety of material systems for almost 30 years [11, 12]. This method can suppress islanding at the initial stage of heteroepitaxy. As the layer thickness increases linearly, misfit relief begins, reducing the strain at the top surface of the layer. Because the growth mode is planar, the dislocations introduced are the long glissile 60° type characteristic of low-misfit layers rather than the short edge dislocation encountered in island growth. Also, since roughening of the surface is at least partly due to strain, this misfit relief delays the onset of roughening and island growth, allowing layers with a high surface misfit but low strain to be grown without generating large numbers of threading dislocations. As the layer becomes even thicker, still more dislocations are introduced, relieving the strain in the lower part of the layer and keeping the strain low at the surface.

Consider a compositional step-graded structure, in which each layer has a misfit strain of $\Delta\varepsilon$ with respect to the layer immediately below it and a thickness Δz . As pointed out by Abrahams *et al* [14], there are two different types of behaviour. First, the misfit dislocation array in each interface may generate new threading dislocations. The threading dislocation density will thus rise in direct proportion to the number of layers grown. Second, the threading dislocations from one interface may act as misfit dislocation sources for the next.

The reduction process of threading dislocations can be described as follows. Each interface mismatch threading dislocation is absorbed by the next interface, then it is re-emitted to produce an array in the next interface, and so on. In comparison with constant composition layers, calculation of the equilibrium distribution of dislocations shows that threading dislocations are subject to greater force and weaker pinning in compositional graded layer, helping them to be swept to the edge of the sample. LeGoues [15] also found anomalous strain relaxation in SiGe thin film and superlattice where relaxation is characterized by the presence of dislocations in the Si substrate, as well as in the lower part of the film. This phenomenon is due to the paucity of nucleation sites and controlled by the specific Ge concentration profile in the linear or stepwise graded layer.

The method, described above, has difficulty keeping 2D growth mode during the compositional graded process, therefore its surface is rather rough up to 150–200 Å. The active layers grown on such buffer do not match with device quality requirements. In order to reduce the threading dislocation density down to 10^6 cm^{-3} , the graded layer would be grown rather thicker or the composition

should not be high. The thick buffer is not suitable for device technology, such as cleavage, wet and dry etching. The compositional graded method can be used for material and substrate, between which there is no intermiscibility, otherwise the substrate become dopant for epilayer.

5.2.4 Patterned substrate

In ideal situations misfit dislocation is nucleated and propagates to the interface and terminates at the edge of the wafer. The probability for dislocation interaction and multiplication is directly dependent on the substrate size and lattice mismatch and only when very small lattice mismatches and small substrate dimensions are used will the dislocation propagate to the growth edge. Therefore, the use of reduced growth areas is effective for reducing threading dislocation density when lattice mismatch is not large.

A more effective method, which can apply to large mismatches and thick epilayers is to combine reduced growth area with composition grading and strained layer superlattices, reported by Beam [16]. In principle, formation of misfit dislocations is from the grid of threading dislocation under the influence of the mismatch stress of the heterointerface. However, when the substrate dislocation density is low and/or the lattice mismatch is large, the number of threading dislocations will not be sufficient to generate the observed densities of misfit dislocations. Reduced growth area provides activity of the edges as heterogeneous nucleation sites, which is favourable for formation of misfit dislocation and prevents dislocation interaction during the misfit accommodation process.

Because the dislocation nucleation rate is dependent on the growth rate and composition, grading could result in larger growth areas free of dislocation interaction since the nucleation rate can be reduced by grading the lattice misfit with growth thickness/time. The experiments show that this hybrid technique can result in a dramatic reduction in threading defect densities with reduced area growth size about 20–30 μm for $\text{In}_x\text{Ga}_{1-x}\text{As}$ with compositions up to $x = 0.53$ on GaAs (001) [15].

5.3 Low-temperature Si and/or GeSi buffer for GeSi alloy strain relaxation

Some of the Ge/Si heterostructures require relaxed GeSi buffers to serve as virtual substrates. We use a new method to avoid some disadvantages of the composition grading technique.

Perovic *et al* [17] discovered that there was a large amount of spherical micro voids in LT-Si, because at low temperatures, the migration rate of the deposited atoms in the surface is not enough to form a sharp surface and micro-holes are induced. These micro-holes are not perfectly covered by the deposited atoms and

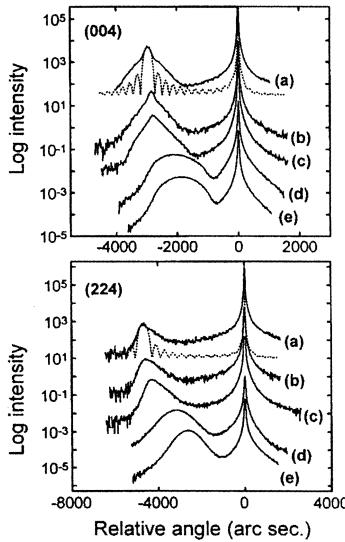


Figure 5.1. X-ray (004) and (224) rocking curves of the five samples with the $\text{Ge}_{0.3}\text{Si}_{0.7}$ epilayer thickness of 80, 100, 150, 300 and 500 nm, respectively. Curves (a)–(e) are for Ma–Me, respectively. The dotted curves in both panels are theoretical simulations of curve (a).

vacant defects are formed and frozen in the epilayer. During the epilayer growth, the surface roughness rises and the number of micro-holes increases. Therefore, the vacant defect density increases toward the surface, which are very easy to aggregate and form stacking faults.

The two groups of samples were grown to study the mechanism of the strain relaxation for low and high Ge composition of GeSi alloy on LT-Si buffer, respectively. Their structures are as follows:

Group 1. Si (001) substrate +50 nm LT-Si + X nm $\text{Ge}_{0.3}\text{Si}_{0.7}$ and $X = 80, 100, 150, 300, 500$ nm corresponding to samples from a1 to a5.

Group 2. Sample b1 is 50 nm LT-Si + 500 nm $\text{Ge}_{0.3}\text{Si}_{0.7}$ on Si (001); sample b2 is 50 nm LT- $\text{Ge}_{0.3}\text{Si}_{0.7}$ + 500 nm $\text{Ge}_{0.6}\text{Si}_{0.4}$ on sample b1; sample b3 is 50 nm LT- $\text{Ge}_{0.6}\text{Si}_{0.4}$ + 500 nm $\text{Ge}_{0.9}\text{Si}_{0.1}$ on sample b2.

Therefore, we obtained high Ge composition SiGe alloy up to $X = 0.9$.

In order to study the mechanism of dislocation induction and strain relaxation, high resolution x-ray diffraction (XRD) and transmission electron microscopy (TEM) were used. The XRD (004) and (224) rocking curves of first group samples (curves (a) to (e) for samples a1 to a5, respectively) are shown in figure 5.1 Results for the strain relaxation were summarized in figure 5.2 in terms

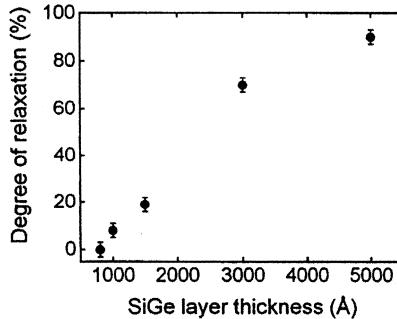


Figure 5.2. Degree of relaxation of the $\text{Ge}_{0.3}\text{Si}_{0.7}$ epilayers versus layer thickness.

of degree of relaxation R which has been plotted as a function of the $\text{Ge}_{0.3}\text{Si}_{0.7}$ layer thickness. R is defined as

$$R = (a_{\parallel} - a_s)/(a_r - a_s).$$

Where a_{\parallel} is the in-plane lattice constant of the epilayer; a_s and a_r are the unconstrained lattice constants of the substrate and the epilayer, respectively. These results show that the relaxation is observable for $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer thickness beyond 80 nm. This thickness is much less than that (150–200 nm) of $\text{Ge}_{0.3}\text{Si}_{0.7}$ grown in the same condition without the LT-Si layer. For the 80 nm thick layer, although there is no observable macroscopic relaxation, evidence of the *onset* of relaxation is observed, as will be discussed later. Further, the sample with a 500 nm thick $\text{Ge}_{0.3}\text{Si}_{0.7}$ layer is already close to full relaxation ($R = 90\%$). Such a thickness is much thinner than that of compositionally graded buffer systems having the same degree of relaxation.

In figure 5.1, theoretically simulated (004) and (224) rocking curves for sample a1 have also been plotted (dotted curves). An interesting feature here is that in curve (a) the main part of the peak of the $\text{Si}_{0.7}\text{Ge}_{0.3}$ epilayers is symmetric and fits quite well to the simulation (note the logarithmic scale), suggesting that a major part of the alloy layer remains pseudomorphic. Whereas, the tail of the epilayer peak becomes *broad and asymmetric* and shifts toward the higher angle side. Therefore, it is reasonable to assume that a small portion of the alloy epilayer is already (partly) relaxed. The implication is that the relaxation is quite inhomogeneous. With the increase of the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer thickness (curves (b) and (c)), the asymmetric part of the layer peak expands, while the symmetry part of the peak shrinks. A further increase of the epilayer thickness makes the whole peak become broad and finally gets more symmetric (curves (d) and (e)), suggesting that the relaxation of the thickest alloy layer turns out to be more homogeneous.

Such a result of relaxation is further confirmed by TEM measurements. Figures 5.3(a) and (b) show the high resolution cross section TEM images of

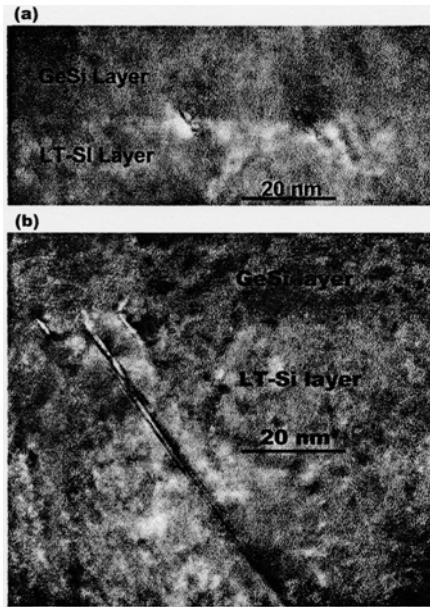


Figure 5.3. The high-resolution cross-section TEM images for samples Ma (a) and Mb (b). There are stacking faults in {111} in the LT-Si layer near the GeSi/LT-Si interface. The number increases and they are grown deeply into the LT-Si layer during the growth.

samples a1 and a2, respectively. When the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer thickness reaches 80 nm, stacking faults are induced in the LT-Si layer near the interface of LT-Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$, then grow into the LT-Si layer. In the LT-Si layer, the density of vacancy near the interface of LT-Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ is the largest [17]. Under the tensile stress and the osmotic force, vacancies near the interface aggregate together first to form stacking faults to relax the strain. At the beginning, the scale of stacking faults is small, and a large part of the interface remains strained. Therefore, the relaxation is inhomogeneous. With further growth, the scale of stacking faults becomes larger and the distribution of them becomes more uniform. At last, as the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer relaxes fully, the relaxation turns out to be homogeneous. Figure 5.4 shows the TEM cross section image of the almost fully relaxed sample a5. The mismatch dislocation distribution is uniform and periodic.

Due to the very large supersaturating of vacancies in LT-Si, the vacancy osmotic force becomes very large. The vacant defects are very easy to aggregate together to form stacking faults only needing very small external force (such as $\text{Si}_{0.7}\text{Ge}_{0.3}/\text{Si}$ mismatch force). Therefore, the critical thickness of strained GeSi alloy grown on LT-Si is much smaller than that on a normal Si substrate.

Linder *et al* [18] reported that there are large numbers of dislocations in the $\text{Ge}_x\text{Si}_{1-x}$ alloy layer grown on LT-Si when the Ge fraction x reaches 40%. It

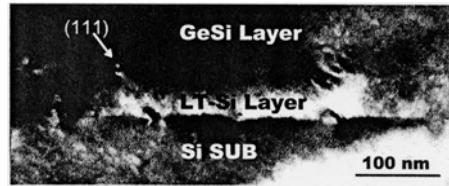


Figure 5.4. The high-resolution cross-section TEM image for sample Me. There are pairs of dislocations in the intersection of {111} and two GeSi/LT-Si/Si interfaces. These pairs arrange periodically along these two interfaces.

Table 5.1. Residual strain, Ge fraction and FWHM of diffraction peak by HRXTD.

	b2			b3		
b1	$x = 0.3$	$x = 0.3$	$x = 0.6$	$x = 0.3$	$x = 0.6$	$x = 0.9$
FWHM (s)	145	193	110	220	176	93
Residual strain (%)	10	8	11	5	9	13
Ge fraction	0.30	0.29	0.59	0.30	0.61	0.89

suggests that there are no more vacancies in the LT-Si layer to form stacking faults when x becomes larger than 30% and the LT-Si layer loses efficacy. To deal with this problem, a stepped-up LT technique was used to relax $\text{Ge}_{0.3}\text{Si}_{0.7}$ on LT-Si, relax $\text{Ge}_{0.6}\text{Si}_{0.4}$ on LT- $\text{Ge}_{0.3}\text{Si}_{0.7}$, and then $\text{Ge}_{0.9}\text{Si}_{0.1}$ on LT- $\text{Ge}_{0.6}\text{Si}_{0.4}$.

Three samples were grown as b1, b2, b3. Bright field cross section (200) TEM images together with layer structure are shown in figure 5.3. It is very clear that the dislocations locate mainly in the LT layers and the layers below them. The density of TDs in the top layer is much lower. The presence of the LT layers has apparently changed the misfit dislocation nucleation mechanism and/or the rate of the dislocation nucleation as proposed above. This mechanism occurs not only in the single step LT-Si technique but also in the multiple steps one. To count the TD densities on the top GeSi layers, plan-view TEM images of the same samples were performed. The results are shown in table 5.1. They are much lower than that in the GeSi layer grown directly on Si substrate.

The surface roughness was measured by AFM. The root mean square (RMS) amplitudes of the cross hatch are only 1.77, 2.34 and 5.78 nm for samples b1, b2 and b3, respectively (see table 5.2). The roughness is much smaller than that obtained from compositionally graded buffers [14, 19]. The density of surface pits [20] measured by AFM is in good agreement with the TD density of the top layer measured by TEM (see table 5.1).

Figure 5.4 shows the high-resolution XTD (004) rocking curves. The open circles stand for the experimental data and full curves for the Gaussian fitting for

Table 5.2. Surface dislocation density and roughness measured by TEM and AFM.

	b1	b2	b3
TD density by TEM (cm^{-2})	$<10^5$	3.8×10^5	3.2×10^6
Pits density by AFM (cm^{-2})	$<2.5 \times 10^5$	$\sim 5.0 \times 10^5$	$\sim 2.5 \times 10^6$
RMS by AFM (\AA)	17.7	23.4	57.8

each peak. The (224) diffraction was also performed in order to determine the residual strain and the alloy composition. Results were summarized in table 5.2, from which all GeSi epilayers are almost fully relaxed and the compositions were successfully controlled at about 0.3, 0.6 and 0.9. The FWHMs of the XTD (004) diffraction are shown in table 5.2. It is clear that the crystal quality of the top layer is much better than the layers below. These results agree well with the TEM analysis.

Therefore, by using multiple steps of the low-temperature growth technique, fully relaxed $\text{Ge}_x\text{Si}_{1-x}$ epilayers with Ge fraction x up to 90% were grown on Si(001) by MBE as a result of low threading dislocation density, smooth surface, and good crystal quality. And, the total layer thickness is relatively small.

5.4 New initial growth method for pure cubic GaN on GaAs (001)

Although successful epitaxial growth of the metastable cubic phase of GaN (c-GaN) has been reported by various groups [21, 22], there still exist three major difficulties in the growth of c-GaN on GaAs substrates. The first originates from the polytypism, as a result, it increases the tendency to form wurtzite-phase subdomains within the cubic lattice. The phase purity of molecular beam epitaxy (MBE) growth c-GaN on GaAs substrates has been the subject of several previous studies [23–25]. The second is a large lattice mismatch between GaAs and GaN which results in a high density of dislocation and stacking faults in the epilayer [26, 27]. The third is the etching GaAs by N^+ -ion, which results in surface roughness, wurtzite-phase subdomain and stacking faults. The common method of c-GaN on GaAs is to deposit a rather thick GaN buffer on GaAs at low substrate temperature and thermal annealing subsequently. It is generally assumed that the buffer layers are amorphous or polycrystalline layers, which crystallize during the annealing. In this section, a new prelayer technique by which a pure cubic GaN on a GaAs substrate with high quality is reported.

The cubic GaN films were grown on semi-insulating GaAs (001) substrates in the MBE system with a radio-frequency (RF) plasma discharge source [28]. The key point for obtaining a pure c-GaN phase is to avoid the appearance of (111) facets during the whole growth process. Prior to each nitridation process, a

thick GaAs buffer of 200 nm was grown at 600 °C under (2 × 4) reconstruction to ensure As-stable condition. Therefore, when nitrogen species impact on the GaAs surface for nitridation of GaAs, the As shutter should be opened to prevent Ga-rich conditions occurring and forming (111) facets. Meanwhile surface reconstruction changes from (2 × 2) to (3 × 3), then the Ga shutter is opened and the As shutter is closed in series. The substrate temperature and initial layer thickness have a great influence on the quality of the cubic GaN layer. As previously reported, if a GaN prelayer is deposited at 400 °C and 500 °C, a spotty RHEED pattern could always be observed due to poor surface migration. Therefore, higher growth temperatures of 600 °C and 700 °C have been chosen. However, 700 °C is too high for nitridation, because etching would occur at such a temperature. This kind of rough surface of prelayer is not suitable for pure cubic GaN growth of high quality, because the mixed phase of GaN would become dominant during the growth. It is found that only at a growth temperature of about 600 °C, can a smooth surface of the prelayer be obtained.

The importance of the prelayer thickness can be understood by following experimental results. The prelayer of two samples were grown at the same growth condition, including substrate temperature, As pressure, growth rate of GaN, ratio of GaN, but with different thickness, i.e. two (sample A) and 20 (sample B) monolayers. Then an active GaN layer of 500 nm was grown on the two prelayers at 700 °C, respectively.

There are significant differences between the two samples in RHEED pattern, x-ray analysis and photoluminescence measurements.

For sample A, the streaky RHEED pattern maintained during growth, the surface reconstruction changed from (1 × 1) to (2 × 2) after only about 4 nm thick deposition of GaN at 700 °C. However, for sample B, a spotty RHEED pattern was observed until about 50 nm thick deposition of GaN at 700 °C, then a (2 × 2) reconstruction occurred after about 70 nm GaN deposition.

X-ray double crystal diffraction was used to analyse the crystal quality of GaN layers. [Figure 5.5](#) shows the x-ray diffraction profile of samples A and B, respectively. In addition to the diffraction peak of GaAs (002), there was only one strong peak at $2\theta = 40.4^\circ$ for both of the samples, which was assigned to the (002) reflection peak from cubic GaN. The full width at half maximum (FWHM) is 10 min for sample A, and 23 min for sample B. The linewidth from the two samples is significantly smaller than that of other reports. For comparison, some published data are listed as below: 27 min (1 μm thick GaN/GaAs [29] for cubic GaN grown by MOVPE), 96 min (400 nm thick c-GaN/GaAs [30]), 60 min (4 μm thick c-GaN/Si [31]), 60 min (4 μm thick c-GaN/Si [32]), and 28 min (1 μm thick c-GaN/Si [33]) for cubic GaN grown by MBE.

The photoluminescence spectra at 12 K for samples A and B are shown in [figure 5.6](#). There is only strong excitonic emission at 3.274 eV from cubic GaN of sample A. However, for sample B there exist three peaks, one at 3.274 eV from excitonic emission, another two peaks at 3.20 eV and 3.14 eV from DA pairs. There was no deep level emission around 2 eV the so-called yellow band from

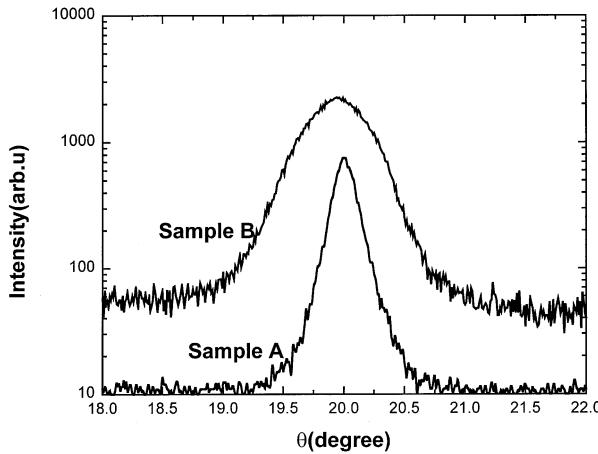


Figure 5.5. X-ray rocking curves of (002) peaks of cubic GaN films for samples A and B, respectively.

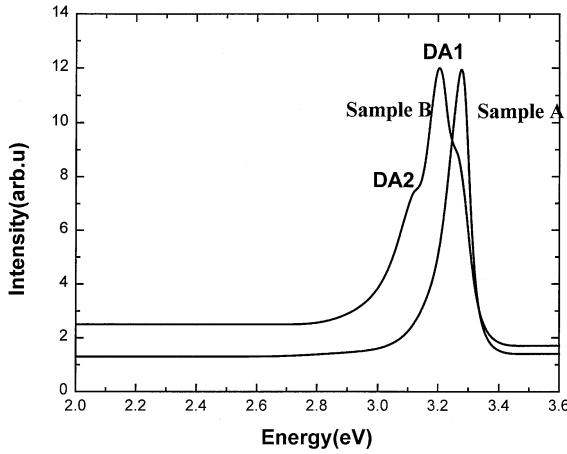


Figure 5.6. 12 K photoluminescence of samples A and B, respectively.

both samples, which indicates a high quality of these samples.

As pointed out by several authors [34, 35], a high density of dislocations and stacking faults will form at the edge of the islands and some in the islands at the initial growth of GaN, and this kind of defect could propagate into the epilayer of GaN during growth. The quality difference between samples A and B is due to the difference in surface topography of the two prelayers. According to the streaky and spotty RHEED pattern during prelayer growth of samples A and B, it may be deduced that layer-by-layer and islands growth mode are dominant at the

initial growth for sample A and sample B, respectively. So the defect in sample A is much lower than sample B and also the FWHM of x-ray rocking curves for sample A is smaller than that of sample B.

Under the same growth condition, Ga and N vacancies and incorporated oxygen and carbon should be similar in undoped samples A and B, and should not contribute to the DA pairs band. Therefore, the significant difference in PL spectra between samples A and B would be related to the difference in the densities of stacking faults and dislocations, which implies that the DA pairs band is associated with these kinds of defects. To our knowledge, this is the first report where there is no peak from DA pairs for the cubic GaN epilayer on GaAs, which is important to light emission efficiency.

The above experiments show clearly that growing heterostructures on layer lattice mismatched substrate does not need very thick ‘buffer’ to obtain a relaxed layer.

5.5 InAs growth on GaAs using new prelayer technology

A variety of specific techniques, including linearly graded or step graded procedures and direct growth, have been used for this purpose. For the first two procedures, the gliding of 0° misfit dislocations is the dominant strain relaxation mechanism. However, there are two major disadvantages. First, these techniques produce a rough cross-hatched surface that is not suitable for devices with submicrometre featurea. Second, these techniques are generally not suitable for material systems with lattice mismatch greater than 3% as the buffer layer thickness also increases proportionally [36, 37].

Direct growth has also been investigated for systems with various amounts of lattice mismatch. It is found that the coexistence of both 60° and pure edge-type (90°) misfit dislocation is observed for the system with moderated lattice mismatch [38, 39]. The resulting epilayers usually have high densities of bulk defects larger than 10^{12} cm^{-2} . It is recognized that these bulk defects, including stacking faults and threading dislocations, originate from sources such as 60° misfit dislocations, while 90° dislocations can relieve strain most effectively at the interface without any threading component. 60° misfit dislocations are highly active sources for the generation of threading dislocations. The surface of InAs grown directly on GaAs is too rough for quantum structure and photolithograph because islands will appear. When 1.7 MLs thick InAs deposited on a GaAs substrate, 60° dislocations and stacking faults will form at the edge of islands [37, 38]. Recently a different case is reported on initial InAs grown on GaP under In-rich conditions, where only 90° misfit dislocations form at the interface [40].

In this section, a strain relaxation InAs layer grown on GaAs under In-rich conditions and subsequent growth of InAs using a relaxed layer such as a prelayer is presented.

The strain relaxed InAs layer grown on a GaAs (001) substrate is investigated

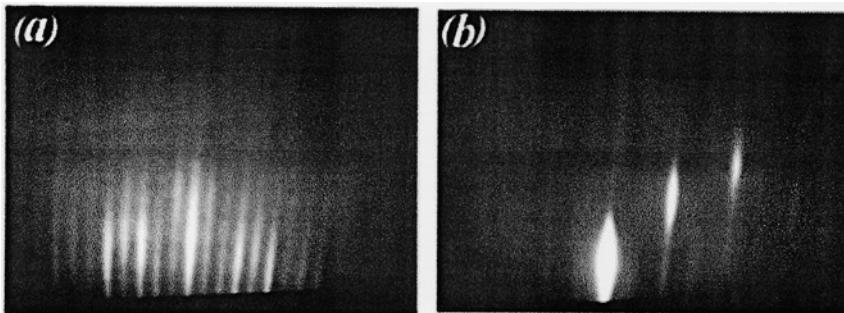


Figure 5.7. The RHEED patterns for InAs grown on (100) GaAs under the In-rich conditions at 380 °C. (a) and (b) correspond to the patterns taken along [110] and [1̄10], respectively.

by x-ray double crystal diffraction. 200 nm GaAs is grown on GaAs (001) substrates at a temperature of 580 °C followed by InAs layers 7 nm, 10 nm and 20 nm thick grown under In-rich conditions at a temperature of 380 °C, respectively. Two-dimensional growth mode maintains throughout the entire growth process of InAs with (4 × 2) reconstruction observed by RHEED, which are streaky as shown in figures 5.7(a) and (b) with the beam incidence along the [110] and [1̄10] directions, respectively. X-ray rocking curves of the (004) peak are shown in figure 5.8 for 7 nm, 10 nm and 20 nm thick InAs. In addition to the diffraction peak of GaAs (004), there are different peaks from 7 nm, 10 nm, 20 nm thick InAs at $\theta = 30.457, 30.450, 30.455$, respectively. The angle difference between the diffraction peaks of GaAs (004) and InAs can be measured precisely to estimate the lattice constants along the [001] direction with 0.6091 nm, 0.6093 nm, 0.6075 nm, for 7 nm, 10 nm, and 20 nm thick InAs, respectively. The same position of peaks from 7 nm, 10 nm and 20 nm thick InAs indicates full strain relaxation in less than a 7 nm thick InAs epilayer. However, in general, full strain relaxation is thicker than the 300 nm thick InAs epilayer grown on GaAs under As-rich conditions [37, 38]. The x-ray diffraction analysis shows that the strain relaxation of the InAs layer grown under In-rich conditions is much faster than that under As-rich conditions. Therefore, such a InAs layer is appropriate for a prelayer to grow high quality InAs.

As is well known, the islands will appear when 1.7 MLs thick InAs is deposited on a GaAs substrate under As-rich conditions, and the stress will be first relaxed through imhomogeneous lattice distortion on the surface of islands and then through misfit dislocation at the interface of InAs/GaAs. As the growth continues, the islands become larger and larger, the bumpy surface tends to smooth. The full relaxation will occur only through the misfit dislocation until the islands disappear, but it is a difficult process to form misfit dislocations to relax stress instead of the lattice distortion on the surface of islands. Under In-rich

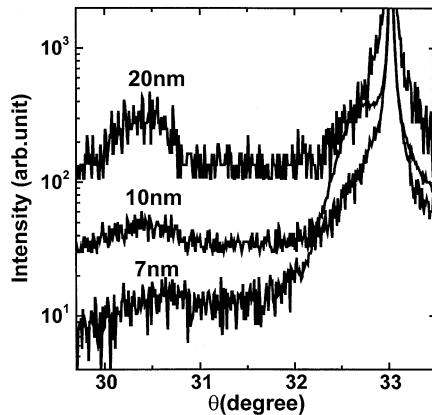


Figure 5.8. X-ray rocking curves of the (004) peak of 4 nm, 7 nm, 20 nm and 10 nm thick InAs.

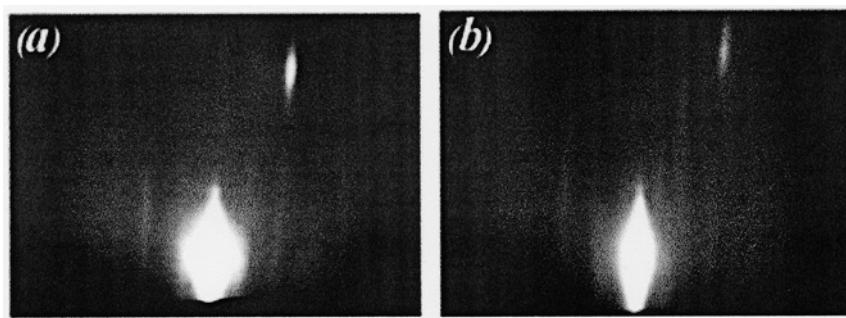


Figure 5.9. The RHEED patterns of (a) and (b) for InAs grown on (100) GaAs under As-rich conditions.

conditions, however, the two-dimensional growth mode maintains throughout the entire growth process, so the stress will be relaxed only through misfit dislocation at the interface without inhomogeneous lattice distortion. This is the reason that the strain relaxation process of an InAs layer grown under In-rich conditions is much faster than that under As-rich conditions.

For comparison, two samples were grown on a 20 nm thick InAs prelayer under As- and In-rich conditions, respectively. For InAs layer growth under As-rich conditions, after InAs prelayer growth, appropriate As was impinged onto the surface of 20 nm thick InAs, then the In-rich (4×2) reconstruction was changed to the As-rich (2×4) one in several seconds, and the streaky RHEED patterns shown in figures 5.9(a) and (b) indicates a two-dimensional growth mode and full strain relaxation. 500 nm and 300 nm thick InAs layers, referred to as samples A and

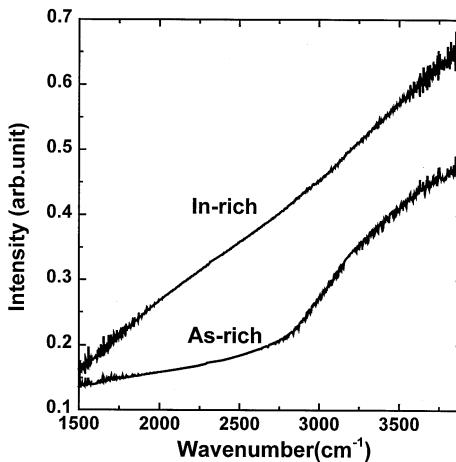


Figure 5.10. The optical absorption spectra for samples A and B.

B, were grown at the temperature under As and In conditions, respectively. Two-dimensional growth mode maintains throughout the entire growth process for both of the samples. The optical absorption spectra for samples A and B is shown in figure 5.10. The InAs grown under As-rich conditions shows an absorption onset at 0.3472 eV \sim 3.57 μ m, which is attributed to the band edge absorption of the InAs layer. However, InAs grown under In-rich conditions shows no absorption of band edge. In this case indium atoms aggregate and form the clusters on the surface of the InAs layer, and indium anti-sites defects would exist in such a InAs layer. Therefore, this layer is not suitable for device application.

5.6 Conclusion

Although heteroepitaxy techniques for large lattice mismatch have been developed greatly, the epilayer with device quality for different heteromaterial systems have not yet been achieved. There exist two kinds of problems to be solved.

First, the origin of misfit dislocation is not clear. In most cases, people consider that the GaAs substrate is responsible for misfit dislocations of heteroepitaxy on GaAs with large lattice mismatch. However, the same statement cannot explain misfit dislocations occurring at the interface of the heteroepilayer on the silicon substrate, where the density of dislocations is very low. Another way to explain the origin is the half-loop of dislocation in the heteroepilayer, that is not observed experimentally in GeSi systems. In addition, we could not understand why misfit dislocations exist in heteroepilayers with small lattice mismatch in their origin.

The second problem is that dislocation movement and interaction is not fully understood. How do we control their movement? So far, we cannot explain why some heteroepilayers are good by some dislocation control techniques, and whether such techniques could be extended to apply to other lattice mismatch material systems. We should also notice that dislocation control is unrepeatable, the reasons could be complicated. Control dislocation generation, movement and interaction and reduction density of dislocations for heteroepitaxy are open questions.

References

- [1] Matthews J W, Mader S and Light T B 1970 *J. Appl. Phys.* **41** 3700
- [2] Matthews J W and Blakesless A E 1974 *J. Cryst. Growth* **27** 118
- [3] Huang Y Y P Y, Charasses M N, Lo Y and Wang S 1987 *Appl. Phys. Lett.* **51** 192
- [4] Tachikawa M and Yamaguchi M 1990 *Appl. Phys. Lett.* **56** 484
- [5] Ayers J E, Chandi S K and Schowalter L J 1991 *J. Cryst. Growth* **113** 430
- [6] Schaffer W J, Lind M D, Kowalezyk S P and Grant R W 1983 *J. Vac. Sci. Technol. B* **1** 688
- [7] Hull R, Beau J C, Lei Benguch R E and Werder D J 1989 *J. Appl. Phys.* **65** 4723
- [8] Matthews J W and Blakesless A E 1975 *J. Cryst. Growth* **29** 273
- [9] Gourley P L, Bieffeld R M and Dawson L R 1985 *Appl. Phys. Lett.* **47** 482
- [10] Rao T S, Nonawa K and Horikoshi Y 1993 *Appl. Phys. Lett.* **62** 154
- [11] EI-Masry N A and Tarh J C L 1989 *Appl. Phys. Lett.* **55** 1442
- [12] Tietjen J J and Amick J A 1966 *J. Electro-Chem. Soc.* **113** 724
- [13] Grenning D A and Herzog A H 1968 *J. Appl. Phys. Lett.* **52** 377
- [14] Abrahams M S, Weisberg L R, Buiochi C J and Blauc L 1969 *J. Mater. Sci.* **4** 223
- [15] LeGoues F K, Meyerson B S and Moras J F 1991 *Phys. Rev. Lett.* **66** 2903
- [16] Beam E A III and Kao Y C 1991 *J. Appl. Phys.* **69** 4253
- [17] Perovic D D, Weatherly G C, Noel J P and Houghton D C 1991 *J. Vac. Sci. Technol. B* **9** 2034
- [18] Linder K K, Zhang F C, Riech J S, Bhattacharya P and Houghton D 1997 *Appl. Phys. Lett.* **70** 3224
- [19] Kvan E P and Namavar F 1991 *Appl. Phys. Lett.* **58** 2357
- [20] Fitzgerald E A, Xie Y H, Green M L, Brasen D, Kortan A R, Michel J, Mii Y J, Weir B E 1991 *Appl. Phys. Lett.* **59** 811
- [21] Hsu J W P, Fitzgerald E A, Xie Y H, Silverman P J and Cardillo M J 1992 *Appl. Phys. Lett.* **61** 1293
- [22] Strite S, Ruan J, Li Z, Salvador A, Chen H, Smith D J, Choyke W J and Morkoc H 1991 *J. Vac. Sci. Technol. B* **9** 1924
- [23] Cheng T S, Jenkins L C, Hooper S E, Foxon C T, Orton J W and Lacklison D E 1995 *Appl. Phys. Lett.* **66** 1509
- [24] Lei T, Ludwig K and Moustakas T D 1993 *J. Appl. Phys.* **74** 4430
- [25] Siegle H, Echey L, Hoffman A, Thomsen C, Meyer B K, Schikora D, Hankeln M and Lischka K 1995 *Solid State Commun.* **96** 943
- [26] Brandt O, Yang H, Jenichen B, Suzuki Y, Däweritz L and Ploog K H 1995 *Phys. Rev. B* **52** R2253

- [27] Trampert A, Brandt O, Yang H and Ploog K H 1997 *Appl. Phys. Lett.* **70** 583
- [28] Ruvimov S, Liliental-Weber Z, Washburn J, Drummond T J, Hafich M and Lee S R 1997 *Appl. Phys. Lett.* **71** 2931
- [29] Nakadaira A and Tanaka H 1997 *Appl. Phys. Lett.* **71** 812
- [30] Strite S, Ruan J, Li Z, Salvador A, Chen H, Smith D J, Choyke W J and Morkoc H 1991 *J. Vac. Sci. Technol. B* **9** 1924
- [31] Lei T, Moustakas T D, Graham R J, He Y and Berkowitz S J 1992 *J. Appl. Phys.* **71** 4933
- [32] Liu H, Frenkel A C, Kim J G and Park R M 1993 *J. Appl. Phys.* **74** 6124
- [33] Powell R C, Lee N E, Kim Y W and Greene J E 1993 *J. Appl. Phys.* **73** 189
- [34] Okumura H, Ohta K, Feuillet G, Bala-Krishnan K, Chichibu S, Hamaguchi H, Hacke P and Yoshida S 1997 *J. Cryst. Growth* **178** B113
- [35] Kuwano N, Nagatomo Y, Kobayashi K, Oki K, Miyoshi S, Yaguchi H, Onabe K and Shiraki Y 1994 *Japan. J. Appl. Phys.* **33** 18
- [36] Chin T P, Hou H Q, Tu C W, Chang J C P and Otsuka N 1994 *Appl. Phys. Lett.* **64** 2001
- [37] Chang J C P, Chen J, Fernandez J M, Wieder H H and Kavanagh K L 1992 *Appl. Phys. Lett.* **60** 1129
- [38] Otsuka N, Choi C, Kolodziejski L A, Gunshor R L, Fischer R, Peng C K, Morkoc H, Nakamura Y and Nagakura S 1986 *J. Vac. Sci. Technol. B* **4** 896
- [39] Choi C 1988 *PhD Dissertation* Purdue University
- [40] Cheng J C P, Chin T P and Woodall J M 1996 *Appl. Phys. Lett.* **69** 931

Chapter 6

Growth of widegap II–VI quantum structures and their optical properties

Takafumi Yao

Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980, Japan

The spontaneous formation of II–VI widegap semiconductors have been reviewed with emphasis placed upon the interface morphology and its effects on photoluminescence properties. It is found that a common cation quantum structure such as ZnS/ZnSe ultrathin QW's lead to an island interface which ensures localization at the interface thereby facilitating the growth of QDs at the heterointerface. Such quantum structures show a broad exciton luminescence band with developed low-energy tail, which is characteristic for a QD system with random distribution of lateral confinement. A lineshape analysis of the luminescence band suggests the formation of QDs with an average height of 3 ML and diameter of 2.9 nm. In contrast, the common anion interface such as ZnSe/CdSe ultrathin QWs leads to interface alloying due to large interdiffusivity of metal atoms. The linewidth of exciton luminescence in ZnSe/CdSe ultrathin QWs becomes narrower as the QW thickness decreases. Temperature-dependent, time-resolved, and micro PL properties of the QDs are investigated.

6.1 Introduction

Widegap II–VI compounds are characterized by the direct bandgap with either zinc blend or wurtzite structures and large exciton binding energy. Table 6.1 summarizes bandgap energy and exciton binding energy of typical II–VI compounds and some other III–V compounds which have similar bandgap energies. As a general tendency, the exciton binding energy of II–VI compounds is larger than that of III–V compounds with similar bandgap energy. Those

Table 6.1. Energy bandgap at RT and exciton binding energy of II–VI compounds and other materials.

	Energy bandgap at RT (eV)	Exciton binding energy (meV)
ZnO	3.37	60
ZnS	3.54	39
ZnSe	2.70	20
CdSe	1.74	16
GaAs	1.43	4.2
GaN	3.39	21

properties make the materials very attractive both from scientific points of views and optical device application aspects.

Excitonic optical transition is characterized by large oscillator strength. Nonlinear optical effects are enhanced when the associated optical transition has a large optical oscillator strength. One of such examples can be seen in room temperature excitonic lasing [1] and high-temperature stimulated emission up to 550 K [2] due to an excitonic mechanism from ZnO epilayers grown by plasma-assisted MBE [3]. The characteristic temperature for the threshold intensity of stimulated emission was as large as 90 K [4], which is very large for a structure without carrier and optical confinement. We note that this large oscillator strength would even be enhanced when biexcitons participated in the optical transition processes due to giant oscillator strength effects.

The binding energies of exciton and biexciton increase due to quantum confinement effects in low-dimensional quantum structures [5], which is favourable for room-temperature operation of nonlinear optical devices. The oscillator strength is also increased due to an increase in density of states at around the bandgap energy. Those properties would contribute to realization of low threshold nonlinear optical devices at room temperature or even at higher temperatures.

The II–VI quantum structures have been fabricated by several methods including (1) micro-fabrication techniques [6]; (2) growth on patterned substrates [7]; and (3) spontaneous formation of quantum structures [8]. The quantum structure contains heterointerfaces which can be classified into the following categories: (1) common cation heterointerface such as ZnSe/ZnS [9]; (2) common anion heterointerface such as CdSe/ZnSe [10]; and (3) heterointerface without common elements such as ZnSe/CdTe.

Most II–VI heterostructures are lattice mismatched systems, in which misfit lattice strain is built up as growth proceeds [11]. The free energy (G) of the system can be expressed as $G = E - TS$. The energy (E) consists of strain energy and electronic energy. In a heterostructure system, there is an increase in energy due

to the presence of lattice misfit strain. This increase in energy can be balanced through intermixing or clustering. The intermixing will increase entropy (S) of the system, which reduces the total free energy. The clustering occurs to minimize the interface energy by reducing the interface area, which leads to the formation of quantum dots. However, we note that kinetics will limit the formation of the interface structure especially in the case of low-temperature growth.

One of the important factors which influence the interface structure is the interdiffusion of atoms through the interface. In most cases, the interdiffusivity of cation atoms is faster than that of anion atoms. For instance, the diffusion length of Cd atoms into bulk ZnSe is estimated to be about 0.04 nm at 280 °C for 20 min annealing, while the diffusion length of Se into bulk ZnS is as small as 0.0002 nm for the same annealing condition [12]. It is natural to assume that the lateral interdiffusion on the growing surface is much faster than that along the growth direction, since the surface can be regarded as a two-dimensional gas/liquid/solid system and the bond energy of the surface atoms are smaller than that of the atoms in bulk crystals.

This paper will overview the spontaneous formation of II-VI quantum dots and their optical properties with special attention placed upon the interface morphology by taking ZnSe/ZnS and CdSe/ZnSe heterostructures as typical examples. The lattice misfits of the ZnSe/ZnS and CdSe/ZnSe heterostructures are estimated to be 4.5% and 7%, respectively. The interface morphology will be effectively characterized by optical properties.

6.2 ZnS/(ZnSe)_n/ZnS quantum dots

6.2.1 Fabrication of ZnS/(ZnSe)_n/ZnS ultrathin quantum structures and their characterization

As described above, a ZnS/(ZnSe)_n/ZnS ultrathin quantum structure has common cation heterointerfaces, in which interface alloying can be neglected at low growth temperature because of low interdiffusivity of chalcogen atoms. ZnS/(ZnSe)_n/ZnS ultrathin quantum structures have been fabricated by molecular beam epitaxy (MBE) or atomic layer epitaxy (ALE) either on GaAs (100) [8] or GaP (100) substrates [13]. It was found that a 0.5 μm thick ZnS buffer layer predeposited on a GaAs (100) substrate was thick enough to fully relax lattice misfit strain, thereby achieving relatively smooth surface morphology [8]. ZnSe quantum well layers with thickness ranging from a few monolayers to fractional monolayers were deposited on a ZnS buffer layer by ALE followed by deposition of a 0.1 μm-thick ZnS barrier layer by MBE. The substrate temperatures for the growth of ZnS and ZnSe were respectively changed as 130 °C and 210 °C to optimize growth conditions. It is possible to grow pseudomorphic ZnS/(ZnSe)_n/ZnS ultrathin quantum well structures on GaP (100) substrates, since the lattice misfit of ZnS with GaP is as low as 0.7% [13]. Pseudomorphic quantum structures consisting of monolayer ZnSe quantum wells and 20 nm thick

ZnS barrier layers were grown on a 50 nm thick ZnS buffer layer.

Prior to the growth of quantum wells, the surface morphology of ZnS buffer layers was evaluated by atomic force microscopy (AFM). Although the reflection high energy electron diffraction (RHEED) pattern showed a well-developed streaky structure during growth, the AFM observation revealed a rather rough surface with a thickness fluctuation of a few nanometres in a submicrometre range and up to 20 nm in a range of 15 μm . The quantum structures were grown on flatter ZnS buffer layers. Such a microscopically rough surface would allow the formation of quantum dot structures at the heterointerface by the deposition of ZnSe ultrathin layers followed by the deposition of a ZnS capping layer. As has been revealed by AFM observation, the surface of the ZnS buffer layer consists of small valleys and islands. At the very beginning of heteroepitaxy on such a rough surface, where only a submonolayer is deposited on the substrate, isolated small islands with an average height of monolayer and various lateral sizes would be formed on the small terraces and valleys of the ZnS buffer layer. As the deposition of ZnSe increases with its amount still being in the submonolayer regime, the neighbouring small islands formed on adjacent terraces with one monolayer height difference would get connected to form coupled islands. However, those coupled islands would be isolated from others due to surface roughness with much larger fluctuation leaving some lateral size fluctuation. With an increase in further deposition, the islands formed on terraces with one monolayer height difference get connected to form extended islands with an average height larger than 1 ML but smaller than 2 ML. Many of those extended islands would still be isolated if the surface roughness is larger than 2 ML. When a ZnS capping layer is subsequently deposited after deposition of ZnSe, the 2 ML thick ZnSe islands formed by the surface roughness are three-dimensionally covered by ZnS, which results in three-dimensional quantum confinement of electrons and holes. Since the lateral dimensions of such quantum dots are distributed almost randomly, the energy spread of the quantum level would become significant. Consequently, the luminescent energy broadening occurs with an upward shift of the quantum level being expected.

Figure 6.1 shows photoluminescence (PL) spectra from the ZnS/(ZnSe)_n/ZnS ultrathin quantum well structures grown on GaAs (001) for various ZnSe QW thickness ranging from 1 ML to 4 ML [8]. The photoluminescence was obtained using the 325.0 nm line from a He–Cd laser at an average power density of 200 mW cm⁻². The integrated PL intensity showed a linear dependence on the excitation strength in the range of 2–800 mW cm⁻². The spectra are dominated by the free excitonic emission line. The blue-shift in energy and broadening of the excitonic emission with decrease in the ZnSe QW thickness are noted.

Figure 6.2 plots the emission peak energy against the ZnSe quantum well width [8]. The linewidth is shown by a vertical bar. As the well width decreases from 20 to 1 ML, the emission energy increases from 2.92 to 3.47 eV. The full curve shows the dependence of the calculated emission energy on the well thickness based on a simple square well potential model taking the strain effect

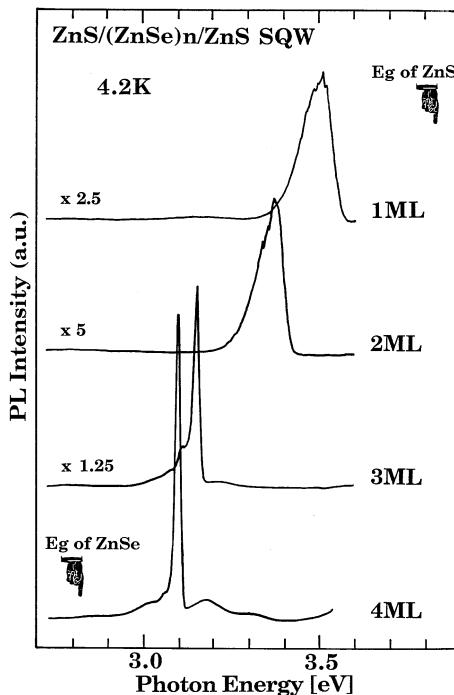


Figure 6.1. The PL spectra from $\text{ZnS}/(\text{ZnSe})_n/\text{ZnS}$ ultrathin SQWs measured at 4.2 K [8].

into account. The linewidth of the emission from QWs thicker than 3 ML varied sample by sample but remained in the range from 15 to 30 meV, which is even narrower than the reported linewidth for ZnSe/ZnS superlattices [20, 21]. However, the emission peak shows a developed band tail on its low-energy side as seen in figure 6.1, which is a typical behaviour of disordered or localized systems. The linewidth increases abruptly as decreasing below 3 ML: 86 meV for 2 ML and 115 meV for 1 ML. This increase in linewidth is even more prominent in fractional quantum structures, where localized 0D nature should be more pronounced. Those facts suggest a transition from a two-dimensional (2D) structure with weak localization to a strongly localized 0D-like structure, however, with varied degree of localization, which will be discussed based on detailed temperature-dependent PL measurements [13].

In order to get more insight into the formation processes of such quantum structures, a series of fractional quantum structures were fabricated, where the ZnSe QW layer was deposited by ALE with the Se shutter being varied from 1–120 s, while the Zn shutter was kept at 60 s and the interval period in between each open period of the shutter was kept at 5 s. The PL emission was dominated by free exciton emission, which indicates high quantum efficiency of the quantum

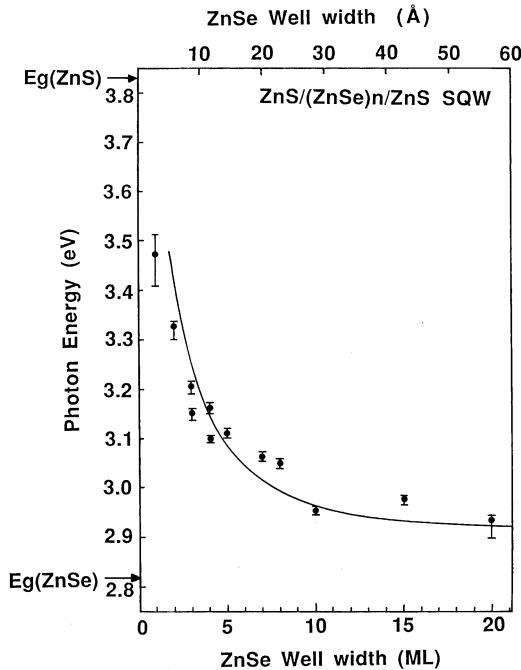


Figure 6.2. The dependence of the emission energy of ZnS/(ZnSe)_n/ZnS ultrathin SQWs on the ZnSe QW thickness. The vertical bar indicates the linewidth of the emission peak [8].

structures. Figure 6.3(a) shows a plot of the emission peak energy of free exciton against the Se shutter period. The vertical bar indicates the linewidth. The emission energy reaches as high as 3.6 eV for a SQW with 1 s Se shutter period and shows a red shift of 80 meV as the Se shutter period increases to 5 s. Interestingly, the emission peak stays at around 3.49 eV for the Se shutter period of 10–60 s. As the Se shutter period increases from 60–120 s, the emission peak shows another red shift to 3.4 eV. The emission peak energy for the shutter period of 10–60 s coincides with that for a 1 ML thick ZnSe QW, while that for 120 s agrees with that of a 2 ML thick ZnSe QW. It should be noted that the emission linewidth varies in the same manner as the emission peak energy, as the Se shutter period is changed, as shown in figure 6.3(b). The emission linewidth for the Se shutter period of 1 s is as broad as 150 meV and decreases to 120 meV as the shutter period is shortened to 5 s. The linewidth shows a plateau for a shutter period of 10–60 s. The linewidth for 5–60 s nearly coincides with that of 1 ML QW and that for 120 s with that of 2 ML QW.

These behaviours can be interpreted in terms of the transition of the surface structure from an extended quantum structures to an isolated quantum structure

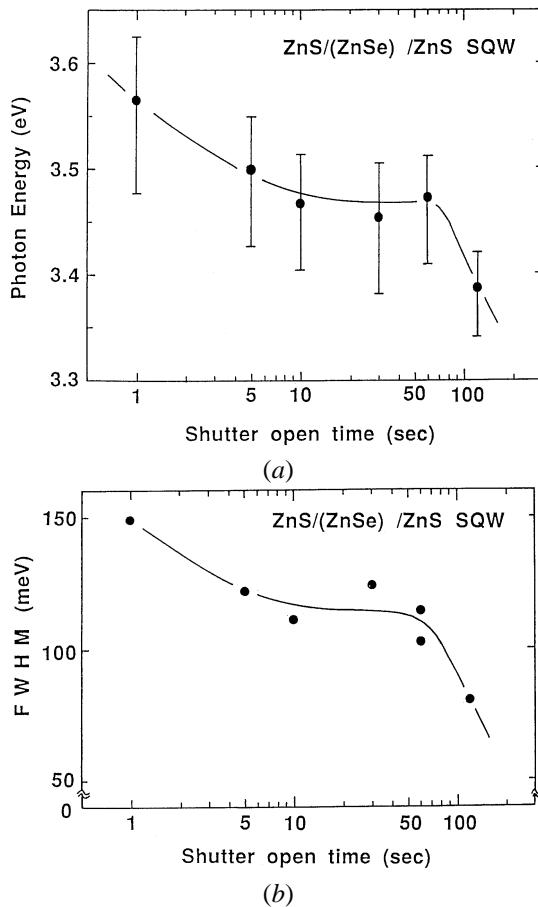


Figure 6.3. (a) Dependence of the free exciton emission energy of $\text{ZnS}/(\text{ZnSe})_n/\text{ZnS}$ ultrathin SQWs on the Se shutter period. The ZnSe QW layers were prepared by ALE with the Zn shutter being kept at 60 s, while the Se shutter was varied from 1–120 s. The vertical bar shows the linewidth of the emission peak. (b) Dependence of the emission linewidth on the Se shutter period.

as the deposition of ZnSe is reduced in the submonolayer regime. In the small deposition limit of ZnSe, isolated ZnSe quantum dots with an average height of one monolayer and various lateral dimensions are formed on the surface of the ZnS buffer layer, which should give rise to the emission energy higher than that of 1 ML QW and broader emission linewidth. As the ZnSe deposition increases, the isolated islands formed on adjacent terraces with one monolayer height difference would get connected. Those coupled islands or coupled quantum dots would give rise to the emission at similar energy as 1 ML QW with still broadened

linewidth reflecting the random distribution of lateral size in addition to thickness fluctuation. With further increases in ZnSe deposition, the average island height becomes larger than one monolayer and the emission from ZnSe quantum dots reaches similar energy as 2 ML QW. Thus, it would be possible to monitor the growth step of the quantum structures by investigating the PL properties at every deposition step.

Since the QDs obtained here are based on potential fluctuations, the size distribution is far from uniform. Although AFM has proved to be a powerful tool for measuring the size of nanostructures, it only gives a correct height of the dot when the underling layer is atomically flat which is not the case here. Instead, it is possible to estimate the lateral dot size through fitting the line shape of PL spectra [14]. For simplicity, the dot is assumed to be cylindrical in shape with the same height in the growth direction and a variable size in the lateral direction. Due to the stochastic nature of the growth process, the lateral size can be assumed to have a Gaussian distribution with a mean value d_0 and a characteristic width σ . The probability distribution of carriers participating in the PL process is given by the product of size distribution and the number of carriers in a specific dot with lateral size of d [16]. The PL spectrum is readily obtained by converting the size distribution into energy distribution in the energy space which reads,

$$I_G(\delta E) = \frac{A}{\delta E^4} \exp \left[-\frac{1}{2} \left(\frac{d_0}{\sigma} \right)^2 \left(\frac{\delta E_0}{\delta E} - 1 \right)^2 \right] \quad (6.1)$$

where A is a constant and δE is the blue-shift caused by the additional confinement in the lateral direction. In deriving equation (6.1), we have assumed that (1) the confinement induced blue-shift is given by $\delta E = B/d$ with B being a properly dimensioned constant¹, (2) $n = 2$ level is far away from the $n = 1$ level so that the carrier distribution factor can be ignored, and (3) the oscillator strength is independent of the dot size. Therefore, δE_0 in equation (6.1) corresponds to the blue-shift of a dot having a lateral size d_0 .

The Gaussian distribution, however, gives a high-energy tail in the PL spectrum which is opposite to the experimental observations of a tail at the low-energy side for ZnSe/ZnS systems [8, 13]. One may attribute the discrepancy to the inappropriate assumption of a Gaussian distribution of the size. However, this is the most reasonable assumption we can make at present given the insufficient evidences for assuming other types of distributions. On the other hand, the ZnSe/ZnS system under investigation essentially belongs to the category of a random or disordered dimensional system in which the density of energy state usually has a low-energy tail [15, 17–19]. By taking this effect into account, the

¹ It is well known that the energy upshift of electrons in an infinite QW is inversely proportional to the square of the well width. In most of the practical case, however, we found that the bandgap upshift is better described by $\delta E = B/d$ with B being a properly dimensioned constant. Through fitting with the calculated bandgaps based on the effective mass model by taking into account the strain effect, we obtained an energy shift $\delta E = 2.1/d$ [eV Å] for ZnSe/ZnS.

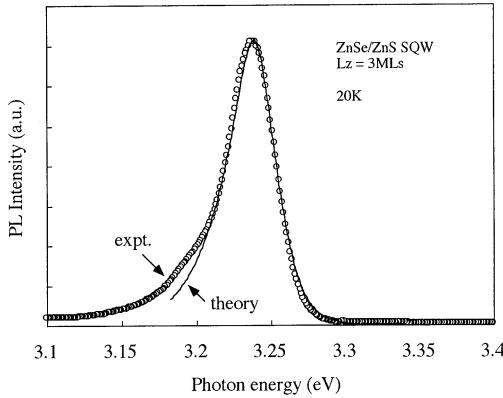


Figure 6.4. The PL spectrum of ZnSe/ZnS QDs with a height of 3 MLs (circles). The full curve shows the calculated result according to equation (6.1) using $E_0 = 5$ meV, $\alpha(E - E') = 1.3$ and a constant g_0 [14].

low-temperature PL spectrum can be written as

$$I_0(E) = \int_0^E I_G(E') g_0(E - E') \exp \left[- \left(\frac{E - E'}{E_0} \right)^{\alpha(E-E')} \right] dE' \quad (6.2)$$

where g_0 , E_0 , and $\alpha(E - E')$ are characteristic parameters associated with the random system, which can be obtained through fitting the experimental spectrum. As shown in figure 6.4, the best fitting to the experimental data was obtained for the 3 ML QW with $d_0 = 2.9$ nm, $\sigma = 0.5$ nm, $E_0 = 5$ meV, $\delta(E - E') = 1.3$, and a constant g_0 ². Although the values obtained here are more or less dependent on the parameters used to calculate the quantum confinement induced energy shift, it is interesting to note that the average size is just half of the exciton diameter in ZnSe. We note that in such kinds of small dots, a strong confinement effect can be expected.

6.2.2 Temperature dependence of PL of ZnS/ZnSe quantum dot structures

The zero-dimensional nature of the randomly distributed dot structures is also reflected in the temperature dependence of the broadened PL lines. The PL intensity and line shape were found to be very sensitive to temperature. Figure 6.6 shows the PL peak energy as a function of temperature for the QW's with widths ranging from 1 to 4 ML [13]. The temperature induced redshift of bulk ZnSe is also given in the figure by broken curves, which are up-shifted along the energy axis to facilitate comparison. Neither of the QW's shows a bulk-like temperature dependence at a temperature ranging from 20–80 K. A

² Refer to [18] for the detailed physical meaning of these parameters.

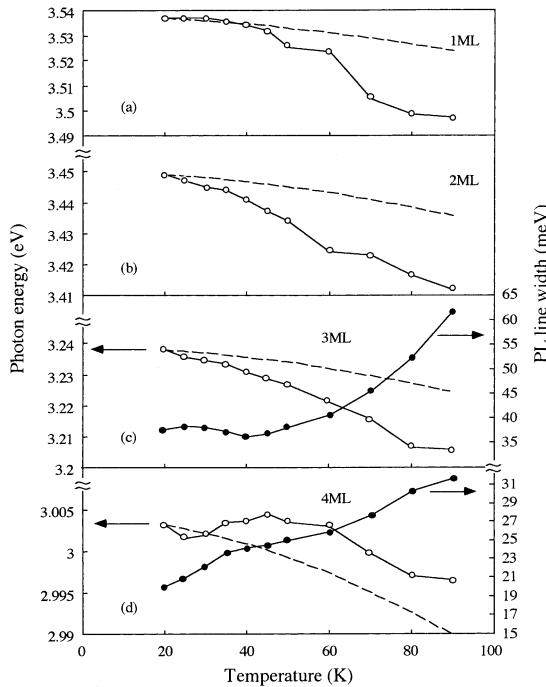


Figure 6.5. Temperature dependence of the PL peak for bulk ZnSe (broken curve) and QW's (circles) with different well widths. Also shown in (c) and (d) are the temperature dependence of the PL linewidth of the 3 ML QW and the 4 ML QW, respectively. Note that the vertical scale of (d) is different from (a)–(c) [13].

rather different temperature dependence was observed for the 4 ML thick QW as compared to the other three narrower QW's.

The temperature dependence of the 4 ML QW shows typical behaviour of excitons localized at the interface due to potential fluctuations which have been observed in a variety of QW's and superlattices [20, 25–27]. In a quantum well with potential fluctuations, excitons will be localized at potential minima at low temperature. As the temperature increases, the localized excitons will be thermally released to become free excitons or hop to other localized sites, resulting in a blue-shift of the PL peak energy. If this shift is larger than the temperature induced redshift of the bandgap, a net blue-shift will be observed in a certain temperature range, as shown in figure 6.5(d) for the 4 ML QW.

In contrast to the blue-shift of the 4 ML QW, a redshift up to 40–50 meV was observed for the other three QW's when temperature is increased from 20–80 K, which cannot be explained by the temperature-induced bandgap change alone. The temperature-induced redshifts of the three narrower QW's are larger than bulk ZnSe by a factor of four. In order to clarify the mechanism of such

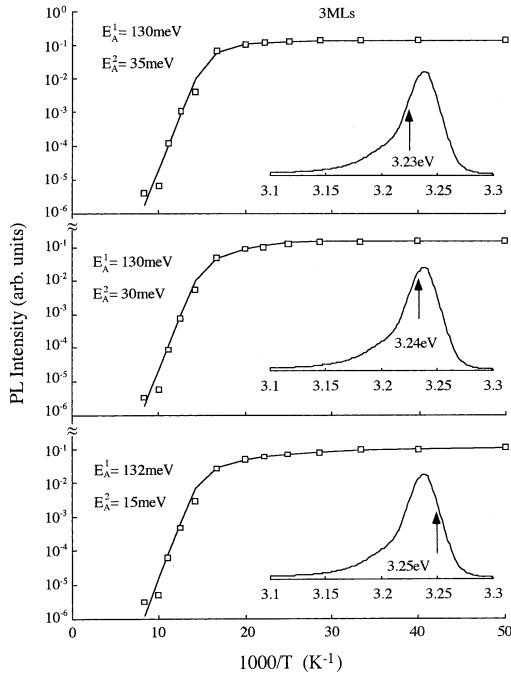


Figure 6.6. Arrhenius plots of the PL intensity of the 3 ML QW at different energy position versus inverse temperature. The full curves are the fitted curves using equation (6.3). The arrows in the insets give the energy positions at 20 K and the temperature induced energy variation is assumed to follow the band gap of ZnSe, i.e. $E_g(T) = E_g(0) - 8.59 \times 10^{-4} T^2 / (T + 405)$ [13].

a large redshift, temperature dependence of the PL intensity for the 3 ML QW was investigated in the temperature range 20–110 K, as shown in figure 6.6 [13]. Since the PL peak energy decreases significantly with increasing temperature, we chose a certain energy position on the spectrum at 20 K and investigated the intensity variation at a photon energy that follows the movement of the bandgap of bulk ZnSe as temperature changes. The intensity variation at three different positions on the PL curve as indicated by the arrows in the insets of figure 6.5 were theoretically fitted using the following equation [22, 23, 28] assuming two non-radiative loss channels with thermal activation energies E_A^1 and E_A^2 :

$$I_{PL}(T) = \frac{I_0}{1 + A_1 \exp(-E_A^1/kT) + A_2 \exp(-E_A^2/kT)} \quad (6.3)$$

where A_i is the ratio of the exciton lifetime in the quantum well over the effective scattering time from the quantum well to the barrier states. The activation energies used for obtaining best fitting to the experimental data are given in the

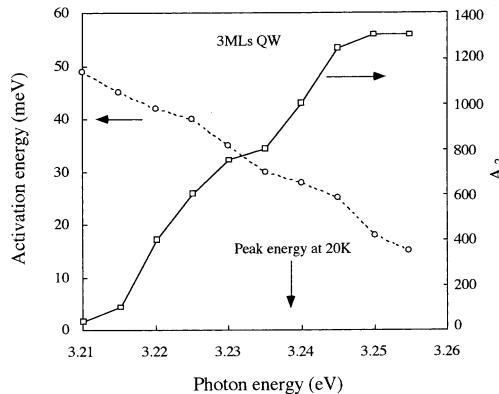


Figure 6.7. Thermal activation energy of one of the non-radiative loss channels and the corresponding characteristic constant A_2 appears in equation (6.3) versus the observation energy [13].

figure. The loss channel that determines the PL quenching process above 60 K has an activation energy of about 130 meV, which is almost independent of the observation energy. On the other hand, the activation energy of the other loss channel that governs the thermalization process in the temperature range 30–60 K is strongly dependent on the observation energy. Similar curve fittings have also been performed at other energy positions.

Figure 6.7 plots the obtained activation energies together with the characteristic parameter A_2 against photon energy. We note that the difference between the PL peak energy and the bandgap of ZnS for the samples under investigation (560 meV) is significantly larger than the activation energies obtained above. This is quite different from most quantum structures, in which the activation energy is close to the difference between the PL photon energy of the quantum well and the barrier bandgap [22–24]. Considering the negligibly small conduction band offset between ZnSe and ZnS in the ZnSe/ZnS system, one may think that electrons rather than electron–hole pairs are thermally emitted out of the quantum well and the thermal activation energy E_A^2 should correspond to the energy difference between the first electron quantized level in the QW and the top of the conduction band of the barrier layer (hereafter we call it δE_{e1-b}). However, this is unlikely because the difference of thermal activation energies between two observation positions is almost equal to the difference of observation energies as shown in figure 6.7, rather than equal to δE_{e1-b} .

The other possible factors responsible for the small thermal activation energy of the ZnSe/ZnS structures may relate to the defects formed at the ZnSe/ZnS interfaces or in the ZnS barriers. Both factors are very probable because (1) thin heteroepilayers grown in a three-dimensional growth mode generally contain a high density of defects, and (2) a large portion of wavefunctions penetrates

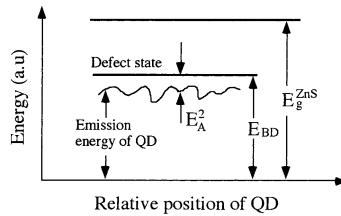


Figure 6.8. Schematic illustration of the defect state supposed to be responsible for the PL quenching process together with the energy levels of QD's with different sizes [13].

into the barrier layer due to the small thickness of the well layer. Following this line of interpretation, we may plausibly assume an effective energy barrier E_{BD} associated with the defect states for excitons in the quantum wells, which are illustrated schematically in figure 6.8. The activation energy is given by the difference between E_{BD} and the emission photon energy. The activation energy differs from dot to dot due to different size. A smaller dot would give a smaller activation energy which is exactly what is shown in figure 6.8. The weaker dependence of E_A^1 on the emission energy suggests that this activation energy is related to the quenching process in the barrier layer. It should be noted here that the explanation given above by no means suggests the existence of a fixed defect state in ZnSe/ZnS heterostructures but the defects might be distributed in a wide range within the bandgap of ZnS and also might be different from sample to sample.

The temperature dependence of PL linewidths for the 3 ML and 4 ML QW's are also shown in figure 6.5, while the characteristic parameter A_2 is given in figure 6.6. The PL linewidth of the 3 ML QW is almost constant from 20–30 K and then decreases by a few meV from 30–40 K. The latter is obviously caused by the dot size dependent activation energies as discussed above, i.e. at a certain temperature, the high-energy side of the PL line quenches faster as compared to the lower-energy side, resulting in a narrowing of the emission line. On the other hand, the increase of A_2 with the emission energy can be understood as caused by the increase in the surface-to-volume ratio with reducing the dot size because A_2 is proportional to the non-radiative rate at elevated temperature. This supports our explanation of the energy-dependent activation energy given above. In contrast to the 3 ML QW, the PL linewidth of the 4 ML QW shows a steady increase with increasing temperature, which is normal for most of the quantum wells.

6.2.3 Time resolved PL

As is well known, the density of electronic states are proportional to $\delta(E)$ for zero-dimensional quantum dot, $1/\sqrt{E}$ for one-dimensional quantum wire, and constant for two-dimensional quantum well. These properties reflect in the radiative lifetime of confined excitons in quantum structures: the life time τ of excitons in a

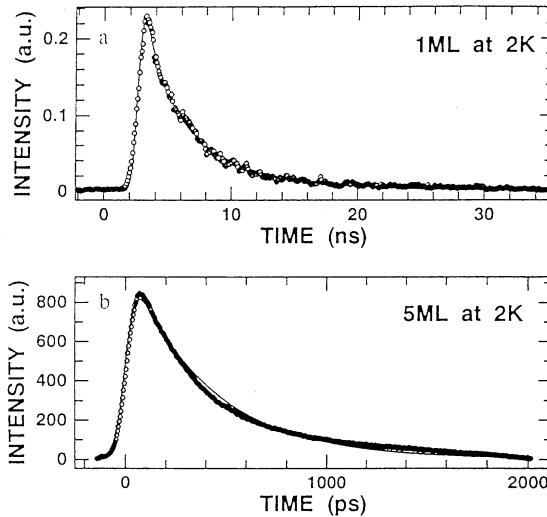


Figure 6.9. The decay curve of the excitonic PL of (a) 1 ML and (b) 5 ML systems at 2 K in the excitation of 315 nm light. The full curve in (a) is the fitting results with the stretched exponential $A \exp\{-(t/t)^\beta\}$, and the full curve in (b) is the fitting result with the exponential, $A \exp\{-(t/t)\}$ [31].

quantum dot is independent of temperature [29], $\tau \propto \sqrt{T}$ for quantum wire [30], and $\tau \propto T$ for quantum well [30]. In order to identify the dimensionality of the electronic states in the ZnSe/ZnS quantum structures, the radiative lifetime of the ZnS/(ZnSe)_n/ZnS ultrathin QW's was investigated using picosecond time resolved PL techniques [31].

Figure 6.9 shows the decay curves of the exciton emission of (a) 1 ML and (b) 5 ML QW's at 2 K excited by the 315 nm light from a frequency doubled dye laser pumped by a frequency doubled YAG laser [31]. The decay curve for 1 ML QW is well fitted by a stretched exponential equation, $A \exp\{-(t/\tau)^\beta\}$ as shown by a full curve in figure 6.9(a), where β is a fitting parameter between zero and one, and τ is the decay time. Similar decay curves were observed at temperatures below 80 K. The obtained β value for a decay curve at 80 K is 0.55. Such a small β value is consistent with a disordered exciton system in 1 ML QW which is characterized by thin quantum dots with various lateral dimensions. In contrast to the 1 ML QW, the decay curve of 5 ML QW is well fitted by a simple exponential equation, $A \exp\{-(t/\tau)\}$ as shown in figure 6.9(b). This is equivalent with $\beta = 1$ in the stretched exponential relationship, thus exhibiting that the exciton system in 5 ML QW has a two-dimensional nature with small randomness. This is consistent with the observed narrow PL linewidth of 5 ML QW shown in figure 6.1.

Figure 6.10 shows the temperature dependence of the PL decay time of

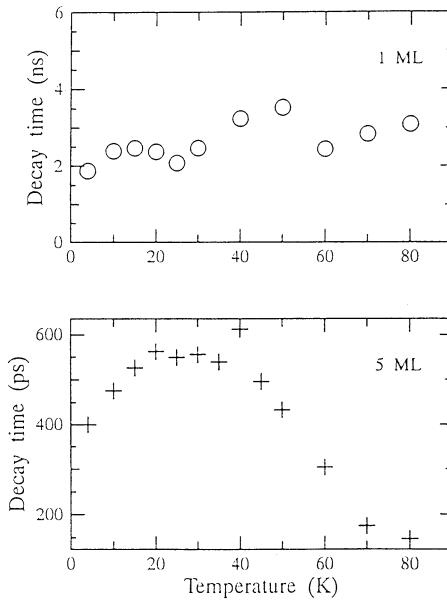


Figure 6.10. Temperature dependence of the measured PL decay time of 1 ML and 5 ML structures [31].

the 1 ML and 5 ML QW structures [31]. The radiative lifetime for the 1 ML system is almost independent of temperature in the low-temperature region below 40 K, while the exciton radiative lifetime for the 5 ML system increases with an increase in temperature below 20 K. The observed different dependence on temperature exemplifies that both systems are different in electronic nature. The temperature dependence for the 1 ML QW is essentially the same as that reported for GaAs quantum dots [29] and indicates that the 1 ML ZnSe/ZnS QW indeed has zero-dimensional nature. In contrast to 1 ML QW, the observed temperature dependence of the exciton radiative lifetime of the 5 ML system is consistent with the two-dimensional nature in 5 ML QW as discussed above. The radiative lifetime of the 1 ML QW is much longer than that of 5 ML. The observed slow lifetime in the 1 ML ZnSe QW may be attributed to the phonon bottleneck problem, which may occur in a zero-dimensional system [38].

6.2.4 The effect of substrate temperature on the formation of quantum structures

As has been discussed, the formation of quantum structures is controlled by the interplay of energetics and kinetics. Since the substrate temperature adopted for the formation of quantum dots has been quite low, a change in substrate temperature may induce substantial change in the formation of quantum

structures. As can be expected from the consideration of the total free energy, the high-temperature growth would promote the spontaneous formation of quantum structures. This is clearly observed in the PL properties of ZnSe/ZnS QW's grown at relatively higher substrate temperatures as shown in figure 6.11 [32], where the ZnSe/ZnS QW is fabricated by ALE with Zn and Se supply times being 12 and 60 s, respectively. Note that the simultaneous supply of Zn and Se for 3 s corresponds to the growth of 1 ML thick ZnSe. At lower substrate temperature, a broad emission with a linewidth of 100 meV at around 3.5 eV dominates PL. In addition, smaller emission bands at around 3.1 eV are observed. Those broad luminescence bands are attributed to the excitonic emission from the ultrathin ZnSe QW's essentially the same as shown in figure 6.1. The emission band at around 3.0 eV is ascribed to the deep level luminescence from the ZnS barrier layers. As the substrate temperature increases, the broad emission bands (labelled 'b', and 'c') intensity decrease and a sharp new emission line (labelled 'a') with a linewidth of 21 meV emerges at around 3.0 eV, which suggests that the quantum structure responsible for this sharp emission line should have an average height of 4 ML. Although the peak positions of the 'b' and 'c' emission bands vary with the substrate temperature, they can be assigned to quantum structures with average thicknesses of 2 ML and 1 ML, respectively, judging from the emission energy. Since the luminescence bands have broad linewidth with an extended tail, it is likely that those quantum structures are localized at the ZnSe/ZnS heterointerface with varied lateral dimensions as mentioned above.

In order to study in more detail the nature of the 'a' emission line, photoluminescence excitation (PLE) spectra were measured with the detection photon energy placed at around this sharp emission peak (figure 6.12) [32]. If the detection is even slightly off-tune from the 'a' emission line, two absorption bands located at 3.4 eV ('b'') and 3.6 eV ('c'') are observed. Those two absorption bands can be observed irrespective of the detection energy and are basically responsible for the emission bands 'b' and 'c', respectively. If the detection is tuned at the 'a' emission band, additional two emission bands appear: one (hereafter called 'a'') is 7 meV above the 'a' emission band and the other is located at around 3.2 eV. It should be noted that the observed sharp emission bands and a small Stokes shift of 7 meV are commonly observed features in well-defined self-organized quantum structures [42]. In addition, spatially-resolved cathodoluminescence (CL) study on this 'a' emission band indicates spot-like spatial distribution of the emission area with a typical diameter of 1 μm . Note that this is comparable with the spatial resolution of 1 μm in this study and that this apparent diameter does not indicate real shape. However, it is true that the observation does support the conjecture that the 'a' emission band is due to quantum dot structures with a typical height of 4 ML. It would be worth commenting on the absorption band located at 3.2 eV which appears at higher energy side of the 'a'' absorption band. Recent theoretical calculation suggests that this emission band can be attributed to excited states of the quantum dots [42].

The effect of the substrate temperature on the growth of ZnSe/ZnS quantum

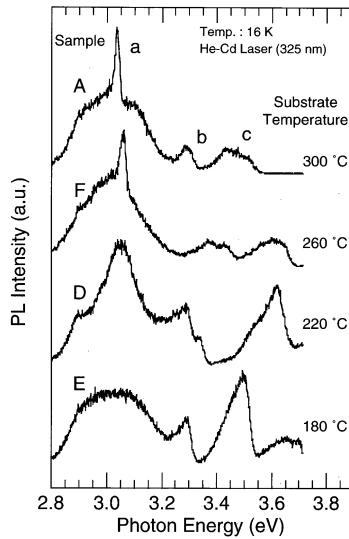


Figure 6.11. PL spectra from ZnSe/(ZnS)₁/ZnSe ultrathin QW structures grown by ALE at different substrate temperatures. The dwell times for Zn and Se in the ALE process were fixed at 12 and 60 s, respectively [32].

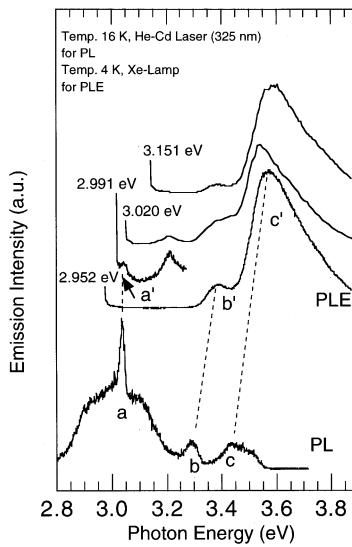


Figure 6.12. PLE spectra of a ZnSe/ZnS QW grown at 300 °C with the detection photon energy placed at around the 'a' emission band [32].

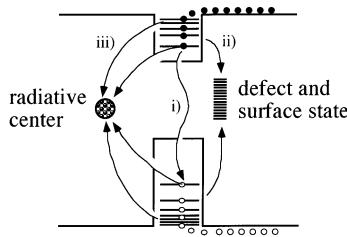


Figure 6.13. The phonon bottleneck problem and three major carrier recombination and loss channels in doped quantum dots [14].

dots can be summarized as follows. As the substrate temperature increases, the migration of adatoms is enhanced to form more energetically favourable interface structures, which results in the formation of larger dots of 4 ML thick with better uniformity in size. We stress again that the linewidth of this emission band is 21 meV which suggests again well-ordered QD's. This is in contrast to the broad emission bands at around 3.5 eV which arise from random QD's.

6.2.5 Mn doped ZnSe quantum dots

Luminescent centres in a semiconductor material can be excited either via direct pumping the centres or indirectly through energy transfer from the host materials [33, 34]. The latter is of more importance for practical applications because it allows the realization of light emitting devices through current injections. Neglecting the non-radiative carrier loss channels, the energy transfer rate should be proportional to $\tau_{rh}/(\tau_{rh} + \tau_T)$ with τ_{rh} as the radiative lifetime of the host carriers and τ_T the energy transfer time, respectively. This relationship implies that there are two ways to enhance the energy transfer, i.e. either increasing τ_{rh} or reducing τ_T . A quantum dot is an ideal system for achieving this purpose because τ_{rh} is expected to be much longer than those of bulk materials if the relaxation time is also taken into account and τ_T can be expected to be shorter due to the strong spatial coupling between the host and the luminescence centres. The latter can even be made shorter through choosing an appropriate host material and controlling the dot size. This idea is depicted by figure 6.13, where the excited carriers should have three major recombination or loss channels, i.e. (1) recombination radiatively and emit photons with an energy equal to the effective bandgap of host materials, (2) trapped by defects, and (3) transfer energy to radiative centres. The phonon bottleneck problem and relatively longer intrinsic lifetime in QDs [38] can be effectively used to enhance the energy transfer from host to luminescent centres.

As was discussed in a previous section, undoped ZnSe/ZnS QW's exhibit strong excitonic emission at low temperature. From temperature-dependent steady-state and time-resolved PL measurements, it is found that dot structures

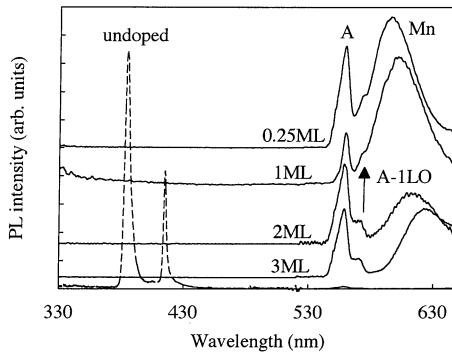


Figure 6.14. PL spectra of Mn-doped ZnSe/ZnS QW's with well widths ranging from 0.25 ML to 3 MLs [36].

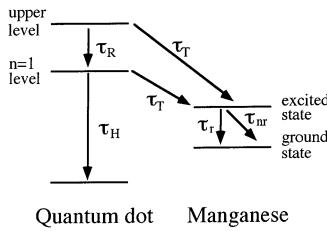


Figure 6.15. Kinetic model for illustrating the carrier generation, relaxation, recombination in host materials and the corresponding energy transfer to Mn impurities [36].

are formed in wells with widths below 3 MLs [8, 13]. On the other hand, a typical quantum well behaviour was obtained for samples with a well width larger than 4 MLs, though the excitons are not completely free but are localized at the interface at low temperature. When Mn was introduced into these structures, the excitonic emission disappeared almost completely. Instead, strong emission from Mn intra-*d* shells was observed. Figure 6.14 shows the PL spectra of Mn-doped samples with well width ranging from 0.25 to 3 MLs [36]. Peak A and the associated phonon replicas are due to the emission of GaP which has been confirmed from the PL measurement of a bare substrate. All of the samples have similar structure which are grown on (001) GaP substrates with a 50 nm thick ZnS buffer layer followed by the Mn-doped ZnSe well layer and finally a 25 nm thick ZnS capping layer. Although the quantum wells are just a few monolayers thick, the Mn-related photoluminescence is two times of magnitude stronger than that of a 1 μm thick ZnSe:Mn epilayer.

The above result is to be compared with Mn doped ZnCdSe:Mn/ZnSe QWs in which the Mn emission was found to be greatly suppressed as compared to the

bulk case [35]. The reason causing this difference might be understood as follows. The Mn atoms can be excited either by the direct absorption of excitation light or indirectly through energy transfer from the host ZnS. Given the ultrathin quantum well, the latter is considered to be dominative in this case. Since carrier capture efficiency in QW's is known to be extremely high, most of the carriers generated in the ZnS barrier will be captured by the ZnSe QW in a very short time. In undoped samples, the captured carriers will first relax to lower quantized levels and then annihilates either radiatively or non-radiatively. For doped samples, however, there is an additional annihilation channel for the captured carriers, i.e. transferring their energy to the luminescence centres. The transfer process can take place either before or after the completion of the relaxation. A kinetic model can thus be developed based on figure 6.15 to explain the dynamics of excitons and/or electron–hole pairs in Mn-doped ZnSe/ZnS QDs. For the sake of simplicity, we use a two-level model to describe the energy relaxation process. The following kinetic equations can be written down for carriers in the upper quantized level, the $n = 1$ level, and the excited state of Mn:

$$\frac{dN_1}{dt} = g_1 - \frac{N_1}{\tau_R} - \frac{N_1}{\tau_T} \quad (6.4)$$

$$\frac{dN_2}{dt} = \frac{N_1}{\tau_R} - \frac{N_2}{\tau_H} - \frac{N_2}{\tau_T} \quad (6.5)$$

$$\frac{dN_3}{dt} = g_2 + \frac{N_1}{\tau_T} + \frac{N_2}{\tau_T} - \frac{N_3}{\tau_r} - \frac{N_3}{\tau_{nr}} \quad (6.6)$$

where N_1 , N_2 and N_3 are the carrier densities at the upper quantized level, the $n = 1$ level, and the Mn excited state, respectively, g_1 is the carrier capture rate of the quantum well and g_2 is the carrier generation rate of Mn via direct optical pumping, τ_R is the carrier relaxation time, τ_T is the energy transfer time, τ_H is the effective lifetime of the well or dot, τ_r and τ_{nr} are the radiative and non-radiative lifetime of Mn d levels, respectively. The time-integrated intensity of the photoluminescence from the QW or dot (I_{QD}) and that from the Mn (I_{Mn}) are thus given by

$$I_{QD} = \frac{g_1}{(1 + \alpha)(1 + \beta)} \quad (6.7)$$

$$I_{Mn} = \left\{ g_2 + \left[\alpha + \frac{\beta}{(1 + \beta)} \right] \frac{g_1}{(1 + \alpha)} \right\} \frac{\tau_{nr}}{(\tau_r + \tau_{nr})} \quad (6.8)$$

where $\alpha = \tau_R/\tau_T$ and $\beta = \tau_H/\tau_T$. Neglecting the non-radiative effect as well as the direct excitation of Mn, the intensity ratio I_{Mn}/I_{QW} is simply given by $\alpha\beta$ when $\alpha, \beta \gg 1$. The energy transfer time in Zn_{0.77}Mn_{0.23}Se has been estimated to be 15 ps from the decay time of excitons [37]. If we assume that τ_T in the Mn-doped ZnSe/ZnS quantum dot is also of the same order of magnitude and take the average lifetime of undoped ZnSe/ZnS QDs to be 2.5 ns [31], β should be around 166. On the other hand, τ_R for an ideal QD is calculated to be in the range of μs –ns, depending strongly on the dot size due to the phonon bottleneck problem [38].

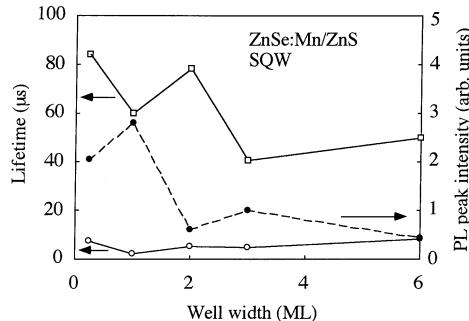


Figure 6.16. Measured lifetime and PL peak intensity of Mn emission as a function of well width. Squares and open circles: longer and shorter lifetimes, respectively; full circles: PL peak intensity [36].

In fact, a much slowed-down rising of the photoluminescence from the dot system as compared to the QW system can be seen in figure 6.9 [31]. However, the exact value of τ_R has not yet been obtained. Nevertheless, even if we assume $\tau_R = \tau_T$, the intensity ratio $I_{\text{Mn}}/I_{\text{QW}}$ is still over 100, which explains well the undetectable excitonic emissions but very strong Mn emissions in ZnSe:Mn/ZnS QD's. Questions may arise here as to why the Mn emissions from wells with width larger than 4 MLs are still quite strong in spite of the fact that samples in this thickness regime are essentially QWs rather than QDs [13]. This might be caused by the exciton localization at the ZnSe/ZnS interfaces, which has been found to be a common feature of this system. For this kind of localized system, it has been found that the relaxation and recombination times are also longer than those of the QWs with abrupt interfaces [39]. Exciton localization in ZnCdSe/ZnSe system has been found to be not as serious as the ZnSe/ZnS system because the latter generally shows a prominent low-energy tail in the PL spectra, while it is almost not noticeable in ZnCdSe/ZnSe QWs presumably due to the common anion interface as discussed in the next section. This might be the reason why Mn emissions have been suppressed in ZnCdSe:Mn/ZnSe QW's.

Figure 6.16 shows the lifetime of Mn emission from ZnSe:Mn/ZnS QD structures together with the peak intensity of steady-state PL as a function of the ZnSe well width. The decay curve of Mn emissions is characterized by two different lifetimes, though a 1 μm thick ZnSe:Mn film exhibits an exponential decay with a lifetime of 21 μs . The non-exponential decay might be caused by different crystal-field environments felt by the Mn atoms. However, if this is the case and neglecting the effect of the second nearest-neighbour atoms, the Mn should experience five types of different crystal-field environment which reduces to three if only the symmetry is concerned; thus one should at least have three different lifetimes, especially when the well width is below 1 ML. Considering the fact that the lifetime of the slow process shows a weak dependence on the

well width, while that of the faster process is almost constant, the latter is likely to be related to defects either at the interfaces or inside the ZnS layers which have been shown to be the main sources of PL quenching in undoped samples [13]. Regardless of the origins, it is worth noting that (1) the lifetime of both processes are much larger than the nanosecond lifetime reported for Mn-doped ZnS nanocrystals [40], and (2) both the lifetime and the PL intensity increases with reducing well width as shown in [figure 6.16](#). If one compares the luminescence intensity per Mn atom for the Mn-doped ZnSe layer and 1 ML thick ZnSe:Mn/ZnS QD, the enhancement of Mn luminescence intensity in the Mn-doped QD is almost three orders of magnitude assuming the absorption depth of 100 nm for the ZnSe:Mn layer. These results imply that the strong Mn emission from the QDs under study is mainly caused by the relatively enhanced energy transfer rather than the lifetime shortening of intra-*d* shell transitions.

6.3 CdSe/ZnSe quantum dots

6.3.1 Alloying at the heterointerface

PL properties of CdSe/ZnSe quantum well structures show striking differences from the common cation ZnSe/ZnS quantum well structures. [Figure 6.17\(a\)](#) shows the PL spectra from ZnSe/(CdSe)_{*n*}/ZnSe (*n* = 1–4, 6 ML) SQWs. The growth of the quantum well structures was carried out at 280 °C under optimized growth conditions in terms of PL properties. The free excitonic emission dominates the PL spectra. The exciton emission energy increased with a decrease in well thickness due to the quantum confinement effect. The linewidths of the excitonic emissions are around 35 meV for quantum wells of 2–6 ML, while that of 1 ML is 25 meV. In contrast to the ZnSe/ZnS QD case, narrowing in emission linewidth is remarkable for 1 ML CdSe QW. This narrowing in emission linewidth becomes more remarkable for submonolayer QWs. [Figure 17\(b\)](#) shows PL spectra from a sample consisting of four ZnSe/CdSe submonolayer SQWs [44]. The schematic of the sample structure is shown in the upper part of the figure. The deposition times for CdSe to form quantum wells were 1, 2, 3 and 4 s, which correspond to 1/4, 1/2, 3/4 and 1 ML of CdSe, respectively. The thickness of the ZnSe barrier layers was 0.1 μm, which is enough to separate the wells from each other. The PL spectrum shows four dominant free exciton emissions from the four separated submonolayer SQWs. A blue-shift in emission energy with decrease in CdSe layer thickness is observed.

The linewidth of the emission indeed decreases from 15 to 5 meV as the CdSe layer decreases from 1 to 1/4 ML. The linewidth of the exciton emission is plotted against the CdSe well thickness in [figure 6.18](#) by open circles. The emissions from CdSe QWs with QW thickness from 2–6 ML show almost the same linewidth of 35 meV, while the linewidth shows pronounced narrowing for submonolayer SQWs. We note the observed narrowing with decreasing the QW thickness in the emission spectra from common anion CdSe/ZnSe QWs is in

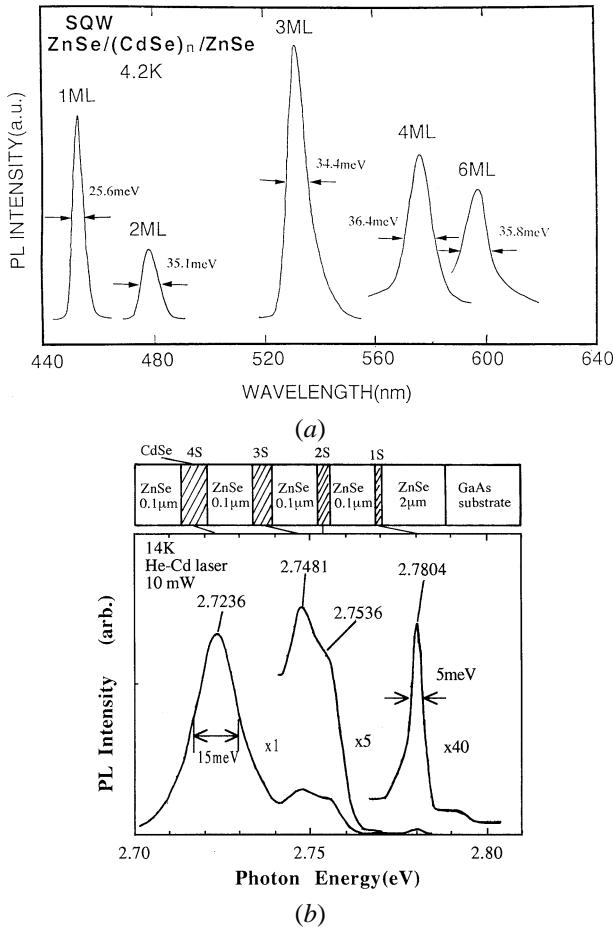


Figure 6.17. (a) PL spectra of ZnSe/(CdSe)_n/ZnSe ($n = 1\text{--}6$) SQWs at 4.2 K. (b) PL spectra of ZnSe/CdSe submonolayer SQWs. The schematic of sample structure is shown in the upper part of the figure. Four peaks of exciton emission from the four submonolayer SQWs appear [44].

contrast to the common anion ZnSe/ZnS QW case, in which dramatic broadening occurs as discussed in the previous section. We note that this behaviour cannot be understood in terms of the probability distribution of lateral dimension of islands and valleys due to interface roughness, but these phenomena can be well interpreted in terms of interface alloying.

We assume the formation of the ZnCdSe layer at the heterointerface due to lateral diffusion of Zn and Cd atoms during growth. When CdSe is deposited on a ZnSe surface with a surface roughness of 1 ML, there would be formation of

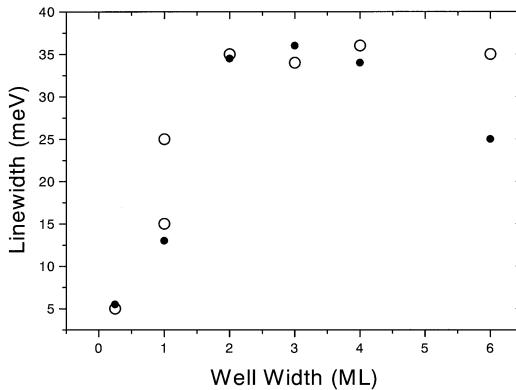


Figure 6.18. Plots of the observed linewidths (open circles) and calculated ones (full circles) of the exciton emission from ZnSe/CdSe QWs against the CdSe QW thickness.

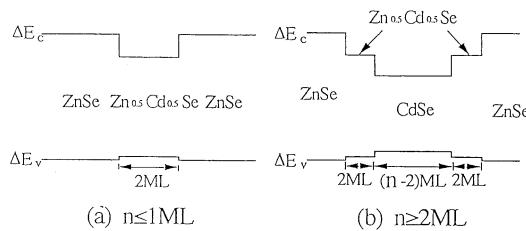


Figure 6.19. An interface model to analyse the observed PL properties of ultrathin ZnSe/CdSe QW structures assuming interface alloying. (a) The formation of $\text{Zn}_{0.5}\text{Cd}_{0.5}\text{Se}$ alloy at the interface is assumed for CdSe QW thicknesses smaller than 1 ML. (b) The formation of a 2 ML thick $\text{Zn}_{0.5}\text{Cd}_{0.5}\text{Se}$ layer on each interface of the QW is assumed in the case of CdSe QW thicker than 2 ML.

CdSe islands and valleys whose thickness are 1 ML. It is likely that $\text{Zn}_x\text{Cd}_{1-x}\text{Se}$ alloying arises within 2 ML thickness in a submolayer QW structure. Hence, it is assumed for simplicity that the well region of submonolayer QW forms a homogeneous $\text{Zn}_x\text{Cd}_{1-x}\text{Se}$ alloy with thickness of 2 ML in the zincblende crystal lattice. Figure 6.19 schematically shows the model of the band structure at the interface to explain the observed PL behaviours. In the case of submonolayer CdSe QW, we assume that the interface alloying ends up with the formation of homogeneous $\text{Zn}_x\text{Cd}_{1-x}\text{Se}$ alloy whose thickness is 2 ML, while the formation of 2 ML thick $\text{Zn}_{0.5}\text{Cd}_{0.5}\text{Se}$ interface layers is assumed in the case of CdSe QW thicker than 1 ML. The linewidth of exciton emission from $\text{ZnSe}/(\text{CdSe})_n/\text{ZnSe}$ SQW's can be estimated based on the formalism developed by Sing and Bajaj [46], where the effect of the probability distribution of Zn or Cd concentration in the QW is calculated. The inhomogeneous distribution of Zn or Cd concentration

in the well region gives rise to the line broadening of excitonic emission spectra. Based on the statistic thermodynamic consideration, the probability of Zn composition (C_{Zn}) is given by [46]

$$P(C_{\text{Zn}}) = \exp \left[\frac{R_{\text{ex}}^3}{r_c^3} \left\{ C_{\text{Zn}} \ln \left(\frac{C_{\text{Zn}}}{x} \right) + (1 - C_{\text{Zn}}) \ln \left(\frac{1 - C_{\text{Zn}}}{1 - x} \right) \right\} \right] \quad (6.9)$$

where R_{ex} , and r_c are the exciton radius, that of cation, and x the averaged Zn composition. The change Δ_{QW} in the excitonic emission energy due to alloy disorder is given by

$$\Delta_{\text{QW}} = \frac{dE^{\text{eh}}}{dC_{\text{Zn}}} \Big|_x \quad (6.10)$$

$$E^{\text{eh}} = E^{\text{e}} - E^{\text{h}} + E_g^{\text{w}} - E_b \quad (6.11)$$

where E^{e} and E^{h} are the electron and hole subband level in the quantum well, E_g^{w} is the bandgap of the well material, and E_b is the exciton binding energy. The linewidth is given from equations (6.9)–(6.11) as

$$\sigma_B^a = 2 \left\{ x(1-x)r_c^3 \frac{1.4}{R_{\text{ex}}^3} \right\}^{1/2} \Delta_{\text{QW}}. \quad (6.12)$$

Full circles in figure 6.18 show calculated linewidths of excitonic emission as a function of the well width based on the above-described procedure. Here we used 1.9 nm as the exciton radius (R_{ex}) of $\text{Zn}_x\text{Cd}_{1-x}\text{Se}$ in the best fit. As can be seen in figure 6.18, the calculated linewidth agrees well with those obtained experimentally at least for the CdSe QW thinner than 4 ML. In the above model, inhomogeneous alloying at the interface is not taken into account. This effect, however, would become serious, as the QW thickness exceeds the critical thickness for the generation of misfit dislocations, which would be the case for thicker QWs.

6.3.2 Self-organized formation of CdSe quantum dots

Self-organized formation of CdSe quantum dots on a ZnSe buffer layer has been achieved by the following two methods [47–51]: (1) growth by MBE at 300 °C, which is a higher temperature than that for the formation of QW with alloyed interface. At this temperature, deposition of CdSe more than the critical thickness leads to 3D QD's as indicated by spotty RHEED pattern; (2) growth by ALE at a lower temperature of 230 °C followed by subsequent annealing at higher temperature. The first technique leads to dot formation in most cases with, however, poor and inhomogeneously broadened PL spectra often of more than 100 meV FWHM. This large FWHM indicates a broad dot size distribution as well as inhomogeneous alloying at the CdSe/ZnSe interface. Figure 6.20

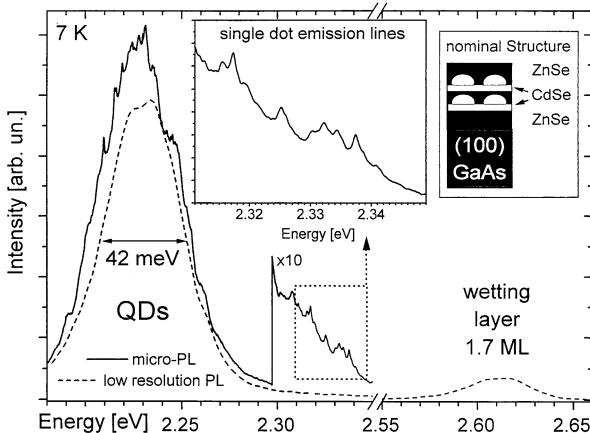


Figure 6.20. Micro-PL spectrum of a sample with a stacked double-layer QD grown by method (1). The sample structure is given in the inset. Distinct single dot emission lines can be observed. Space-integrated PL shows emission from the wetting layer at 2.62 eV corresponding to a 1.7 ML thickness [51].

shows the micro-PL spectrum of a sample consisting of two similar, subsequently stacked CdSe QD sheets with a 10 nm ZnSe spacer [51]. The low-energy region of the spectrum shows a large number of overlapping lines with a linewidth of 1 meV at 7 K, while those emission lines could not be observed under low-resolution measurements. The energy of the sharp peaks changes in a jumping way when the sample position is moved slowly [52]. This is strong evidence that the observed spectra originate truly from CdSe quantum dots. The line density observed for an illuminated area of about $1 \mu\text{m}$ in diameter suggests a rather high density of dots of the order of $100 (\mu\text{m})^{-2}$. Emission from the wetting layer is observed at 2.62 eV, which corresponds to a quantum well of 1.7 ML.

A detailed drawing of the second technique is given in figure 6.21 [51]. This method allows a considerably improved control over the dot formation, owing to the initial layer-by-layer deposition of the CdSe at a low temperature. After the growth of a ZnSe buffer layer at 300°C , the sample was cooled down to 230°C under a Se flux, and CdSe was deposited by ALE. Despite a slight decrease in intensity, no pronounced change in the RHEED pattern was observed even for a total deposition up to seven ALE (shutter) cycles which correspond approximately to 7 ML. Note that at a higher growth temperature of 300°C , one ALE-cycle corresponds to only about 0.5 ML. An obvious 3D transition is regularly observed when exceeding the critical thickness of CdSe (about 3.5 ML). However, TEM showed that generation of misfit dislocations is favoured over dot formation for strain relaxation. The ALE growth at 300°C can be employed to obtain QWs of high optical quality up to a thickness of about 4 MLs. The low-temperature

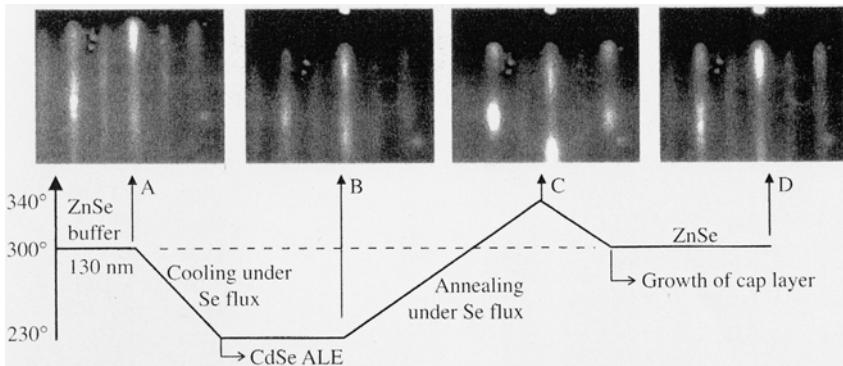


Figure 6.21. Growth cycle for the second growth method and corresponding RHEED evolution. A: after buffer layer deposition at 300 °C; B: after deposition of five ALE-cycles of CdSe at 230 °C; C: annealing at 340 °C. The 3D spotty pattern clearly demonstrates the reorganization of the CdSe layer. D: after growth of a 10 nm ZnSe cap at 300 °C. The RHEED recovers quickly to a 2D pattern indicating the embedding of the islands by ZnSe [51].

deposited CdSe layers are reorganized under annealing at temperatures above 320 °C, leading to a spotty RHEED pattern that indicates the island formation (figure 6.21(c)). Growing a subsequent cap layer of ZnSe (figure 6.21(d)), the RHEED recovers quickly to 2D (typically after the growth of 5–10 nm thick capping layer) indicating that the ZnSe fills in the gaps between the islands. Thus, the complete growth cycle can be repeated several times leading to a reproducible stacking of QD sheets.

Figure 6.22 shows the evolution of the position of the (0–1) RHEED rod with respect to the (00) rod as observed in the ⟨110⟩ direction during ALE deposition of CdSe and after annealing [51]. The rod position gradually approaches the position of the unstrained CdSe, demonstrating the gradual relaxation of the layer with further deposition. Interestingly, the strain relaxation on Cd and Se surfaces differs with more relaxation on Cd surfaces. After the annealing step, the relative position of the transmission spots almost match that of bulk CdSe, thus, almost a complete strain relaxation or even a slight over-relaxation takes place via the island formation.

6.3.3 Spectral diffusion in CdSe QDs

The spectral diffusion was first observed in single-molecular spectra [46]. The spectra diffusion and the on/off behaviour may be caused by changes of (1) temperature, (2) stress field, (3) magnetic field, (4) many-body effects, and/or (5) electric field. Similar phenomena have been reported in single CdSe nanocrystallites synthesized as colloids [53, 54]. The Stark effect has been

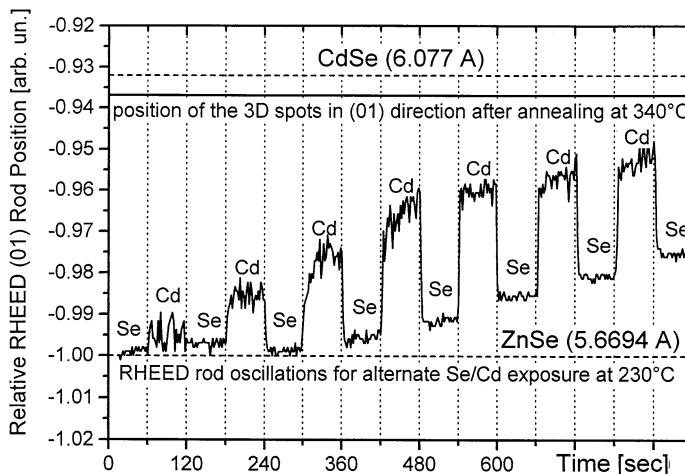


Figure 6.22. Evolution of the position of the (0–1) RHEED rod seen in the $\langle 110 \rangle$ direction during the deposition of CdSe by ALE. After the annealing process, the position is close to the one expected for bulk CdSe, indicating an almost complete relaxation of the layer through island formation [51].

proposed to be responsible to the spectral diffusion and on/off behaviour in the CdSe nanocrystallites [53]. If this is the case, the spectral diffusion should be commonly observed in any quantum dot system.

Figure 6.23 shows the evolution of 1 s integrated micro-PL spectra at the higher-energy side of the main emission peak from a CdSe QD sample at 7 K [52]. The spectra was taken every 15 s using the 476.5 nm line from an Ar^+ ion laser at an excitation power density of 5 kW cm^{-2} . The sharp peaks with the linewidth less than 1 meV is clearly observed. The peak becomes sharper, when the excitation intensity decreases. Both the energy and the intensity of the sharp peak show random drift with time. The spectral diffusion was observed even at 77 K with somewhat broadened spectral features and wider spectral diffusion range. Much quicker temporal variation in spectrum was observed at 77 K. Similar fluctuation in spectrum was reported on CdSe quantum dots synthesized as colloids, however with a larger spectral diffusion range than 10 meV [53], which should be compared with the present diffusing spectral range of 1 meV.

The spectral diffusion and the on/off behaviours of a single quantum dot have been qualitatively explained in terms of the Stark effect [53]. The Stark effect results from the Auger ionization process which causes the photo-darkening effect in microcrystallites [54]. When an electron (or hole) escapes from a quantum dot, photo-generated electrons and holes in the dot are influenced by the electric field created by the free electron in the vicinity of the dot. If the free electron (or hole) is recaptured by the dot, the electric field disappears. The migration of

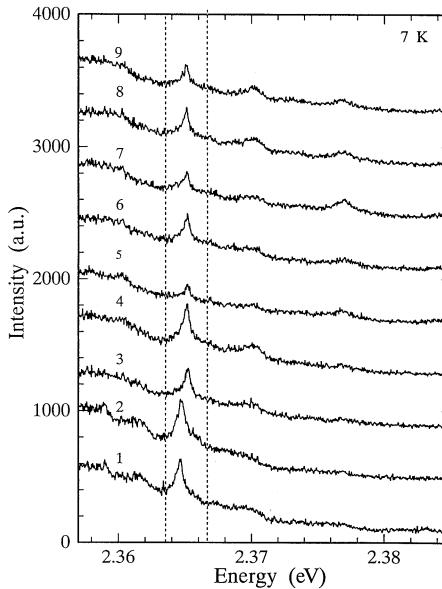


Figure 6.23. Evolution of 1 s integrated micro-PL spectra on the high-energy side of the main emission peak shown in [figure 6.20](#) [52].

the free electron (or hole) outside the ionized dot thus cause fluctuation in the energy state in the quantum dots. As temperature rises, the thermal energy of the free electron (or hole) exceeds the trapping energy of quantum dots. Then, the electron (or hole) migrates easily, and thus moves back and forth more quickly at higher temperature, which makes the spectral diffusion much faster at 77 K than 2 K as has been observed. In addition, the change of the local environment also becomes faster, as the temperature rises, which makes the sharp peak broader at higher temperature. The spectral diffusion in ZnSe/CdSe QD is found to be faster than that in CdSe nanocrystallites synthesized as colloids [53]. This is presumably ascribed to the difference in the matrix material: single crystal layer of ZnSe in the ZnSe/CdSe QD and organic material in the CdSe nanocrystallites. It is most likely that the free electrons outside the dot move much faster in the ZnSe single crystal layer than in the organic material. Although the spectral diffusion has not been reported in other quantum structures, this effect should be commonly observed in semiconductor QDs, if the Auger ionization is involved in carrier relaxation processes.

6.4 Summary

The spontaneous formation of II–VI widegap semiconductors have been reviewed with emphasis placed upon the interface morphology. It is found that a common cation quantum structure such as ZnS/ZnSe ultrathin QW's lead to an island interface which ensures localization at the interface thereby facilitating the growth of QDs at the heterointerface. Such quantum structures show a broad exciton luminescence band with developed low-energy tail, which is characteristic for a QD system with random distribution of lateral confinement. A lineshape analysis of the luminescence band suggests the formation of QDs with an average height of 3 ML and diameter of 2.9 nm. High-temperature growth is found to promote the self-organization of QDs. The temperature dependence of PL properties suggest the existence of a defect-related carrier loss mechanism which plays an important role in the PL quenching. The energy-dependent thermal activation energies are obtained, which clearly shows the zero-dimensional nature and size fluctuation in the ZnSe/ZnS QDs. The zero-dimensional nature of ZnSe dots are also supported by the presence of temperature insensitive exciton lifetime and slower decay time compared to a ZnSe QW sample. Mn is doped into ZnSe QDs to solve the phonon bottleneck problem, thereby enhancing the luminescence efficiency. It is found that the luminescence efficiency of each Mn atom increases three orders of magnitude in QD compared to Mn doped in bulk ZnSe.

In contrast to the common cation quantum structures, the common anion interface such as ZnSe/CdSe ultrathin QWs lead to interface alloying due to large interdiffusivity of metal atoms. PL properties of ZnSe/CdSe QWs show quite different behaviours in comparison with the ZnSe/ZnS system. The linewidth of exciton luminescence becomes narrower as the QW thickness decreases in the ZnSe/CdSe system, while a broader luminescence band with extended low-energy tail appeared as the ZnSe QW thickness enters into the submonlayer regime. The self-organized growth of CdSe QDs are achieved by Stranski–Krastanow growth mechanism either by MBE or ALE. Strain relaxation during the formation of QDs is confirmed by RHEED investigation. Micro-PL investigation reveals a number of sharp emission lines originating from single QDs of different sizes. It is found that these sharp lines show spectral diffusion presumably caused by the Stark effect.

Acknowledgments

This is a summary of a part of the works done in collaboration with Z Zhu, Y-h Wu, E Kurtz, K Arai, T Goto and M Y Shen. The author's sincere gratitude should be extended to all of the collaborators.

References

- [1] Bagnall D M, Chen Y F, Zhu Z, Yao T, Koyama S, Shen M Y and Goto T 1997 *Appl. Phys. Lett.* **70** 2230
- [2] Bagnall D M, Chen Y F, Zhu Z, Yao T and Goto T 1998 *Appl. Phys. Lett.* **73** 1038
- [3] Chen Y F, Bagnall D M, Ko H, Park K-t, Hiraga K, Zhu Z and Yao T 1998 *J. Appl. Phys.* **84** 3912
- [4] Bagnall D M, Chen Y F, Zhu Z, Yao T, Shen M Y and Goto T 1997 *Nonlinear Opt.* **18** 243
- [5] Takagahara T 1989 *Phys. Rev. B* **39** 10 206
- [6] Kummell T, Bacher G, Forchel A, Lermann G, Kiefer W, Jobst B, Hommel D and Landwehr G 1998 *Phys. Rev. B* **57** 15 439
- [7] Heiss W, Sitter D, Prechtel G, Sitter H, Liu J, Toth L and Barna A 1998 *Appl. Phys. Lett.* **72** 575
- [8] Yao T, Fujimoto M, Chang S K and Tanino H 1991 *J. Cryst. Growth* **111** 823
- [9] Arai K, Zhu Z Q, Sekiguchi T, Yasuda T, Lu F, Segawa Y, Kuroda N and Yao T 1997 *Nonlinear Opt.* **18** 307
- [10] Kurtz E, Jung H D, Hanada T, Zhu Z and Yao T 1998 *Appl. Surf. Sci.* **130-132** 755
- [11] Yao T 1986 *Opt.-Dev. Technol.* **6** 37
- [12] Bhargava R N 1997 *Properties of Wide Bandgap II-VI Semiconductors (EMIS Datereviews Series 17)* ed R N Bhargava, ch 1.5
- [13] Wu Y-h, Arai K and Yao T 1996 *Phys. Rev. B* **53** R10 485
- [14] Yao T and Wu Y-h 1997 *J. Korean Phys. Soc.* **31** 465
- [15] Delalande C, Meynadier M H and Voos M 1985 *Phys. Rev. B* **31** 2497
- [16] John G C and Singh V A 1994 *Phys. Rev. B* **50** 5329
- [17] Daly E M, Glynn T J, Lambkin J D, Considine L and Walsh S 1995 *Phys. Rev. B* **52** 3696
- [18] Cohen E and Sturge M 1982 *Phys. Rev. B* **25** 3828
- [19] Ouadjaout D and Marfaing Y 1990 *Phys. Rev. B* **41** 12 096
- [20] Kawakami Y, Taguchi T and Hiraki A 1988 *J. Cryst. Growth* **93** 714
- [21] Mohamed K, Olego D J, Newbury P, Cammack D A, Dalby R and Cornelissen H 1987 *Appl. Phys. Lett.* **50** 1820
- [22] Vening M *et al* 1993 *Phys. Rev. B* **48** 2412
- [23] Bacher G *et al* 1993 *Phys. Rev. B* **47** 9545
- [24] Weber S *et al* 1995 *Phys. Rev. B* **52** 14 739
- [25] Delalande C *et al* 1985 *Phys. Rev. B* **31** 2497
- [26] Gil B *et al* 1994 *Phys. Rev. B* **50** 18 231
- [27] Day E M *et al* 1995 *Phys. Rev. B* **52** 4696
- [28] Tournie E *et al* 1995 *Appl. Phys. Lett.* **67** 103
- [29] Kono T, Naganuma Y, Nishioka M and Arakawa Y 1995 *Superlatt. Microstruct.* **17** 73
- [30] Akiyama H, Koshiba S, Someya T, Wada K, Noge H, Nakamura Y, Inoshita T, Shimizu A and Sakaki H 1994 *Phys. Rev. Lett.* **72** 924
- [31] Shen M Y, Koyama S, Goto T, Arai K, Kasuya A, Wu Y-h and Yao T 1996 *Mater. Sci. Eng. A* **217/218** 189
- [32] Arai K, Zhu Z Q, Sekiguchi T, Yasuda T, Lu F, Kuroda N, Segawa Y and Yao T 1998 *J. Cryst. Growth* **184/185** 254
- [33] Masterov V F and Zakharenkov L F 1990 *Sov. Phys. Semicond.* **24** 383

- [34] Schmitt-Rink S, Varma C M and Levi A F J 1991 *Phys. Rev. Lett.* **66** 2782
- [35] Heimbrot W, Geode O, Weinhold V, Happ M, Knoch R, Hieke K, Hoffmann N, Griesche J, Jacos K, Neugebauer F, Suisky D and Roseler J 1994 *J. Cryst. Growth* **138** 601
- [36] Wu Y-h, Arai K, Kuroda N, Yao T, Yamamoto A, Shen M Y and Goto T 1997 *Japan. J. Appl. Phys.* **36** L1648
- [37] Hefets Y, Goltsos W T, Nurmiikko A V, Kolodziejski L K and Gunshor R L 1986 *Appl. Phys. Lett.* **48** 372
- [38] Benisty H 1995 *Phys. Rev. B* **51** 13 281
- [39] Takahara T 1985 *Phys. Rev. B* **31** 6552
- [40] Bhargava R N, Gallagher D, Hong X and Nurmiikko A 1994 *Phys. Rev. Lett.* **72** 416
- [41] Feldmann J, Peter G, Gobel E O, Dawson P, Moore K, Foxon C T and Elliott R J 1987 *Phys. Rev. Lett.* **59** 2337
- [42] Tenmyo
- [43] Zhu J L, Zhu S, Zhu Z, Kawazoe Y and Yao T 1998 *J. Phys.: Condens. Matter* **10** L583
- [44] Zhu Z, Yoshihara H, Takebayashi K and Yao T *Appl. Phys. Lett.*
- [45] Zhu Z, Yoshihara H, Takebayashi K and Yao T *J. Cryst. Growth.*
- [46] Singh J and Bajaj K K 1985 *J. Appl. Phys.* **57** 5433
- [47] Kurtz E, Jung H D, Hanada T, Zhu Z, Sekiguchi T and Yao T 1997 *Nonlinear Opt.* **18** 93
- [48] Zhu Z, Kurtz E, Arai K, Chen Y F, Bagnall D M, Tomashini P, Lu F, Sekiguchi T, Yao T, Yasuda T and Segawa Y 1997 *Phys. Status Solidi b* **202** 827
- [49] Kurtz E, Jung H D, Hanada T, Zhu Z and Yao T 1998 *Appl. Surf. Sci.* **130–132** 755
- [50] Kurtz E, Jung H D, Hanada T, Zhu Z, Sekiguchi T and Yao T 1998 *J. Cryst. Growth* **184/185** 242
- [51] Kurtz E, Sekiguchi T, Zhu Z, Yao T, Shen J X, Oka Y, Shen M Y and Goto T 1999 *Superlatt. Microstruct.* **25**
- [52] Shen M Y, Goto T, Kurtz E, Zhu Z and Yao T 1998 *J. Phys.: Condens. Matter* **10** L171
- [53] Empedocles S A, Norris D J and Bawendi M G 1996 *Phys. Rev. Lett.* **77** 3873
- [54] Nirmal M, Dabbousi B O, Bawendi M G, Macklin J J, Harris T D and Brus L E 1996 *Nature* **383** 802

PART 3

TRANSPORT PROPERTIES IN QUANTUM STRUCTURES

Chapter 7

Theory of quantum transport in mesoscopic systems: antidot lattices

Tsuneya Ando¹, Seiji Uryu¹, Takeshi Nakanishi² and Satoshi Ishizaka³

¹ *Institute for Solid State Physics, University of Tokyo, 7-22-1 Roppongi, Minato-ku, Tokyo 106-8666, Japan*

² *Institute of Physics and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-0198*

³ *NEC Fundamental Research Laboratories, 34 Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan*

A review of magnetotransport in antidot lattices mainly from a theoretical point of view is given. The topics include various mechanisms proposed for the explanation of the origin of the commensurability peaks such as pinned orbits, runaway orbits, and diffusive orbits combined with a magnetic focusing effect. Roles of quantum effects and antidot disorder, the difference between square and triangular lattices, and localization effects are also discussed.

7.1 Introduction

Recent advances in microfabrication and crystal growth technology have enabled the preparation of lateral superlattices with submicrometre structures on top of two-dimensional (2D) electron systems with mean free paths as large as 100 μm . The 2D system modulated by a periodic strong repulsive potential is called antidot lattices. The transport in this system is ballistic, i.e. electrons are scattered from the antidot potential itself rather than impurities. The purpose of this paper is to give a brief review on transport properties in antidot lattices in the presence of magnetic fields mainly from a theoretical point of view.

Various interesting phenomena have been observed in antidot lattices in magnetic fields. They are quenching of the Hall effect [1, 2], Altshuler–Aronov–Spivak oscillation near vanishing field [3–6], so-called commensurability peaks in the magnetoresistance [1, 7–16] and fine oscillations around them [17, 18].

Antidot lattices are introduced in section 7.2. In section 7.3 the origin of the commensurability peaks is discussed with emphasis on roles of classical chaotic motion. In section 7.4 the Aharonov–Bohm-type oscillation superimposed on the commensurability peak is analysed based on semiclassical quantization of periodic orbits existing in the chaotic sea. Triangular lattices are discussed in section 7.5 with the emphasis on differences from square lattices. In section 7.6 a kind of the Altshuler–Aronov–Spivak (AAS) oscillation appearing in weak magnetic fields is discussed in relation to roles of inherent disorder in the antidot potential arising during the fabrication process. A new method based on a scattering matrix to calculate electronic states and transport properties is introduced in section 7.7, together with some examples of the results obtained. The electron localization is discussed in section 7.8. A summary and conclusion are given in section 7.9.

7.2 Antidot lattices

Figure 7.1 shows an illustration of the potential of a square antidot lattice. It is characterized by the period a and the diameter d of the antidot. We have typically $d \gtrsim 1000$ Å, larger than the Fermi wavelength of the 2D electron system, $\lambda_F \sim 500$ Å for a typical electron concentration less than 5×10^{11} cm $^{-2}$ in a GaAs/AlGaAs heterostructure. Correspondingly, the period of the antidot is usually $a \geq 2000$ Å. This means that the system is in the boundary between quantum and classical regimes. In classical mechanics the electron motion is fully chaotic and in fact the antidot lattice is a kind of a Sinai billiard, which is a typical system where the chaotic motion has been intensively studied.

The model antidot potential used often for a square antidot is given by

$$U(\mathbf{r}) = U_0 \left| \cos\left(\frac{\pi x}{a}\right) \cos\left(\frac{\pi y}{a}\right) \right|^{2\beta} \quad (7.1)$$

where U_0 is a potential maximum and the parameter β characterizes its steepness. The antidot diameter can be defined as $d/2 = |\mathbf{r}|$ with $U(\mathbf{r}) = E_F$, where E_F is the Fermi energy and \mathbf{r} is chosen in the direction of a nearest-neighbour. In a hexagonal lattice, the corresponding potential is given by

$$U(\mathbf{r}) = U_0 \left| \cos\left(\frac{\pi \mathbf{a}_1 \cdot \mathbf{r}}{a^2}\right) \cos\left(\frac{\pi \mathbf{a}_2 \cdot \mathbf{r}}{a^2}\right) \cos\left(\frac{\pi(\mathbf{a}_1 - \mathbf{a}_2) \cdot \mathbf{r}}{a^2}\right) \right|^{4\beta/3}, \quad (7.2)$$

with $\mathbf{a}_1 = (\sqrt{3}a/2, a/2)$ and $\mathbf{a}_2 = (0, a)$. Sometimes, we use the following

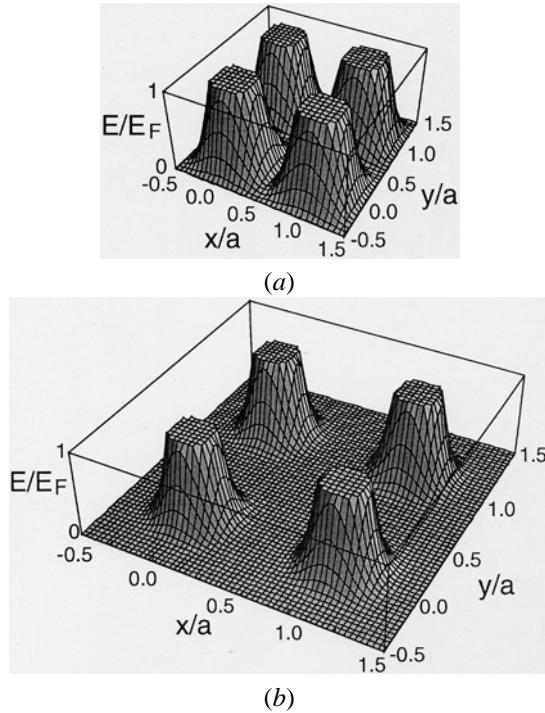


Figure 7.1. A model antidot potential in a square lattice. (a) $d/a = 0.5$ and $\beta = 1$. (b) $d/a = 0.3$ and $\beta = 4$. The potential is cut off at the Fermi energy.

model potential

$$U(\mathbf{r}) = E_F \left(\frac{d/2 + \Delta - r}{\Delta} \right)^2 \theta(d/2 + \Delta - r) \quad (7.3)$$

where r is the distance from the centre of an antidot, Δ is the parameter characterizing the steepness of the potential, and $\theta(t)$ is the step function defined by $\theta(t) = 1$ for $t > 0$ and 0 for $t < 0$.

A self-consistent calculation in quantum wires at GaAs/AlGaAs heterostructures [19, 20] suggests that the potential is nearly parabolic for a wire with small width and consists of a flat central region and a parabolic increase near the edge for a wider wire. The width Δ of the region where the potential increases from the bottom to the Fermi energy is of the same order as the Fermi wavelength λ_F for typical electron concentrations. This leads to roughly $\beta \sim 1$ for $d/a \sim 0.5$, $\beta \sim 2$ for $d/a \sim 0.4$, and $\beta \sim 4$ for $d/a \sim 0.3$ in the case $d \sim 1000$ Å.

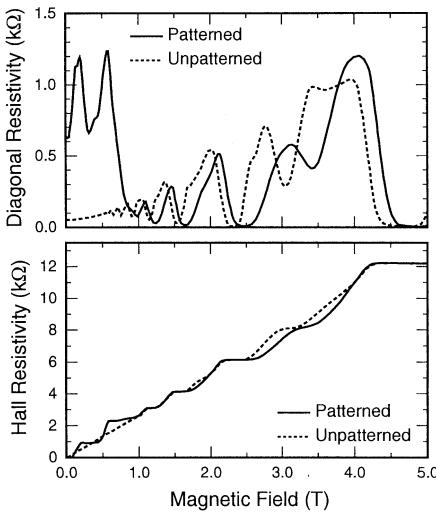


Figure 7.2. An example of ρ_{xx} and ρ_{xy} observed in patterned (full curves) and unpatterned (dotted curves) [1]. Two prominent peaks and step structures appear in the diagonal and Hall resistivity, respectively, in antidot lattices.

7.3 Commensurability peaks

Figure 7.2 shows an example of observed diagonal and Hall resistivity in a square antidot lattice [1]. Two prominent peaks appear in the diagonal resistivity ρ_{zz} in weak magnetic fields and correspondingly some steplike structures are present for the off-diagonal Hall resistivity ρ_{xy} . Cyclotron orbits can be classified into those affected and not affected by antidots [1]. At certain magnetic fields electrons can move on the commensurate classical orbit encircling a specific number of antidots as shown in figure 7.3 (pinned orbits). The magnetoresistance was expected to increase at this magnetic field.

On the other hand, a numerical simulation based on the classical electron motion showed that the change in the volume of the pinned orbits in the phase space is not enough to cause the commensurability oscillation and suggested that chaotic orbits which have a character close to the periodic orbits give larger contribution [21]. The importance of the ‘runaway’ orbit, which skips regularly from an antidot to its neighbouring antidot in the same direction was also proposed [22]. Experiments on rectangular and/or disordered antidots [23–27] and numerical simulations [27–29] provided evidence showing the importance of such orbits.

A better way of understanding the origin of the commensurability peaks is through diffusive orbits and magnetic focusing. Consider first the limit of small aspect ratio $d/a \ll 1$. In this case, the electron loses its previous memory of the

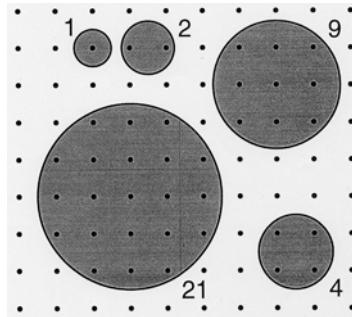


Figure 7.3. Some examples of pinned orbits. It was suggested that the resistivity exhibits a peak at the magnetic fields where the classical cyclotron orbit encircles 1, 2, 4, 9, 21, ... antidots.

direction of the velocity when it collides with an antidot, and therefore successive scattering with antidots can be approximately regarded as independent of each other [30]. This means that antidots are nothing but independent scatterers.

In magnetic fields, transport is possible through the migration of the centre of the cyclotron motion and therefore the conductivity vanishes in the absence of scattering. When $2R_c < a$, the scattering of an electron from an antidot cannot give rise to diffusion or conduction because the electron is trapped by the antidot.

Scattering from antidots starts to contribute to the conductivity when $2R_c > a - d$. The migration of the centre of the cyclotron orbit occurs most frequently due to successive scattering from nearest-neighbour antidots at the magnetic field corresponding to $2R_c = a$. At this magnetic field the measure of such orbits becomes maximum in the phase space due to a kind of magnetic focusing effect. In fact, a slight change θ in the direction from the normal to the line connecting neighbouring antidot leads to a change only of the order of $2R_c\theta^2$ in position when the electron collides with a neighbouring antidot as is shown in figure 7.4. This leads to an increase of the phase-space volume of the orbits contributing to the increase of the diffusion coefficient at $2R_c \approx a$. This magnetic focusing is different from mechanisms suggested in other works [25, 27, 31].

As shown in figure 7.5, the orbit corresponding to $2R_c = a$ can be denoted as $(n_x, n_y) = (\pm 1, 0)$ or $(0, \pm 1)$, where the line segment connecting the point (n_x, n_y) and the origin constitutes the diameter of the circle. With a further decrease in the magnetic field, successive scattering with next nearest-neighbour antidots becomes possible and the conductivity has a peak around $2R_c = \sqrt{2}a$ corresponding to $(\pm 1, \pm 1)$. This contribution becomes less prominent, however, because the orbit passes through the position of a nearest-neighbour antidot. The next peak arises from $(\pm 2, 0)$ or $(0, \pm 2)$ and $(\pm 2, \pm 1)$ or $(\pm 1, \pm 2)$, which lie close to each other. The latter contribution should be larger because its measure is twice as large as that of the former and therefore the peak occurs roughly around

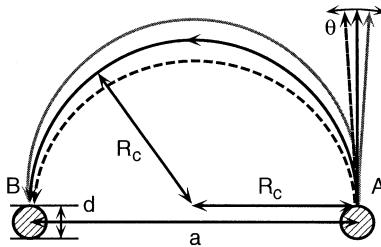


Figure 7.4. Schematic illustration of magnetic focusing leading to the fundamental commensurability peak.

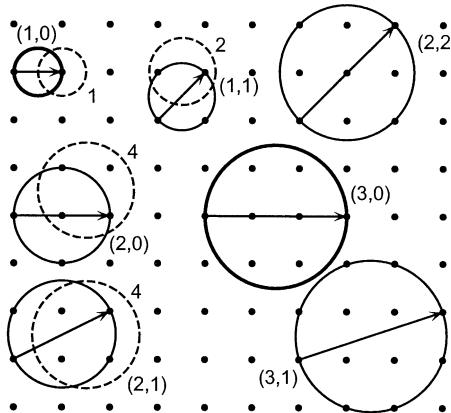


Figure 7.5. Some examples of diffusive cyclotron orbits contributing to the migration of the guiding centre. The broken circles correspond to pinned orbits (the number denotes antidots in the orbit).

$2R_c \sim \sqrt{5}a$. This peak should be weaker, however, as these orbits also pass through the position of other antidots. The next prominent peak is expected to be given by $(\pm 2, \pm 2)$, $(\pm 3, 0)$, $(0, \pm 3)$, $(\pm 3, \pm 1)$, and $(\pm 1, \pm 3)$. The peak is expected to be at $2R_c \approx 3a$, because the orbits $(\pm 3, 0)$ and $(0, \pm 3)$ are not disturbed by other antidots. This explains most of the commensurability peaks observed experimentally.

When the diameter d is no longer negligible in comparison with the period a , the commensurability peak in the diagonal conductivity σ_{xx} deviates from the condition $2R_c = (n_x^2 + n_y^2)^{1/2}a$. In the following, we confine ourselves to the fundamental peak given by $(\pm 1, 0)$ and $(0, \pm 1)$. A cyclotron orbit starting at an antidot and colliding with a neighbouring antidot has a centre in the region schematically illustrated in figure 7.6, to which the phase-space volume of such orbits is proportional. Figure 7.7 shows the area as a function of the magnetic field. The area has a maximum at $2R_c = a + fd$ with $f \sim 0.3$ being a constant

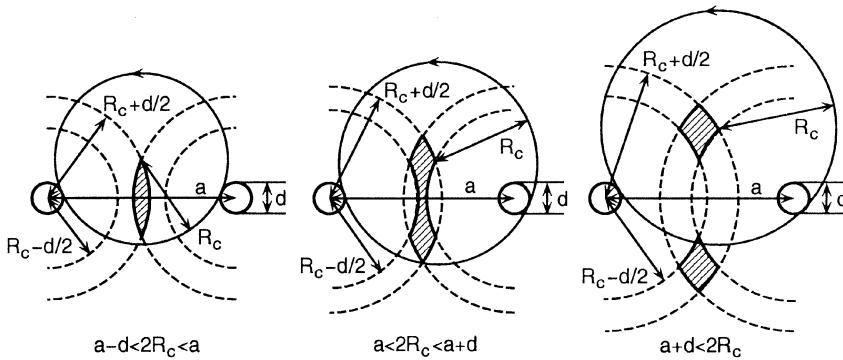


Figure 7.6. The region (shaded) of the centre of the cyclotron orbits passing through two nearest-neighbour antidots.

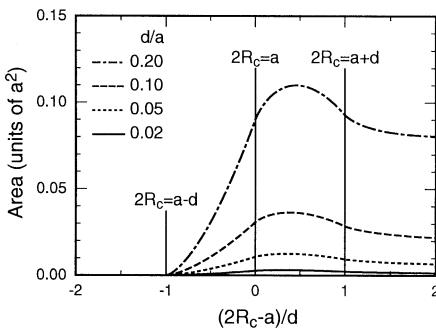


Figure 7.7. The area of the region where cyclotron orbits passing through two nearest-neighbour antidots. The diagonal conductivity is roughly proportional to the area at $2R_c \sim a$.

weakly dependent on d/a . This means that the commensurability peak is shifted to the weak-field side roughly in proportion to d/a .

For a non-negligible d/a , the correlation among successive scattering from antidots becomes important. In fact, it manifests itself most directly in the Hall conductivity. As illustrated in figure 7.8, it tends to enhance the left circular motion of electrons in the weak-field region $2R_c \gtrsim a$ and reduce it in the high-field region $2R_c \lesssim a$. This leads to an enhancement of $|\sigma_{xy}|$ for $2R_c \lesssim a$ and a reduction for $2R_c \gtrsim a$, giving rise to a steplike structure around $2R_c \approx a$.

With the increase of d/a , the measure of the diffusive orbits increases as shown in figure 7.7, and the diagonal conductivity becomes larger and comparable to σ_{xy} (numerical calculations show that $\sigma_{xx} \sim |\sigma_{xy}|$ around $d/a \sim 0.5$ [30]). Therefore, the commensurability peaks in the resistivity become different from those in the diagonal conductivity. Let $\Delta\sigma_{xx}$ and $\Delta\sigma_{xy}$ be a small change around

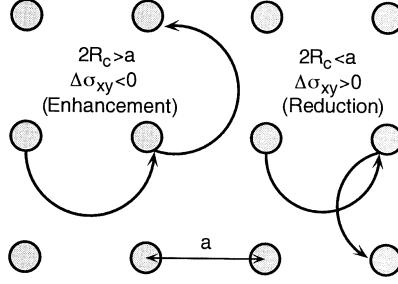


Figure 7.8. A schematic illustration of scattering from two adjacent antidots. This leads to an enhancement of $|\sigma_{xy}|$ for $2R_c \gtrsim a$ and a reduction for $2R_c \lesssim a$.

the field corresponding to the commensurability peak and $\bar{\sigma}_{xx}$ and $\bar{\sigma}_{xy}$ be the average ($\Delta\sigma_{xx} \ll \bar{\sigma}_{xx}$, $|\Delta\sigma_{xy}| \ll |\bar{\sigma}_{xy}|$). Then, we have

$$\Delta\rho_{xx} = \frac{(\bar{\sigma}_{xy}^2 - \bar{\sigma}_{xx}^2)\Delta\sigma_{xx} - 2\bar{\sigma}_{xx}\bar{\sigma}_{xy}\Delta\sigma_{xy}}{(\bar{\sigma}_{xx}^2 + \bar{\sigma}_{xy}^2)^2}. \quad (7.4)$$

This means that $\Delta\sigma_{xy}$ determines the structure of ρ_{xx} when $\bar{\sigma}_{xx} \sim |\bar{\sigma}_{xy}|$, i.e. the reduction of $|\sigma_{xy}|$ in the high-field side gives the peak in ρ_{xx} for large d/a .

Figure 7.9 shows an example of the results of recent numerical simulations in classical mechanics [30]. It shows that the peak in σ_{xx} is broadened considerably with the increase of d/a and at the same time shifted to the lower magnetic field side. A step structure corresponding to the first derivative of σ_{xx} with the magnetic field appears in σ_{xy} . It is interesting, however, that the commensurability peak in the resistivity ρ_{xx} remains around $2R_c = a$ independent of d/a . For $d/a \geq 0.5$, the structures in σ_{xx} and σ_{xy} disappear almost completely and ρ_{xx} does not show a clear commensurability peak.

A large aspect ratio usually corresponds to a small antidot period a and to a narrow spacing between the nearest-neighbour antidot comparable to the Fermi wavelength λ_F . Therefore, quantum effects can be important for antidot lattices with large d/a . In fact, a quantum mechanical calculation [32] in a self-consistent Born approximation (SCBA) [33, 34] shown in figure 7.10 gave the result for $d/a = 0.5$ that the diagonal conductivity σ_{xx} exhibits essentially no structure and the off-diagonal Hall conductivity has a small dip at $2R_c \sim a$, leading to the fundamental commensurability peak at $2R_c \sim a$. In this calculation use has been made of $\beta = 2$ which is larger than more realistic $\beta = 1$. Some experiments have shown that σ_{xx} even has a dip at $2R_c \sim a$ for antidot lattices with larger d/a [27, 35]. The structure appearing in the off-diagonal Hall resistivity at the fundamental commensurability peak was analysed experimentally and theoretically [36].

In high magnetic fields, the structure of the energy bands is relatively simple and can be understood as combinations of Landau bands which have non-zero

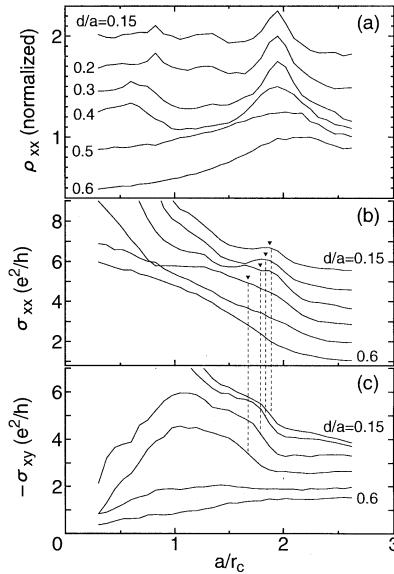


Figure 7.9. Some examples of (a) ρ_{xx} , (b) σ_{xx} and (c) σ_{xy} calculated in classical mechanics. Each curve of ρ_{xx} is normalized by its maximum value and shifted successively. The results for σ_{xx} are shifted successively by e^2/h . The peak positions of σ_{xx} are indicated by triangles and vertical dotted curves. The mean free path is $\Lambda = 3.3a$. After [30].

dispersion in the presence of antidot lattices and quantized edge states localized along an antidot boundary. With the decrease of the magnetic field, these magnetic minibands start getting mixed with each other. This mixing gives rise to many bands having a large quantized value of the Hall conductivity, both negative and positive [37]. Extremely large values are usually cancelled out if they are summed over adjacent minibands. This cancellation is not complete and the Hall conductivity tends to have a dip around this magnetic field. Therefore, the Hall conductivity is sensitive to parameters such as the Fermi energy and sometimes exhibits even a signature change.

7.4 Aharonov–Bohm-type oscillation

A fine oscillation was observed superimposed on commensurability peaks of the magnetoresistance [5, 17, 18]. Figures 7.11 and 7.12 give examples of such Aharonov–Bohm (AB)-type oscillation reported in [17, 18]. The period is roughly given by $\Delta B \sim \Phi_0/a^2$ with $\Phi_0 = ch/e$ being the magnetic flux quantum. Numerical calculations have been performed on this AB-type oscillation [32, 38, 39]. In the following some of the results [32, 40] obtained

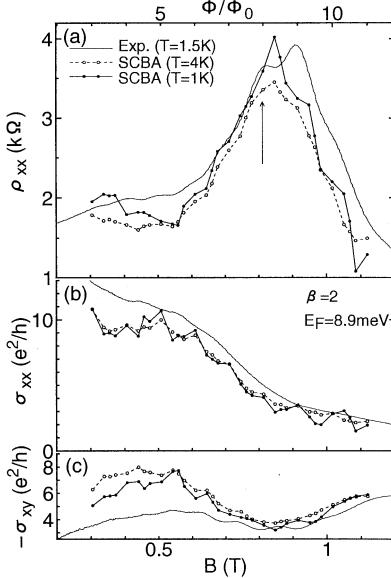


Figure 7.10. (a) ρ_{xx} , (b) σ_{xx} and (c) σ_{xy} calculated in SCBA as a function of the magnetic field for $\beta = 2$. The experimental data [17] are also shown for comparison. After [32].

in SCBA are reviewed briefly.

The calculated diagonal and Hall conductivity at $T = 4\text{ K}$ shown in figure 10 oscillate with an exact period of Φ_0 over the entire range of magnetic fields considered. This oscillation originates from Hofstadter's butterfly band structure [41] as has been suggested theoretically [42]. At a lower temperature ($T = 1\text{ K}$), another kind of oscillation appears in addition to this butterfly oscillation. The difference between this new oscillation and the butterfly oscillation becomes much clearer for the more realistic parameter $\beta = 1$.

The density of states in classically chaotic systems is given semiclassically in the so-called periodic orbit theory [44–46]. According to the trace formula, the density of states is given by the sum of a classical contribution and a quantum correction. The latter is given by semiclassically quantized energy levels associated with periodic orbits and the former is proportional to the classical phase-space volume at an energy. The contribution of a particular periodic orbit can be written as the summation of δ functions for a stable orbit,

$$D_{\text{osc}}^{\text{po}}(E) = \frac{T}{\hbar} \sum_{m=0}^{\infty} \left[\sum_{n=-\infty}^{\infty} \delta \left(\frac{S_c}{\hbar} - \left(m + \frac{1}{2} \right) v - 2n\pi \right) - \frac{1}{2\pi} \right] \quad (7.5)$$

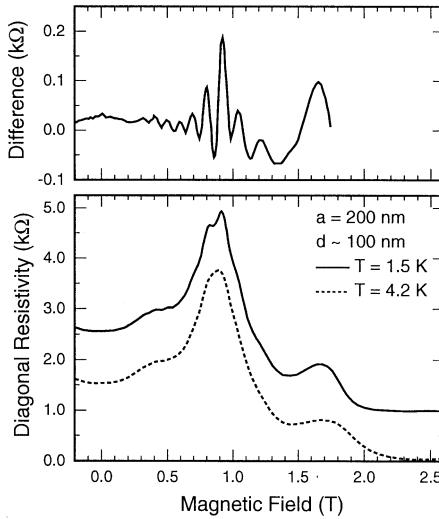


Figure 7.11. An example of observed quantum oscillations superimposed on the fundamental commensurability peak [17]. The oscillation becomes clear in the difference of the resistivity at $T = 4.2$ K and that at $T = 1.5$ K (shifted vertically by $1\text{ k}\Omega$).

and the summation of Lorentzian functions for an unstable orbit,

$$D_{\text{osc}}^{\text{po}}(E) = \frac{T}{\hbar} \sum_{m=0}^{\infty} (\pm 1)^m \times \left[\sum_{n=-\infty}^{\infty} \frac{1}{\pi} \frac{(2m+1)u/2}{[S_c/\hbar - 2\pi(\mu/4+n)]^2 + [(2m+1)u/2]^2} - \frac{1}{2\pi} \right], \quad (7.6)$$

where S_c is the classical action, T is the period and v is the stability angle. For unstable orbits, u is the Lyapunov exponent, μ is the Maslov index, and the sign becomes \pm according as the neighbourhood of a periodic orbit is hyperbolic or inverse hyperbolic. The integer m is considered as a quantum number for a motion transverse to the periodic orbit [47]. The quantization conditions are

$$S_c = \begin{cases} 2\pi\hbar \left[n + \frac{1}{2\pi} \left(m + \frac{1}{2} \right) v \right] & \text{for stable orbit,} \\ 2\pi\hbar \left(n + \frac{\mu}{4} \right) & \text{for unstable orbit,} \end{cases} \quad (7.7)$$

with integer n and non-negative integer m .

Figure 7.13 shows some examples of periodic orbits having a short trajectory [43, 48]. The orbit (a) is localized in the potential-well region surrounded by

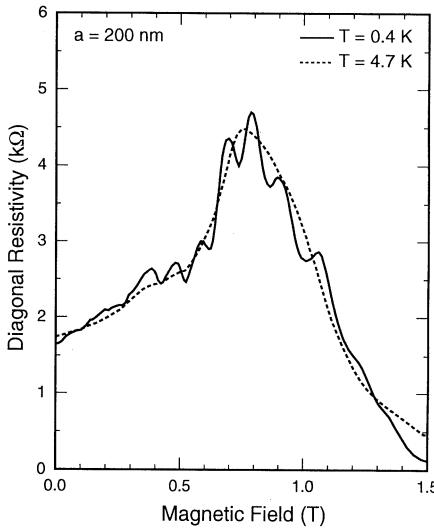


Figure 7.12. An example of observed quantum oscillations superimposed on the fundamental commensurability peak [18].

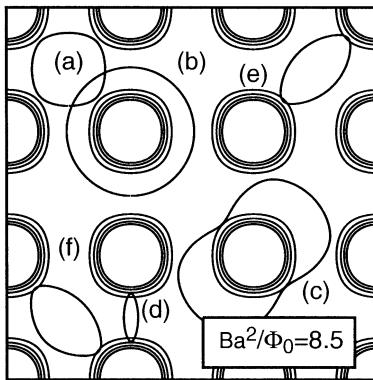


Figure 7.13. Some examples of periodic orbits having a short trajectory in a square antidot in magnetic field $Ba^2/\Phi_0 = 8.5$ roughly corresponding to the main commensurability peak. The parameters are $a/\lambda_F = 4.31$, $U_0/E_F = 8.4$, $d/a = 0.6$ and $\beta = 2$. After [43].

four neighbouring antidots. It turns continuously into a normal cyclotron orbit in sufficiently high magnetic fields. The orbit (b) circling around an antidot is considered to be most important in giving rise to the quantum oscillation.

Figure 7.14 shows an example of the calculated density of states together with quantized energy levels associated with the orbits (a) and (b) for $m = 0$ [32, 40]. The density of states is correlated well with the semiclassical energy

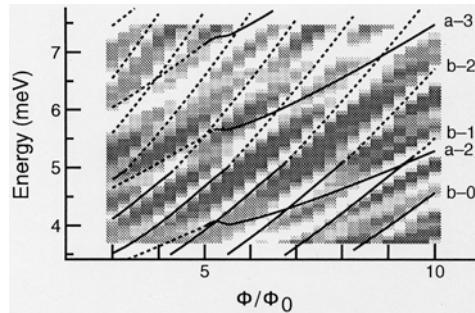


Figure 7.14. The density of states calculated in SCBA for $\hbar/\tau = 0.18$ meV and $\beta = 1$. Its values are large in the light region and small in the dark region. The energy levels for the orbits (a) and (b) calculated semiclassically with the use of the trace formula for $m = 0$ are given by curves (full curves for stable orbits and dotted curves for unstable orbits). After [40].

levels, although many features remain to be explained.

A semiclassical expression can be obtained for the conductivity tensor [32, 49, 50], but has proved to be unsuccessful for the parameter range corresponding to the experiments mentioned above [32]. As discussed in the previous section, the diffusive orbits have a dominant contribution to the transport in the magnetic-field region near the commensurability peak. These orbits are perturbed by the presence of quantized periodic orbits, which is likely to lead to a quantum oscillation. The semiclassical expression for the conductivity does not take into account such effects.

7.5 Triangular antidot lattices

Commensurability peaks have also been observed in hexagonal antidot lattices. Their origin is the same as in square lattices and is understood in terms of diffusive orbits and magnetic focusing. Figure 7.15 shows some examples of diffusive orbits contributing to migration of guiding centres of cyclotron orbits. They are specified by two integers (n_1, n_2) such that $n_1\mathbf{a}_1 + n_2\mathbf{a}_2$ corresponds to a vector whose length is equal to the cyclotron diameter, i.e. $2R_c/a = \sqrt{n_1^2 + n_2^2 + n_1 n_2}$.

There are some important differences from the square lattice. First, diffusive orbits are strongly disturbed by the presence of other antidots with increasing antidot diameter. As schematically illustrated in figure 7.16, the diffusive orbits start to be severely influenced for $d/a > (\sqrt{3} - 1)/2 \approx 0.37$ in the hexagonal case in contrast to $d/a > 2/3 \approx 0.67$ in the square case. Therefore, the commensurability peaks are much more sensitive to the aspect ratio.

Second, there is a wide gap in the magnetic field between the fundamental peak given by $a/R_c = 2[(n_1, n_2) = (\pm 1, 0) \text{ or } (0, \pm 1)]$ and the next peak given

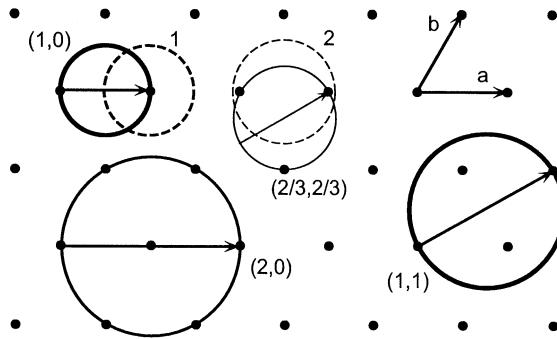


Figure 7.15. Some examples of diffusive cyclotron orbits contributing to the migration of the guiding centre in a hexagonal lattice. The broken circles correspond to pinned orbits (the number denotes antidots in the orbit). The measure of the diffusive orbits has a dip around the field corresponding to $(n_1, n_2) = (2/3, 2/3)$.

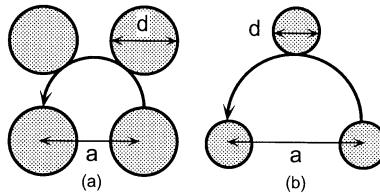


Figure 7.16. Schematic illustration of diffusive orbits giving rise to a fundamental commensurability peak in (a) square and (b) hexagonal lattices. The orbit starts to be disturbed by other antidots for smaller d/a in the hexagonal case.

by $a/R_c = 2/\sqrt{3}[(n_1, n_2) = (\pm 1, \pm 1)]$. As is shown in figure 7.17, the area of the region corresponding to a guiding centre of a cyclotron orbit passing through a pair of nearest-neighbour antidots starts to overlap with that of another pair at a magnetic field given by the condition $R_c + d/2 = a/\sqrt{3}$. This leads to a reduction in the measure of the diffusive orbits around $a/R_c \sim \sqrt{3}$, giving rise to a small dip in the diffusion coefficient at this magnetic field.

Figure 7.18 shows some examples of results of quantum and classical calculations [37, 51]. The commensurability peaks are rapidly shifted to the lower magnetic field side with increasing d/a , which is more appreciable in quantum results. In fact, the peak is shifted down to the vanishing magnetic field and therefore only a large negative magnetoresistance can be seen for $d/a \geq 0.5$. These results can be well understood in terms of the differences between the square and hexagonal lattices mentioned above (a weak dip may be identified in the resistivity at the magnetic field given by $a/R_c = \sqrt{3}$ in the resistivity for $d/a = 0.2$).

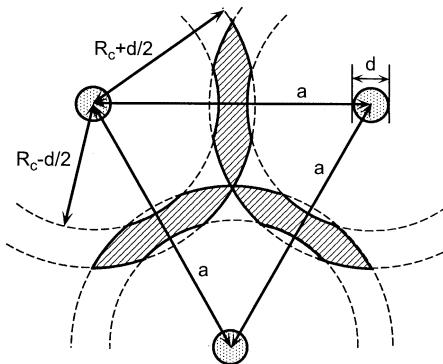


Figure 7.17. The region (shaded) of the guiding centre of the cyclotron orbits passing through two nearest-neighbour antidots in a hexagonal lattice when $R_c + d/2 \approx a/\sqrt{3}$.

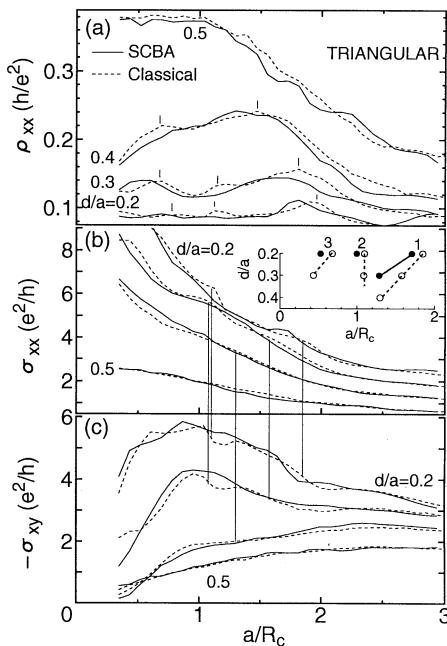


Figure 7.18. Calculated resistivity and conductivities in hexagonal antidot lattices. Quantum and classical results are shown by full and dotted curves, respectively. The peak positions of σ_{xx} (full and open circles for quantum and classical, respectively) are shown in the inset. After [37].

The strong disturbance of the diffusive orbits in hexagonal lattices with a large aspect ratio suggests that the number of events that an electron is scattered by the same antidot is larger than that in square lattices. This means that

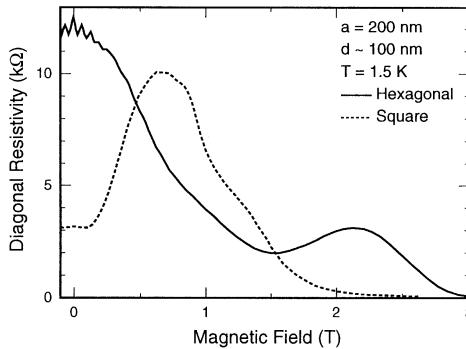


Figure 7.19. Comparison of the resistivity in a hexagonal (full curve) and square (dotted curve) antidot observed experimentally. A prominent AAS oscillation with period determined by $\Phi_0/2$ appears in the hexagonal lattice.

the electron trajectory is much more complicated in hexagonal lattices than in square lattices, explaining the reason that the conductivity itself is considerably smaller in hexagonal lattices and that the amplitude of the AAS oscillation is much larger as will be discussed in section 7.6 [52, 53]. This also explains the similar behaviour of the AAS oscillation of the localization length which was recently observed experimentally [54] and also demonstrated theoretically as will be discussed in section 7.8 [55, 56].

7.6 Altshuler–Aronov–Spivak oscillation

The conductance of a conductor in the form of a hollow cylinder oscillates as a function of the magnetic flux passing through its cross section with period given by half of the flux quantum $\Phi_0 = ch/e$. This oscillation is a result of interference effects and is called the Altshuler–Aronov–Spivak (AAS) oscillation [57]. This AAS oscillation was experimentally observed in cylindrical metal films [58] and networks of thin metallic wires [59] in the diffusive regime where the mean free path is much smaller than the sample size or the wire width.

AAS oscillations have been observed also in antidot lattices [3–6]. Figure 7.19 shows an example [6]. The essential features of the experiments can be summarized as follows: (a) The oscillations are observed clearly in triangular lattices but are almost invisible in square lattices [5, 6]. (b) The negative magnetoresistance is observed together with the AAS oscillation. In the following, a numerical study [52, 53] is reviewed briefly.

Real antidot lattices have inherent randomness arising from fabrication processes in addition to impurities whose strength is characterized by the mean free path Λ . This disorder is introduced through fluctuations in antidot diameters around the mean value d . The amount of fluctuations is characterized by the root-

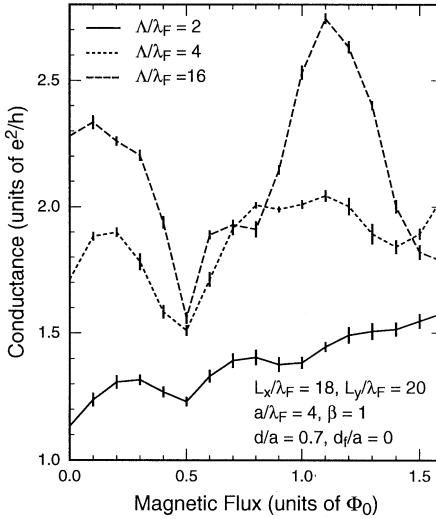


Figure 7.20. Conductance of a triangular antidot lattice as a function of magnetic flux passing through a unit cell for different values of the mean free path. The system size is given by $L_x \times L_y$. After [52, 53].

mean-square deviation d_f . The conductance is calculated by a Green function technique [60] for systems with length L_x and width L_y containing a finite number of antidots (~ 25).

Figure 7.20 gives calculated conductance of a triangular antidot lattice with $a/\lambda_F = 4$ in the absence of disorder in the antidot potential. In the ballistic regime $\Lambda/\lambda_F = 16$, where Λ is larger than the antidot period, the conductance exhibits an irregular oscillation. This is a consequence of the fact that an irregular oscillation with period Φ_0 due to interference effects has a dominant contribution. In fact, both phase and form of the oscillation vary sensitively depending on parameters like d/a , a/λ_F and β . With the decrease in the mean free path, the amplitude of the irregular oscillation becomes smaller and that of an AAS oscillation characterized by period $\Phi_0/2$ becomes more and more dominant. In particular, a beautiful AAS oscillation manifests itself in the diffusive regime where $\Lambda/\lambda_F = 2$ or $\Lambda/a = 1/2$.

Similarly, the contribution of the irregular oscillation diminishes with the increase of disorder in the antidot potential, as shown in figure 7.21. The oscillation with period $\Phi_0/2$ becomes dominant for disorder as large as $d_f/a = 0.07$. Figure 7.21 shows that the AAS oscillation is accompanied by a weak negative magnetoresistance or a slight increase in the conductance as a function of a magnetic field. This is in agreement with one of the features of the experiments discussed above. The amplitude of the AAS oscillations saturates for $d_f/a > 0.07$.

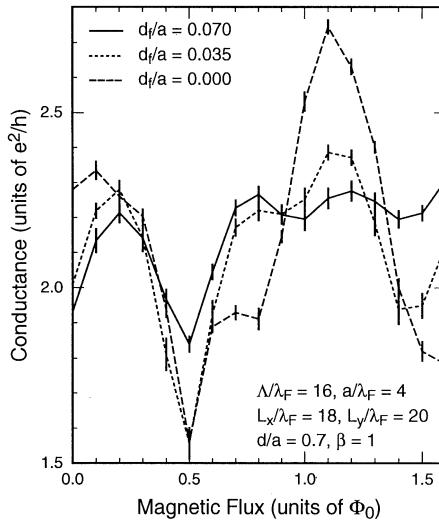


Figure 7.21. Conductance of a triangular antidot lattice in the presence of fluctuations in antidot diameter. After [52, 53].

In the square lattice the AAS oscillation has a smaller amplitude than that in the triangular lattice, again in excellent agreement with experiments though not shown here explicitly. Note that the conductance of the triangular lattice is much smaller than that of the square lattice. Electrons perform more complicated motion and stay longer in the triangular lattice than in the square lattice [5, 6]. This is presumably the main origin of the enhancement of the oscillation in the triangular lattice as has been discussed in this section.

7.7 Scattering matrix formalism

7.7.1 Quantum-wire junction

Difficulties of quantum mechanical calculations of energy levels and transport coefficients in antidot lattices lie in the fact that we have to consider a magnetic unit cell which can become very large depending on the magnetic flux passing through a unit cell. In this section a new method of full quantum calculations of energy bands and transport quantities [61, 62] is reviewed.

We shall replace an antidot lattice by a 2D array of quantum-wire junctions with width W_x in the x direction and W_y in the y direction, as illustrated in figure 7.22. Each junction is connected to neighbouring junctions through a quantum wire with length L_x and L_y in the x and y directions, respectively. The lattice period in the x direction is a and that in the y direction is b . For the calculation of an S matrix, we consider the model shown in figure 7.23. We

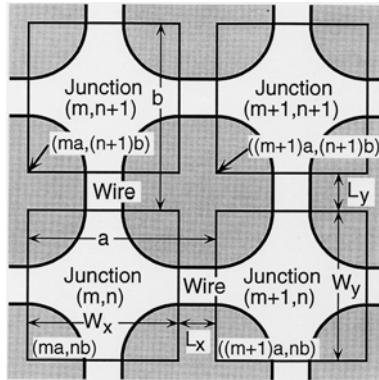


Figure 7.22. Two-dimensional array of quantum wire junctions.

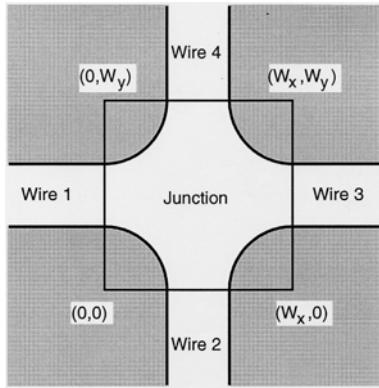


Figure 7.23. A quantum wire junction for calculation of an S matrix.

divide this system into five regions, i.e. four wires and a junction.

The vector potential is chosen as $(-By, 0)$. The modes in the wires 1–4 are given by

$$\begin{aligned} \psi_{1v}^{\pm}(x, y) &= (v_v^x)^{-1/2} \exp[i(\pm\kappa_v^x + W_y/2l^2)(x + L_x/2)]\eta_{v\pm}(y), \\ \psi_{2v}^{\pm}(x, y) &= (v_v^y)^{-1/2} \exp[i(\pm\kappa_v^y - W_x/2l^2)(y + L_y/2)] \\ &\quad \times \exp(-ixy/l^2)\xi_{v\pm}(x), \\ \psi_{3v}^{\pm}(x, y) &= (v_v^x)^{-1/2} \exp[i(\pm\kappa_v^x + W_y/2l^2)(x - W_x - L_x/2)]\eta_{v\pm}(y), \\ \psi_{4v}^{\pm}(x, y) &= (v_v^y)^{-1/2} \exp[i(\pm\kappa_v^y - W_x/2l^2)(y - W_y - L_y/2)] \\ &\quad \times \exp(-ixy/l^2)\xi_{v\pm}(x), \end{aligned} \quad (7.8)$$

where $l = \sqrt{eB/c\hbar}$ is the magnetic length, v denotes modes, v_v corresponds to

velocities, $\eta(y)$ is an eigenfunction for a wire in the x direction and $\xi(x)$ is that for a wire in the y direction for the vector potential $(0, Bx)$.

There are two kinds of modes. One of them is a travelling mode whose wavenumber κ is real and the other is an evanescent mode whose wavenumber has an imaginary part. The sign $+(-)$ denotes a mode travelling or decaying in the positive (negative) direction. The wavefunction in a wire region is represented using expansion coefficients C with respect to modes in the wires.

$$\Psi_i(x, y) = \sum_v [C_{iv}^+ \psi_{iv}^+(x, y) + C_{iv}^- \psi_{iv}^-(x, y)]. \quad (7.9)$$

The S matrix gives the following relation between incoming and outgoing waves:

$$\begin{pmatrix} \mathbf{C}_1^- \\ \mathbf{C}_2^- \\ \mathbf{C}_3^+ \\ \mathbf{C}_4^+ \end{pmatrix} = S(E) \begin{pmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_2^+ \\ \mathbf{C}_3^- \\ \mathbf{C}_4^- \end{pmatrix}, \quad (7.10)$$

where \mathbf{C} consists of expansion coefficients C and therefore its dimension is the number of modes.

7.7.2 Energy bands and density of states

We can calculate energy bands of an infinitely large array of junctions using Bloch's theorem. In the following, we confine ourselves to the case in the absence of the magnetic field for simplicity. The extension to the case in magnetic fields is discussed in [62]. The Bloch condition for $\Psi_i(x, y)$ gives the conditions for expansion coefficients, $C_{1v} = \exp(-ik_x a)C_{3v}$ and $C_{2v} = \exp(-ik_y b)C_{4v}$. These are rewritten as the following relation between incoming and outgoing waves.

$$\begin{pmatrix} \mathbf{C}_1^- \\ \mathbf{C}_2^- \\ \mathbf{C}_3^+ \\ \mathbf{C}_4^+ \end{pmatrix} = P(k_x, k_y) \begin{pmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_2^+ \\ \mathbf{C}_3^- \\ \mathbf{C}_4^- \end{pmatrix}, \quad (7.11)$$

where

$$P(k_x, k_y) = \begin{pmatrix} 0 & 0 & P_x^{-1} & 0 \\ 0 & 0 & 0 & P_y^{-1} \\ P_x & 0 & 0 & 0 \\ 0 & P_y & 0 & 0 \end{pmatrix}, \quad (7.12)$$

with $P_x = \exp(ik_x a)$ and $P_y = \exp(ik_y b)$. Energy bands are given by the condition that (7.10) and (7.11) have non-trivial solutions for \mathbf{C} :

$$\det[P(k_x, k_y) - S(E)] = 0. \quad (7.13)$$

This equation gives the energy bands as equi-energy lines in \mathbf{k} space at energy E .

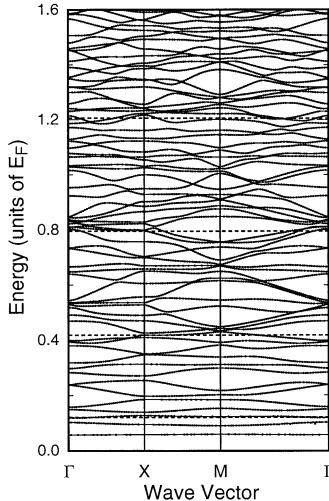


Figure 7.24. Calculated energy bands in a magnetic field $Ba^2/\Phi_0 = 4$. In the S matrix method, all travelling modes and only a single evanescent mode are taken into account. After [62].

Figure 7.24 shows some examples of calculated energy bands for a square lattice ($b = a$). The parameters are $a/\lambda_F = 4.31$, $U_0/E_F = 8.4$, $d/a = 0.6$, and $\beta = 2$ (this choice of β is likely to be too large for realistic antidots). The corresponding equipotential lines are given in [figure 7.13](#). The magnetic field $Ba^2/\Phi_0 = 4$ corresponds to about 0.4 T for a typical antidot lattice with $a \sim 2000$ Å. The bands are calculated by including all travelling modes and a single evanescent mode having the smallest imaginary wavevector. The horizontal dotted curves denote the bottom of one-dimensional subbands in the wire region.

The results using the S matrix formalism with only a single evanescent mode agree quite well with the exact results in which all the evanescent modes are included. This shows that the inclusion of a few evanescent modes is sufficient for an accurate description of energy bands of the antidot lattices. The bands for $Ba^2/\Phi_0 = 4$ corresponding to $2R_c/a \sim 2$ may be regarded as a mixture of a band with a large dispersion (runaway orbit) and those having a small dispersion for $0.4 \lesssim E/E_F \lesssim 1.0$ [32]. The energy bands become more complicated for higher energies.

[Figure 7.25](#) shows the corresponding energy dependence of the density of states and the conductivity calculated using the Boltzmann transport equation. The density of states is measured in units of the 2D value $m^*/\pi\hbar^2$ and the conductivity in units of $ne^2\tau/m^*$ where n is the electron concentration in the absence of the antidot potential and m^* is the effective mass. They are averaged over the Gaussian distribution with broadening $\Gamma/E_F = 0.02$. Some structures of

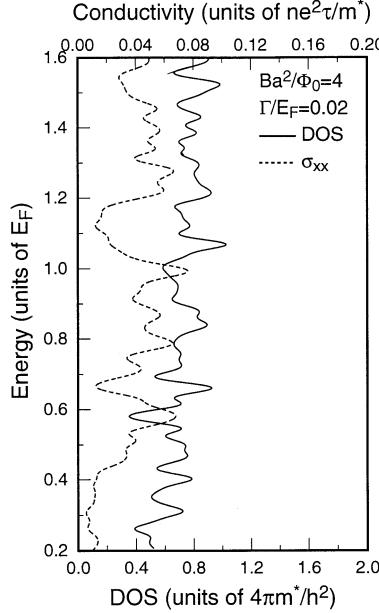


Figure 7.25. The energy dependence of the density of states (full curve) and the conductivity (dotted curve) in the magnetic field $Ba^2/\Phi_0 = 4$. After [62].

the conductivity correspond to those of the density of states, but most of them do not.

A semiclassical analysis has been made for the density of states by considering six periodic orbits (a)-(f) given in figure 7.13 [43, 62, 63]. As a cutoff of a quantum number m in equation (7.5), the factor $\exp(-\alpha m)$ has been introduced, where m is the quantum number for a motion transverse to each periodic orbit. Figure 7.26 shows some examples in the magnetic field corresponding to $Ba^2/\Phi_0 = 8.5$, where the condition of the fundamental commensurability peak is nearly satisfied, i.e. $2R_c \sim a$ at the Fermi level.

The orbit (a) is stable in the energy region $E/E_F \leq 0.43$, the orbit (b) $0.72 \leq E/E_F \leq 1.41$ and $E/E_F \leq 0.52$, orbit (c) $0.7 \leq E/E_F \leq 0.74$, and orbit (e) $E/E_F \leq 0.68$. Orbit (c) merges into orbit (b) and disappears at $E/E_F = 0.7$. The quantized levels associated with such stable orbits give clear peaks in the semiclassical density of states, many of which are in agreement with those of the quantum-mechanical density of states. Particularly, near the Fermi energy, the orbit (b) circling around an antidot gives a major contribution in agreement with the analysis mentioned in section 7.4. Inclusion of higher transverse modes does not considerably affect the structures of the semiclassical density of states. However, peaks at $E/E_F = 0.97$ and 1.1 , which roughly correspond to peaks of quantum-mechanical density of states, is shifted to the higher energy side by

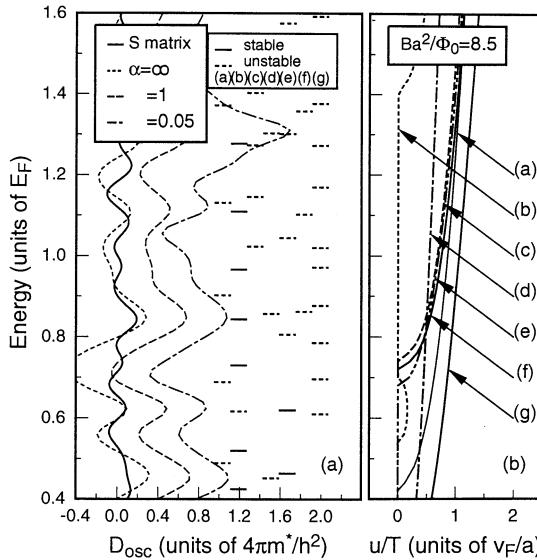


Figure 7.26. (a) The oscillatory part of the density of states at $Ba^2/\Phi_0 = 8.5$ corresponding to $2R_c \approx 2$. The broken curve is shifted to the right by 0.4 and the chain curve is shifted by 0.8. (b) The corresponding Lyapunov exponents. After [43].

including higher transverse modes and the agreement becomes worse. A more elaborate work in which many periodic orbits are taken into account is necessary for a full understanding of the correspondence between the quantum-mechanical and semiclassical results for the density of states.

7.7.3 Commensurability peaks and Aharonov–Bohm oscillation

In order to calculate transport quantities, we take a quantum wire junction array as a model of an antidot lattice as shown in figure 7.27 [39]. This is a kind of four-terminal system where constrictions at edges are connected respectively to four reservoirs through ideal leads. Transport coefficients are calculated based on the Büttiker–Landauer formula [66, 67] from an S matrix for a junction array obtained by connecting S matrices for a single junction.

In the following, square antidot lattices are considered and parameters are chosen as $d/a = 0.5$. The parameter Δ for the model potential (7.3) characterizing the steepness of the potential is chosen as $\Delta/a = 0.24$. This model potential becomes close to that given by (7.1) for $\beta = 1$ and $d/a \sim 0.5$. Effects of impurities are introduced through random short-range scatterers and we shall consider fluctuations in the diameter of antidots.

First, we consider the commensurability peak and the superimposed Aharonov–Bohm oscillation. Figure 7.28 shows the dependence of the calculated

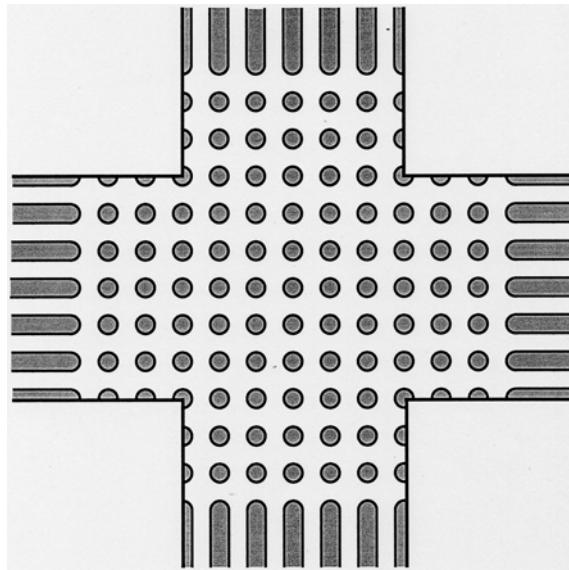


Figure 7.27. Schematic illustration of a finite antidot lattice for which the numerical calculation is performed. It has eight sites in the direction of the wire width and 16 sites in the direction of the wire length.

magnetoresistance on the Fermi wavelength λ_F for the mean free path $\Lambda/\pi a = 2$ in the absence of the antidot disorder [63]. The AB-type oscillation is clearly seen superimposed on the commensurability peak. The vertical full and broken curves indicate semiclassical quantized levels of the stable and unstable periodic orbits, respectively, encircling a dot, which are calculated based on equation (7.7). The AB-type oscillation varies with the Fermi energy in a somewhat irregular manner, but is correlated well with the semiclassical levels.

For the model potential (7.3), the periodic orbit encircling an antidot is quite sensitive to the aspect ratio d/a and disappears when d/a exceeds a critical value depending on the magnetic field. Figure 7.29 shows the boundary of the existence of the periodic orbit encircling an antidot for a model potential given by equation (7.3) [64]. The insets on the right-hand side of the figure show the periodic orbit (full curve) for $d/a = 0.6$ and 0.677 . With the increase of d/a it merges into another orbit (dotted curve), which is unstable and has a longer trajectory, and disappears. It contains the boundary in the case of $\Delta/a = 0.18$ and 0.36 as well. The critical aspect ratio becomes smaller with the increase of Δ/a , i.e. as the antidot potential becomes broader.

Figure 7.30 shows the calculated results of the magnetoresistance for various values of the aspect ratio in the case of dominant impurity scattering [64]. The AB-type oscillation changes as a function of the magnetic field in an irregular

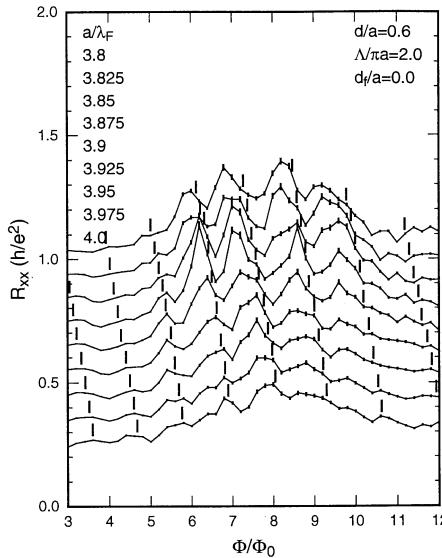


Figure 7.28. Diagonal resistance as a function of the magnetic field for $d/a = 0.6$ and the mean free path $\Lambda/\pi a = 2$. After [63].

manner, but the dependence on the field is, on the whole, in good agreement with that of the position of quantized levels of the periodic orbit. The oscillation disappears or becomes unrecognizable near the field corresponding to $2R_c = a$ for large aspect ratio $d/a = 0.7$. This field corresponds to the disappearance of the periodic orbit. In fact, the periodic orbit is absent in the hatched region, where no clear AB-type oscillation is visible. This provides a clear evidence that the AB-type oscillation is due to the specific periodic orbit (b) in figure 7.13 encircling an antidot.

Figure 7.31 shows the resistivity for the mean free path $\Lambda/\pi a = 2$ in the absence of the antidot disorder at several values of d/a and a/λ_F [65]. Both commensurability peak and AB-type oscillation vary in an irregular manner with the change in the Fermi energy and the aspect ratio. This can be understood in terms of the anomalous behaviour of the Hall conductivity due to the complicated band mixing as mentioned in section 7.3.

Figure 7.32 shows the resistivity in the presence of disorder $d_f/a = 0.07$ in the antidot diameter without impurities [65]. The irregular dependence of the commensurability peak and the AB-type oscillation on the Fermi energy and the aspect ratio has disappeared completely. In particular, the resistivity shows a smooth AB-type oscillation much closer to experiments. This suggests that such band structure is destroyed almost completely by the presence of the antidot disorder.

Many periodic orbits can contribute to the AB-type oscillation in ideal

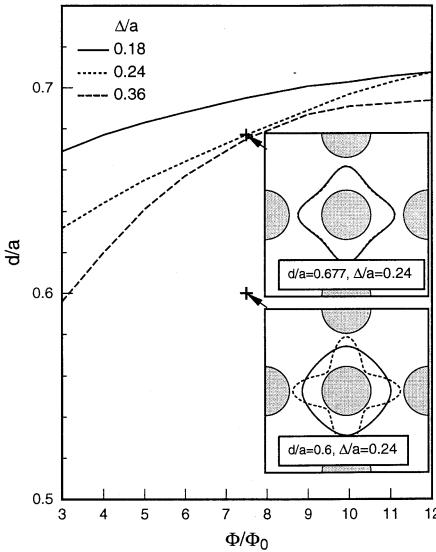


Figure 7.29. Critical values of the aspect ratio at which the periodic orbit encircling an antidot disappears. $\Delta/a = 0.18$ (full curve), 0.24 (dotted curve, realistic), and 0.36 (broken curve). The right upper and lower insets show the periodic orbit (full curve) giving the AB-type oscillation and the other (dotted curve) that is unstable. These orbits merge and disappear when the aspect ratio approaches a critical value. After [64].

antidot lattices. In the presence of disorder in the antidot potential, the shortest orbit encircling a single antidot is expected to be less influenced and its quantized energy levels can survive, while those of other orbits are strongly disturbed and smeared out. In other words, the antidot disorder tends to enhance the contribution of the energy levels of this orbit relative to those of other orbits. This explains the fact that the AB-type oscillation is sensitive to the Fermi energy for impurity scattering but not for scattering from antidot disorder.

7.8 Anderson localization

7.8.1 Experiments

One important feature of a hexagonal antidot lattice is the appearance of an AAS oscillation in weak magnetic fields at low temperatures (1.6 K) as has been discussed in section 7.6. Quite recently, a behaviour suggesting a strong localization was observed at lower temperatures [54]. Figure 7.33 shows an example of observed resistivity below 1.6 K. The resistivity increases considerably with decreasing temperature at low magnetic fields, while it increases at high magnetic fields. At a critical magnetic field, $B_c \approx 0.47$ T,

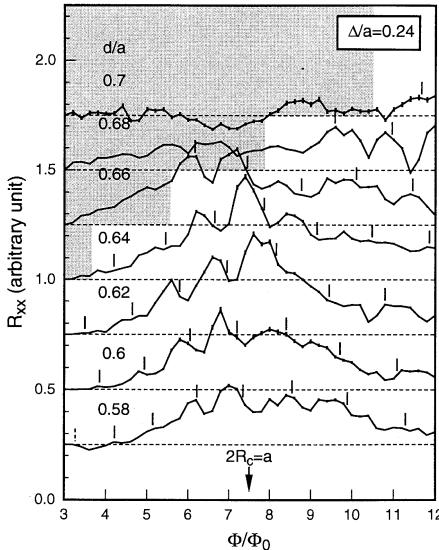


Figure 7.30. Calculated resistance as a function of the total flux Φ passing through a unit cell divided by the flux quantum Φ_0 for varying Fermi wavelength. Short vertical lines represent semiclassically quantized energy levels associated with the periodic orbit encircling an antidot (full and dotted curves for stable and unstable orbits, respectively). After [64].

the resistivity remains almost independent of the temperature and is nearly equal to h/e^2 .

Figure 7.34 shows that the resistivity for $B < B_c$ exhibits a temperature dependence consistent with variable range hopping below 1 K. Further, the diagonal conductivity also shows a temperature dependence reminiscent of variable range hopping for $B > B_c$. The localization length estimated from the fitting to variable range hopping turns out to be close to the lattice constant of the hexagonal lattice, showing that electrons may become completely confined in each ‘quantum dot’ surrounded by six antidots.

Another interesting feature of the results given in figure 7.33 is the fine quantum oscillation. At lower magnetic fields ($B < B_c$) the period is determined by half of the flux quantum per unit cell and by the flux quantum at higher magnetic fields ($B > B_c$). This leads to the conclusion that the localization length oscillates with a period given by half of the flux quantum at $B < B_c$ and by the flux quantum at $B > B_c$. The origin of the AAS oscillation is known to be quantum interferences of a path encircling an antidot with its time-reversal path in the weak localization regime. The AAS oscillation of the conductance alone does not mean the oscillation of the localization length itself, however.

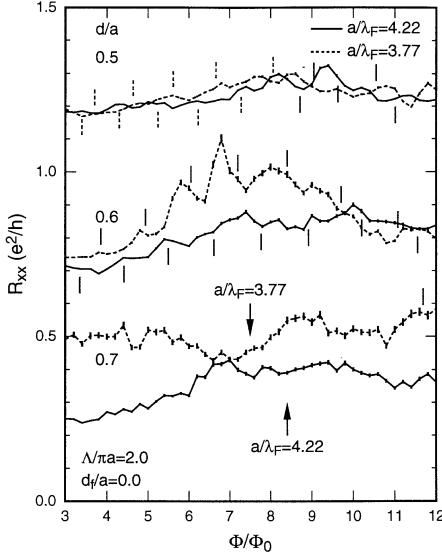


Figure 7.31. Calculated magnetoresistance in the case of dominant impurity scattering characterized by the mean free path Λ . The arrows indicate the field corresponding to $2R_c/a = 1$ for two different values of the Fermi wavelength λ_F . Short vertical lines represent semiclassically quantized energy levels associated with the periodic orbit encircling an antidot (full and dotted curves for stable and unstable orbits, respectively). After [65].

7.8.2 Thouless-number method

In order to study localization effects in antidot lattices, we shall use a conventional Thouless-number method [68, 69]. This method is known to be useful in magnetic fields [70]. According to a scaling argument [71], the Thouless number $g(L)$ is a dimensionless conductance defined by the ratio of the strength of effective coupling $V(L)$ and energy difference $W(L)$ of two systems with size L when the systems are combined.

The effective coupling strength is estimated from a geometric average of the curvature of energy bands $\partial^2 E / \partial k^2$ for periodic systems having a unit cell with size L and the energy difference is determined from the average energy spacing given by $D(E)^{-1}$ where $D(E)$ is the density of states at energy E . We then have

$$g(L) = \frac{V(L)}{W(L)} = \frac{\overline{\partial^2 E}}{\partial k^2} D(E). \quad (7.14)$$

The S matrix method [62] described in section 7.7 is ideal for this study because calculations in large-size systems are possible. For a hexagonal lattice we shall use quantum-wire junctions illustrated in figure 7.35, whose area is twice as

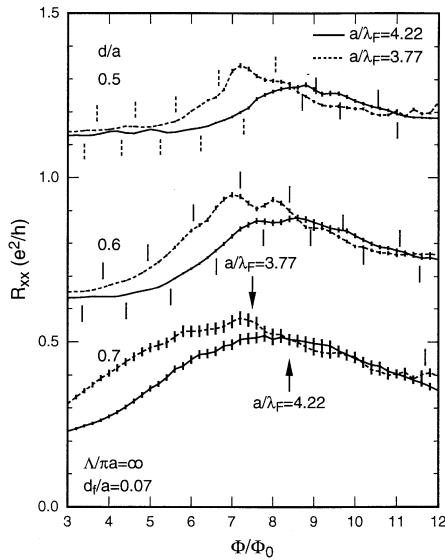


Figure 7.32. Calculated resistance in the case of dominant scattering with antidot disorder. After [65].

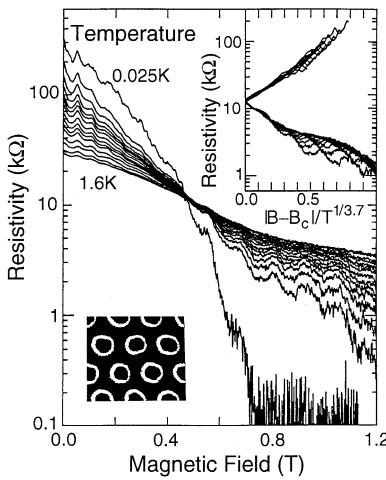


Figure 7.33. Resistivity as a function of a magnetic field in a hexagonal antidot lattice at low temperatures (0.025, 0.15, 0.2, 0.3, 0.4, 0.6, 1.0 and 1.6 K). After [54].

large as a unit cell. In actual numerical calculations, $g(L)$ is calculated for $2n \times n$ systems with $n = 2, 4, 6$ and 8. The inverse localization length α is determined

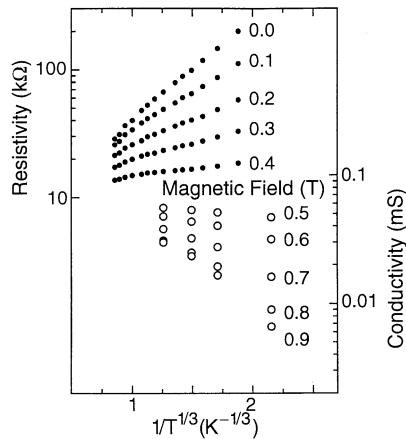


Figure 7.34. Temperature variation of the resistivity at lower magnetic fields (black dots) and diagonal conductivity at higher magnetic fields (open circles) can be both explained by variable range hopping. After [54].

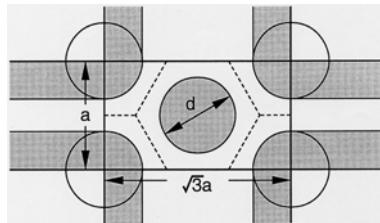


Figure 7.35. Schematic illustration of the quantum-wire junction for a hexagonal antidot lattice. Broken curves indicate boundaries of the Wigner–Seitz cell.

by fitting the results to

$$g(L) = g(0) \exp(-\alpha L), \quad (7.15)$$

with $L = 2na$, where a is a lattice constant.

We use the model potential given by equation (7.2) with $\beta' = 1$ and $a/\lambda_F = 3.77$. This corresponds to the electron concentration $n_s = 2.2 \times 10^{11} \text{ cm}^{-2}$ for $a = 2000 \text{ \AA}$ which is comparable to $n_s = 1.8 \times 10^{11} \text{ cm}^{-2}$ in the experiments [54]. We shall introduce a disorder in the antidot potential, i.e. fluctuations in the diameter, and use the mean free path $\Lambda/a = 4$ which is consistent with a real system with high mobility (but still much smaller) where the mean free path is much larger than the lattice constant.

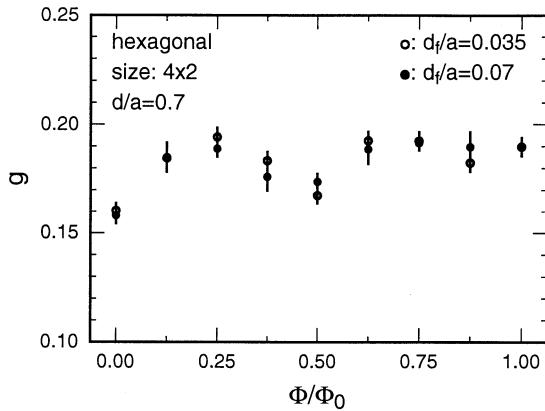


Figure 7.36. Calculated Thouless numbers as a function of a magnetic field in hexagonal antidot lattices with $d/a = 0.7$ and the $4a \times 2a$. After [56].

7.8.3 Numerical results

Figure 7.36 shows some examples of the calculated Thouless number g as a function of a magnetic field in hexagonal antidot lattices with size 4×2 ($L = 4a$). They show an AAS oscillation with the period $\Phi_0/2$ as a function of the total magnetic flux passing through a unit cell given by $\Phi = \sqrt{3}Ba^2/2$, where B is the strength of the magnetic field. The amplitude of the oscillation Δg normalized by the Thouless number at zero magnetic field g_0 is about $\Delta g/g_0 \sim 0.19$ almost independent of d_f . This is consistent with the results $\Delta g/g_0 \sim 0.15$ calculated for similar systems in a different method, i.e. through a direct calculation of the conductance from the transmission probability [52, 53], as discussed in section 7.6.

Figure 7.37 shows calculated results of (a) the Thouless number and (b) the inverse localization length for $d_f/a = 0.035$ in the absence of a magnetic field. It contains the energy of a one-dimensional subband in the quantum-wire region between nearest-neighbour antidots in the absence of the disorder. The figure shows that the localization is quite sensitive to the aspect ratio d/a . In fact, the localization length is reduced from $\alpha^{-1} \sim 50a$ for $d/a = 0.6$ to $\alpha^{-1} \sim 8a$ for $d/a = 0.8$.

Figure 7.38 shows the magnetic-field dependence of the Thouless number and the inverse localization length for $d/a = 0.7$ and $d_f/a = 0.035$. The Thouless number for each system size oscillates with the period of about $\Phi_0/2$ like an AAS oscillation. The localization length also exhibits a clear oscillation with the same period in good qualitative agreement with experiments [54]. The localization length oscillates in the range between $\sim 20a$ and $\sim 70a$.

Figure 7.39 shows the calculated results for $d/a = 0.8$ and $d_f/a = 0.035$. In this case, the channel width between neighbouring antidots is small and only

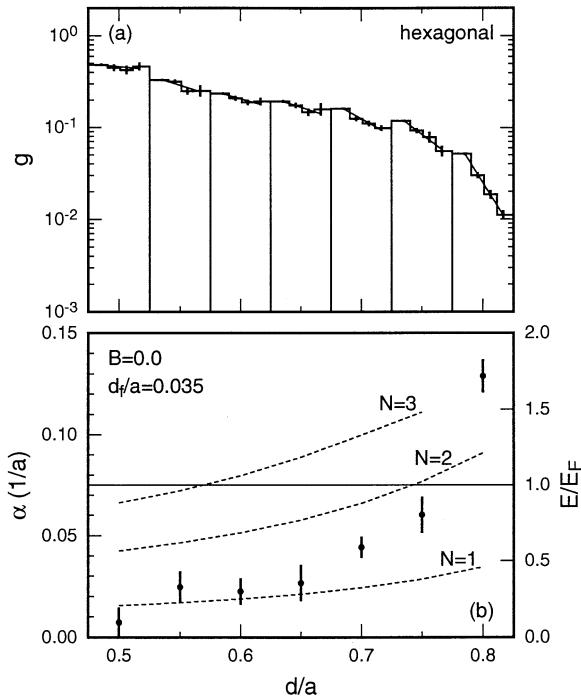


Figure 7.37. Calculated dependence of (a) the Thouless number g and (b) the inverse localization length α on the aspect ratio d/a in the absence of a magnetic field. In (a) the histograms in the region $0.475 < d/a < 0.525$, for example, indicate the Thouless number for a system with size $L = 4a, 8a, 12a$ and $16a$ from the left to right in antidot lattices with $d/a = 0.5$. In (b) the horizontal full line indicates the Fermi energy and the dotted curves show the energy of one-dimensional subbands in the ‘quantum wire’ between neighbouring antidots. After [56].

a single travelling channel is present (see figure 7.37(b)). The localization effect is enhanced considerably and the AAS oscillation of the localization length is reduced significantly. In fact, the localization length varies only in the range between $\sim 8a$ and $\sim 10a$ and the tendency corresponding to a negative magnetoresistance that the localization effect becomes weaker with the increase of the magnetic field can be more clearly identified.

7.8.4 Localization oscillation

The localization depends strongly on the symmetry of the system. In usual systems in the absence of a magnetic field, the Hamiltonian is chosen as a real symmetric matrix and the corresponding wavefunction is real except for an unimportant phase factor. In this orthogonal case the effective coupling $V(L)$ is

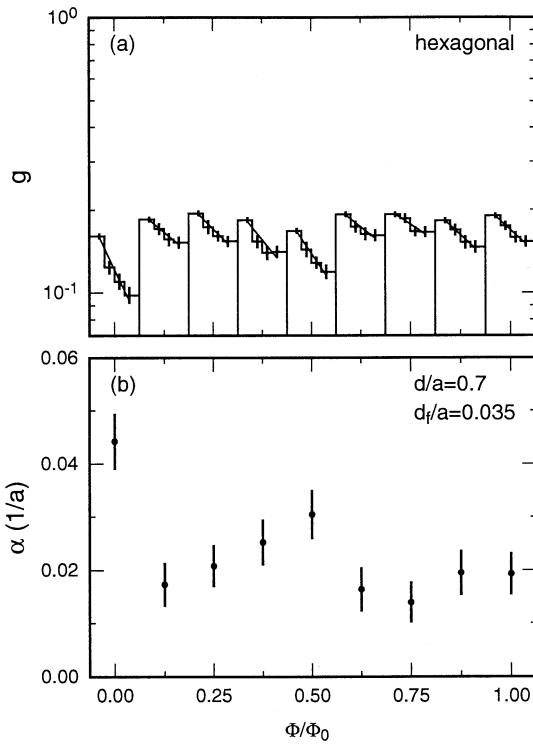


Figure 7.38. Magnetic field dependence of (a) the Thouless number and (b) the inverse localization length in a hexagonal lattice with $d/a = 0.7$ and $d_f/a = 0.035$. After [56].

given by a real number and an effective dimension of $V(L)$ is given by $\eta = 1$. In the presence of a magnetic field, the Hamiltonian is given by a complex unitary matrix and the corresponding wavefunction becomes complex. In this unitary case, $V(L)$ is given by a complex number and consequently $\eta = 2$. In the presence of a strong spin-orbit interaction in the absence of a magnetic field, states are always doubly degenerate (Kramers degeneracy) and $V(L)$ is given by a quaternion number. In this symplectic case we have $\eta = 4$. This η is the exponent characterizing level repulsion effects in the random-matrix theory [72].

The localization effect becomes weaker with the increase of η . In fact, the localization length in quasi-one-dimensional systems is known to be approximately given by $N_c \eta \Lambda$ [73–75], where N_c is the number of channels and Λ is the mean free path. The AAS oscillation of the conductivity is closely related to the fact that the system belongs to an orthogonal ensemble for $\Phi = n\Phi_0/2$ with an integer n and to a unitary ensemble in other cases.

The amplitude of localization-length oscillation for strong localization shown in figure 7.39 is much smaller than that for weaker localization shown in figure 7.38. This is consistent with the fact that the AAS oscillation requires

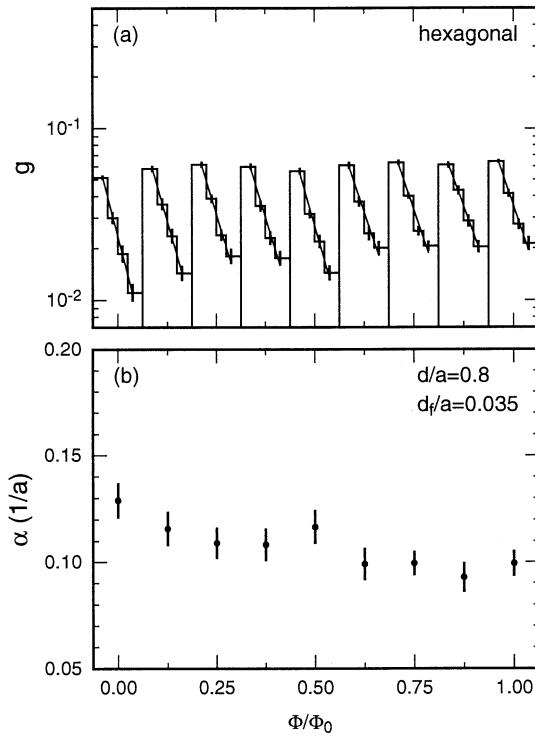


Figure 7.39. Magnetic field dependence of (a) the Thouless number and (b) the inverse localization length in a hexagonal lattice with $d/a = 0.8$ and $d_f/a = 0.035$. After [56].

interferences between the paths encircling an antidot in different directions. With the increase in the localization effect the probability amplitude for an electron to circle around an antidot is reduced.

According to our results on hexagonal antidot lattices with $d_f/a = 0.035$, the oscillation amplitude $\Delta\alpha$ of the inverse localization length normalized by that at zero magnetic field α_0 is $\Delta\alpha/\alpha_0 \sim 0.35$ for $d/a = 0.7$. This is in reasonable agreement with experimental results $\Delta\alpha/\alpha_0 \sim 0.20$ [54] mentioned above. However, the absolute magnitude of the localization length obtained in the present calculation is $\alpha_0^{-1} \sim 20a$, which is much larger than $\alpha_0^{-1} \sim a$ estimated experimentally.

7.9 Summary and conclusion

A review of magnetotransport in antidot lattices mainly from a theoretical point of view has been given. The commensurability peaks appearing in the magnetoresistivity can be understood quite well in terms of an enhancement

in the measure of diffusive orbits contributing to the diffusion coefficient due to a magnetic focusing effect. Differences existing between square and hexagonal lattices and roles of quantum effects have been understood by the same mechanism. It has also been shown that realistic antidot lattices have a large amount of disorder in the potential, which accounts for the appearance of the Altshuler–Aronov–Spivak oscillation and manifests itself in the commensurability peak and the superimposed Aharonov–Bohm-type oscillation. The disorder in the antidot potential gives rise to the localization of electron wavefunctions and the localization length oscillates as a function of a magnetic field with the period determined by half of the magnetic flux quantum.

Acknowledgments

This work is supported in part by the Japan Society for the Promotion of Science ('Research for the Future' program JSPS-RFTF96P00103). One of the authors (SU) has been supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. One of the authors (TN) acknowledges the support of a fellowship from the Special Postdoctoral Researches Program at RIKEN. Some numerical calculations were performed on FACOM VPP500 in the Supercomputer Center, Institute for Solid State Physics, University of Tokyo.

References

- [1] Weiss D, Roukes M L, Menschig A, Grambow P, von Klitzing K and Weimann G 1991 *Phys. Rev. Lett.* **66** 2790
- [2] Fleischmann R, Geisel T and Ketzmerick K 1994 *Europhys. Lett.* **25** 219
- [3] Gusev G M, Kvon Z D, Litvin L V, Nastaushev Yu V, Kalagin A K and Toropov A I 1992 *JETP Lett.* **55** 123
- [4] Gusev G M, Basmaji P, Kvon Z D, Litvin L V, Nastaushev Yu V and Toropov A I 1994 *J. Phys.: Condens. Matter* **6** 73
- [5] Nakamura K, Ishizaka S and Nihey F 1994 *Physica B* **197** 144
- [6] Nihey F, Hwang S W and Nakamura K 1995 *Phys. Rev. B* **51** 4649
- [7] Lorke A, Kotthaus J P and Ploog K 1991 *Phys. Rev. B* **44** 3447
Lorke A, Kotthaus J P and Ploog K 1991 *Superlatt. Microstruct.* **9** 103
- [8] Yamashiro T, Takahara J, Takagaki Y, Gamo K, Namba S, Takaoka S and Murase K 1991 *Solid State Commun.* **79** 885
- [9] Takahara J, Kakuta T, Yamashiro T, Takagaki Y, Shiokawa T, Gamo K, Namba S, Takaoka S and Murase K 1991 *Japan. J. Appl. Phys.* **30** 3250
- [10] Deruelle T, Ensslin K, Petroff P M, Efros A L and Pikus F G 1992 *Phys. Rev. B* **45** 9082
- [11] Berthold G, Smoliner J, Rosskopf V, Gornik E, Böhm G and Weimann G 1993 *Phys. Rev. B* **47** 10 383
- [12] Schuster R, Ensslin K, Kotthaus J P, Holland M and Stanley C 1993 *Phys. Rev. B* **47** 6843

- [13] Schuster R, Ensslin K, Wharam D, Kühn S, Kotthaus J P, Böhm G, Klein W, Tränkle T and Weimann G 1994 *Phys. Rev. B* **49** 8510
- [14] Weiss D, Richter K, Vasiliadou E and Lütjering G 1994 *Surf. Sci.* **305** 408
- [15] Deruelle T, Meurer B, Guldner Y, Vieren J P, Riek M, Weiss D, von Klitzing K, Eberl K and Ploog K 1994 *Phys. Rev. B* **49** 16561
- [16] Tsukagoshi K, Wakayama S, Oto K, Takaoka S, Murase K and Gamo K 1994 *Superlatt. Microstruct.* **16** 295
- [17] Nihey F and Nakamura K 1993 *Physica B* **184** 398
- [18] Weiss D, Richter K, Menschig A, Bergmann R, Schweizer H, von Klitzing K and Weimann G 1993 *Phys. Rev. Lett.* **70** 4118
- [19] Kumar A, Laux S E and Stern F 1990 *Phys. Rev. B* **42** 5166
- [20] Suzuki T and Ando T 1993 *J. Phys. Soc. Japan* **62** 2986
Suzuki T and Ando T 1994 *Physica B* **201** 345
Suzuki T and Ando T 1996 *Physica B* **227** 46
- [21] Fleischmann R, Geisel T and Ketzmerick R 1992 *Phys. Rev. Lett.* **68** 1367
- [22] Baskin E M, Gusev G M, Kvon Z D, Pogosov A G and Entin M V 1992 *JETP Lett.* **55** 678
- [23] Ensslin K, Sasa S, Deruelle T and Petroff P M 1992 *Surf. Sci.* **263** 319
- [24] Takahara J, Nomura A, Gamo K, Takaoka S, Murase K and Ahmed H 1995 *Japan. J. Appl. Phys.* **34** 4325
- [25] Tsukagoshi K, Wakayama S, Oto K, Takaoka S, Murase K and Gamo K 1995 *Phys. Rev. B* **52** 8344
- [26] Tsukagoshi K, Haraguchi M, Oto K, Takaoka S, Murase K and Gamo K 1995 *Japan. J. Appl. Phys.* **34** 4335
- [27] Tsukagoshi K, Haraguchi M, Takaoka S and Murase K 1996 *J. Phys. Soc. Japan* **65** 811
- [28] Nagao T 1995 *J. Phys. Soc. Japan* **64** 4097
- [29] Nagao T 1996 *J. Phys. Soc. Japan* **65** 2606
- [30] Ishizaka S and Ando T 1997 *Phys. Rev. B* **55** 16331
- [31] Lu W-C and Anderson O K 1996 *Proc. 23rd Int. Conf. on Physics of Semiconductors* ed M Scheffler and R Zimmermann (Singapore: World Scientific) p 1497
- [32] Ishizaka S, Nihey F, Nakamura K, Sone J and Ando T 1995 *Japan. J. Appl. Phys.* **34** 4317
- [33] Ando T and Uemura Y 1974 *J. Phys. Soc. Japan* **36** 959
- [34] Ando T 1974 *J. Phys. Soc. Japan* **36** 1521
Ando T 1974 *J. Phys. Soc. Japan* **37** 622
Ando T 1974 *J. Phys. Soc. Japan* **37** 1233
- [35] Schuster R, Ernst G, Ensslin K, Entin M, Holland M, Böhm G and Klein W 1994 *Phys. Rev. B* **50** 8090
- [36] Tsukagoshi K, Nagao T, Haraguchi M, Takaoka S, Murase K and Gamo K 1996 *J. Phys. Soc. Japan* **65** 1914
- [37] Ishizaka S and Ando T 1997 *Phys. Rev. B* **56** 15195
- [38] Silberbauer H and Rössler U 1994 *Phys. Rev. B* **50** 11911
- [39] Uryu S and Ando T 1996 *Proc. 23rd Int. Conf. on Physics of Semiconductors* ed M Scheffler and R Zimmermann (Singapore: World Scientific) p 1505
- [40] Ishizaka S, Nihey F, Nakamura K, Sone J and Ando T 1995 *Phys. Rev. B* **51** 9881
- [41] Hofstadter D 1976 *Phys. Rev. B* **14** 2239
- [42] Oakshott R B S and MacKinnon A 1993 *J. Phys.: Condens. Matter* **5** 6971

- Oakshott R B S and MacKinnon A 1993 *J. Phys.: Condens. Matter* **5** 6983
 Oakshott R B S and MacKinnon A 1993 *J. Phys.: Condens. Matter* **5** 6991
 Oakshott R B S and MacKinnon A 1994 *J. Phys.: Condens. Matter* **6** 1519
- [43] Uryu S and Ando T 1996 *Physica C* **227** 138
 [44] Gutzwiller M C 1967 *J. Math. Phys.* **8** 1979
 Gutzwiller M C 1969 *J. Math. Phys.* **10** 1004
 Gutzwiller M C 1970 *J. Math. Phys.* **11** 1791
 Gutzwiller M C 1971 *J. Math. Phys.* **12** 343
 [45] Gutzwiller M C 1978 *Path Integrals* ed G J Papadopoulos and J T Devrees (New York: Plenum) p 163
 [46] Gutzwiller M C 1990 *Chaos in Classical and Quantum Mechanics* (New York: Springer)
 [47] Miller W H 1975 *J. Chem. Phys.* **63** 996
 [48] Richter K 1995 *Europhys. Lett.* **29** 7
 [49] Wilkinson M 1987 *J. Phys. A: Math. Gen.* **20** 2415
 [50] Hackenbroich G and von Oppen F 1995 *Europhys. Lett.* **29** 151
 [51] Ishizaka S and Ando T 1998 *Solid State Electron.* **42** 1147
 [52] Nakanishi T and Ando T 1996 *Phys. Rev. B* **54** 8021
 [53] Nakanishi T and Ando T 1996 *Physica B* **227** 127
 [54] Nihey F, Kastner M A and Nakamura K 1997 *Phys. Rev. B* **55** 4085
 [55] Uryu S and Ando T 1998 *Physica B* **249–251** 308
 [56] Uryu S and Ando T 1998 *Phys. Rev. B* **58** 10 583
 [57] Altshuler B L, Aronov A G and Spivak B Z 1981 *JETP Lett.* **33** 1255
 [58] Sharvin D Yu and Sharvin Yu V 1981 *JETP Lett.* **34** 272
 [59] Dolan G J, Licini J C and Bishop D J 1986 *Phys. Rev. Lett.* **56** 1493
 [60] Ando T 1991 *Phys. Rev. B* **44** 8017
 [61] Uryu S and Ando T 1995 *Japan. J. Appl. Phys.* **34** 4295
 [62] Uryu S and Ando T 1996 *Phys. Rev. B* **53** 13 613
 [63] Uryu S 1998 *Doctor Thesis* (University of Tokyo)
 [64] Uryu S and Ando T *Physica B* in press
 [65] Uryu S and Ando T 1998 *Solid State Electron.* **42** 1141
 [66] Landauer R 1957 *IBM J. Res. Dev.* **1** 223
 Landauer R 1970 *Phil. Mag.* **21** 863
 [67] Büttiker M 1986 *Phys. Rev. Lett.* **57** 1761
 Büttiker M 1988 *IBM J. Res. Dev.* **32** 317
 [68] Edwards J T and Thouless D J 1972 *J. Phys. C: Solid State Phys.* **5** 807
 [69] Licciardello D C and Thouless D J 1975 *J. Phys. C: Solid State Phys.* **8** 4157
 [70] Ando T 1983 *J. Phys. Soc. Japan* **52** 1740
 Ando T 1984 *J. Phys. Soc. Japan* **53** 3101
 Ando T 1984 *J. Phys. Soc. Japan* **53** 3126
 [71] Abrahams E, Anderson P W, Licciardello D C and Ramakrishnan T V 1979 *Phys. Rev. Lett.* **42** 673
 [72] Dyson F J 1962 *J. Math. Phys.* **3** 140
 [73] Dorokhov O N 1983 *Pis. Zh. Eksp. Teor. Fiz.* **85** 1040
 Dorokhov O N 1983 *Sov. Phys.–JETP* **58** 606
 [74] Muttalib K, Pichard J L and Stone A D 1990 *Phys. Rev. Lett.* **65** 1812
 [75] Tamura H and Ando T 1993 *Physica B* **184** 355

Chapter 8

Edge states in magnetic quantum structures and composite fermion systems

G Ihm¹, H-S Sim², K J Chang² and S J Lee³

¹ *Department of Physics, Chungnam National University, Taejon 305-764, Korea*

² *Department of Physics, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea*

³ *Department of Physics, Korea Military Academy, Seoul 139-799, Korea*

We study the nature of edge states in magnetic quantum structures which are formed by inhomogeneous magnetic fields in two-dimensional electron systems. For a magnetic quantum dot, which is a simple magnetic structure, we find that the edge states circulate either clockwise or counterclockwise along the boundary region of the dot, depending on the number of missing flux quanta, and exhibit quite different properties, as compared to those induced by electrostatic potentials in the quantum Hall system. We also investigate composite fermion edge states in spatially varying electrostatic potentials, where the effective magnetic field in the context of composite fermions is inhomogeneous. Considering composite fermion edge states near an antidot potential, we explain the origin of the asymmetric commensurability peaks in recent experiments. The current carried by composite fermion edge channels in the edge of the fractional quantum Hall systems is also found to be consistent with recent experimental data.

8.1 Introduction

In past decades, a great deal of interest has been focused on the electron transport behaviour in two-dimensional electron gas (2DEG) systems. In the quantum Hall regime, the current-carrying edge states, formed near the boundary region, play an

important role in describing the resonant tunnelling, Aharonov–Bohm oscillation and non-local magnetoresistance, etc [1, 2]. Recently, with the application of spatially inhomogeneous magnetic fields, a number of alternative magnetic structures were proposed on 2DEG, such as magnetic quantum dots using a scanning tunnelling microscope lithographic technique [3], magnetic superlattices by the patterning of ferromagnetic materials integrated by semiconductors [4], type-II superconducting materials deposited on conventional heterostructures [5], and non-planar 2DEG systems grown by a molecular beam epitaxy [6]. In those magnetic structures, one can find a close analogy with the conventional quantum structures, which are modulated by electrostatic potentials, but expect quite different electron transport properties. In the inhomogeneous magnetic field, which varies linearly across a 2DEG, electrons propagate in the direction perpendicular to the field gradient [7]. For a 2DEG in a weak periodic modulation of magnetic fields commensurability phenomena come into play [8, 9]. Moreover, composite fermion (CF) systems [10] in slowly varying electrostatic potentials are closely related with the magnetic quantum structures, because the effective magnetic field in the context of CFs depends on electron density [11, 12]. Thus, a variety of new phenomena associated with the magnetic structures are expected in the electron transport.

In this paper, we investigate the nature of edge states in magnetic quantum structures which are formed by inhomogeneous magnetic fields in 2DEG. First, we consider a magnetic quantum dot; electrons are apparently confined to a plane and within that plane the magnetic field is zero within a circular disc and constant B outside it [13–15]. We calculate exactly the single electron eigenstates and energies of a magnetic quantum dot as a function of magnetic field, using a single scaled parameter $s = \pi r_0^2 B / \phi_0$, which represents the number of missing magnetic flux quanta within the dot, where r_0 is the radius of the quantum dot and $\phi_0 (= h/e)$ is the flux quantum. We find two types of edge states which circulate in opposite directions to each other along the boundary of the magnetic dot and exhibit quite different energy dependences on angular momentum. For a small conductor with a magnetic quantum dot at the centre, the calculated magnetoconductances show aperiodic oscillations instead of the Aharonov–Bohm-type of periodic oscillations [16], and this behaviour is attributed to the characteristics of the magnetic edge states, which is absent in the conventional ones.

Second, we consider CFs near an antidot potential. Recently, Kang *et al* observed remarkable commensurability magnetoresistance peaks in antidot superlattices near $v = 1/2$, at which the effective CF magnetic field B_{eff} is zero [17]. Here, v is the electron filling factor. Compared with the commensurability peaks of electrons near zero external magnetic field ($B = 0$), the peaks near $v = 1/2$ show strong asymmetry in their heights in contrast to the symmetrical peaks near $B = 0$; the lower field peak ($B_{\text{eff}} < 0, v > 1/2$) exceeds the higher field peak ($B_{\text{eff}} > 0, v < 1/2$) by as much as a factor of 10. It is the motivation of the second part of this work to explain the physical origin of this asymmetry. Previous

theories [18] based on the quasi-classical dynamics of CFs explained this feature through the different set of chaotic trajectories but does not give a clear view as will be presented here. We consider a CF system in an antidot as a magnetic quantum dot; the magnetic field is B_0^* within a circular disk of radius r_0 while $B_0 (\neq B_0^*)$ outside it. We show that the strong asymmetry of commensurability peaks arise primarily from quite different eigenenergy dispersions between the cases of the lower field peak ($\alpha < 0$) and the higher field peak ($\alpha > 0$), where $\alpha (= B_0^*/B_0)$ is a field ratio.

Finally, we consider CF edge channels in the fractional quantum Hall systems. Motivated from the successful description in a bulk region using CFs, it is natural to extend the CF approach to the edge states in the fractional quantum Hall regime [19, 20]. However, to our knowledge, the CF approach in describing the electron transport near edges has not been well established and not compared with the interacting-electron picture. In the third part of this paper, we investigate the CF edge states in the fractional quantum Hall system which consists of alternating compressible and incompressible strips [21, 22]. We derive the change of the effective CF chemical potential $\Delta\mu_{\text{eff}}$ due to a small variation of the electron chemical potential $\Delta\mu$, where μ_{eff} is the energy cost to add one CF into the system. We also find the change of current carried by CFs in a compressible strip to be $\Delta I = (-e/h)\Delta\nu\Delta\mu$, where $\Delta\nu$ is the difference of the electron filling factors between the two adjacent incompressible regions. This result is consistent with previous theoretical work [22] and recent experimental data [23, 24], thus, the CF edge channels are well defined.

This paper is organized as follows. In section 8.2 we study the magnetic quantum dot. In section 8.3, we apply the magnetic quantum dot to the CF commensurability peak problem. In section 8.4, we investigate the CF edge channels in a slowly varying confinement potential. We finish in section 8.5 with a summary.

8.2 Magnetic quantum dot

In this section, we study the nature of edge states in a magnetic quantum dot which is formed by inhomogeneous magnetic fields. Hereafter, these edge states will be termed the magnetic edge states in close analogy with electrostatically induced conventional ones¹. The single particle Schrödinger equation for a two-dimensional magnetic quantum dot is $(\vec{p} + e\vec{A})^2/(2m^*)\psi'(\vec{r}) = E\psi(\vec{r})$, where m^* is the effective mass of electron and e is the absolute value of the electron charge. In polar coordinates (r, θ) on the plane, the vector potential \vec{A} can be chosen as 0 for $r < r_0$ and $(r^2 - r_0^2)B/(2r)\hat{\theta}$ for $r > r_0$, so that $B = 0$ for $r < r_0$ and non-zero $B\hat{z}$ otherwise. The wavefunctions and the energies are easily determined by the continuity of the wavefunctions and their derivatives at the boundary of the dot.

¹ These conventional channels are often referred to as the magnetic edge channels to distinguish them from the narrow channels in the absence of magnetic fields. In our work this is not the case.

Since the wavefunctions are separable, i.e. $\psi_{nm}(\vec{r}) = R_{nm}(r)e^{im\theta}$, where m is the angular momentum quantum number and $n (= 0, 1, 2 \dots)$ is the number of nodes in the radial wavefunction, the equation for the radial part is written as

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{m^2}{r^2} + 2E \right) R_{nm}(r) = 0 \quad (r < r_0) \quad (8.1)$$

$$\left\{ \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{(m-s)^2}{r^2} - r^2 + 2[E - (m-s)] \right\} R_{nm}(r) = 0 \quad (r \geq r_0). \quad (8.2)$$

Here

$$R_{nm}(r) = C_1 J_{|m|}(\sqrt{2Er}) \quad \text{for } r < r_0$$

and

$$R_{nm}(r) = C_2 r^{|m-s|} e^{-r^2/2} U(a, b; r^2) \quad \text{for } r \geq r_0.$$

In this case, all quantities are expressed in dimensionless units by letting $\hbar\omega_L [= \hbar eB/(2m^*)]$ and the inverse length $\beta = \sqrt{m^*\omega_L/\hbar}$ be 1. Then, since $\hbar^2/m^* = \hbar\omega_L/\beta^2 \rightarrow 1$ and $r_0 \rightarrow \sqrt{s}$, $s = B\pi r_0^2 e/h$ is only the relevant parameter. The function J_m is the Bessel function of order m and U is the confluent hypergeometric function with $a = -(E - m_{\text{eff}} - |m_{\text{eff}}| - 1)/2$, $b = |m_{\text{eff}}| + 1$ and $m_{\text{eff}} = m - s$. It is noted that equation (8.2) has the same form as that of the uniform magnetic field case, except that the angular momentum m is replaced by the effective angular momentum m_{eff} .

In the magnetic quantum dot, the Landau level degeneracy is lifted for the states near the dot. From equations (8.1) and (8.2), if the effective potential $V_{\text{eff}}(r)$ is defined as

$$V_{\text{eff}}(r) = \begin{cases} \frac{m^2}{2r^2} & (r < r_0) \\ \frac{m_{\text{eff}}^2}{2r^2} + \frac{r^2}{2} + m_{\text{eff}} & (r \geq r_0), \end{cases} \quad (8.3)$$

the minimum of $V_{\text{eff}}(r)$ always occurs at $r = r_0 (= \sqrt{s})$ for the states with $|m_{\text{eff}}| < s$, i.e. $0 < m < 2s$, which correspond to the magnetic edge states circulating counterclockwise, as we will see later. The $m = 0$ state is widely distributed over the dot due to the lack of the centrifugal force, and the minimum of $V_{\text{eff}}(r)$ for the states with $|m_{\text{eff}}| > s$, i.e. $m < 0$ or $m > 2s$, is located at $r = \sqrt{|m_{\text{eff}}|}$ outside the quantum dot, similar to the case of uniform magnetic fields. The states with $m < 0$, which exist near the dot, give rise to the magnetic edge states circulating clockwise. [Figure 8.1](#) shows the energy levels of the magnetic quantum dot for different values of m at $s = 5$, the radius of which is about 500 Å for magnetic fields of teslas. The lowest energy state occurs at $m = 0$ and the degeneracy of the Landau levels are removed, as shown in figure 8.1. This result indicates that the inhomogeneity of magnetic fields mostly perturbs the states near the boundary of the quantum dot, and this perturbation is caused

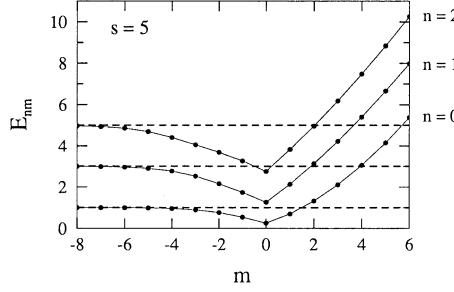


Figure 8.1. Dependence of the energy eigenvalues E_{nm} on the angular momentum m for $s = 5$. Broken curves represent the bulk Landau levels.

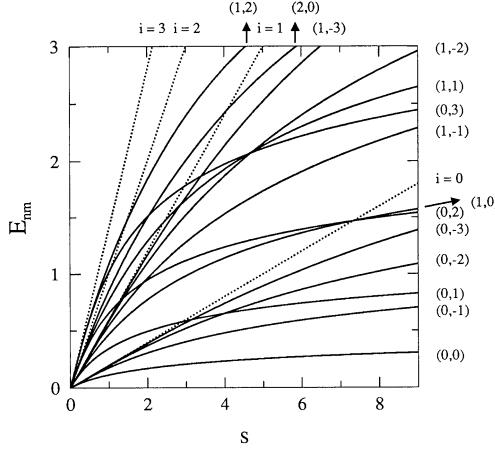


Figure 8.2. Energy spectra as a function of s . The energy unit of $\hbar\omega_L = 1$ at $s = 5$ is used. Dotted curves represent the Landau levels.

by the missing of s flux quanta. The probability currents [25] $I_{nm} = \frac{1}{\hbar} \frac{\partial E_{nm}}{\partial m}$ for the perturbed states are found to be non-zero, resulting in the magnetic edge states; for $m > 0$, I_{nm} have positive values for counterclockwise circulations whereas for $m < 0$, I_{nm} have negative values for clockwise ones. In figure 8.2, the energy levels are plotted as a function of magnetic field s for different values of (n, m) , with the energy $\hbar\omega_L$ set to one at $s = 5$ and the radius r_0 fixed. As the magnetic field increases, the deviations of energies from the bulk Landau levels become significant, which lead to the magnetic edge states near the boundary of the quantum dot. In the limit of $B \rightarrow \infty$, we find that the energies approach those for the conventional circular dot which is electrostatically confined by hard walls without magnetic fields.

Now, we discuss the difference between the magnetic edge states and

conventional ones, by investigating the quantum interference effect in the magnetic quantum dot. We consider a small two-dimensional conductor with a magnetic quantum dot at the centre. For magnetic fields which give the quantum Hall plateaux, the transport along the boundary of the sample, which is usually promoted by conventional edge states, can be backscattered by the resonant tunnelling into the magnetic edge states along the boundary of the dot, because of the impurity effect in the narrow region between two boundaries. In usual quantum dots or ring structures, the resonant tunnelling effect in magnetoresistance measurements gives rise to the Aharonov–Bohm oscillations [16, 26], which are periodic with magnetic field. In the magnetic quantum dot considered here, we do not see such periodic oscillations. We calculate the two-terminal conductance, which is the inverse of the sum of magnetoresistance and Hall resistance, taking into account the resonant backscattering via the magnetic edge channels as follows:

$$G(B) = \frac{2e^2}{h} \left[1 - \sum_{n,m} \frac{\Gamma^2}{(E_F - E_{nm}(B))^2 + \Gamma^2} \right], \quad (8.4)$$

where Γ is the elastic resonance width and a constant value of $\Gamma = 0.005$ is used for simplicity. The calculated conductance is plotted as a function of magnetic field in figure 8.3, with the Fermi energy of $E_F = 2$ in units of figure 8.2. In this case, the magnetic fields represented by s are in the $\nu = 2$ quantum Hall plateau region, where ν is the Landau level filling factor. We find that the oscillations are not periodic, in contrast to the Aharonov–Bohm-type of oscillations. The first dip in the conductance around $s = 3.7$ is due to the resonant backscattering via the $(1, -3)$ magnetic edge state. The other dips are found to be associated with the $(0, 3)$, $(1, 1)$, $(1, -2)$, and $(1, -1)$ states in the increasing order of s . In the narrow ring structure of Jain [16], the intervals between the dips were shown to be periodic, which indicates the subsequent change of one flux quantum passing through the inner boundary. In our magnetic dot structure, the resonances occur via the two different magnetic edge states circulating in different directions, depending on the sign of m . Since there is no magnetic field inside the magnetic dot, the magnetic edge states may not enclose the magnetic flux, resulting in the missing of flux quanta, which is absent in the edge states formed by electrostatic confinements.

The result of this section can be extended to more complex systems and applied to the edge states of CFs [27] in the fractional quantum Hall system with a spatially varying electrostatic potential. In the next section, we will study CFs in an antidot using a modulated magnetic dot.

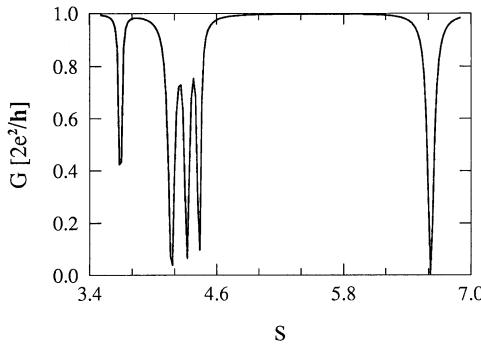


Figure 8.3. Magnetoconductance as a function of s .

8.3 Composite fermions in an antidot: application of magnetic quantum dot

In this section, we study CFs in an antidot potential. A composite fermion (CF) consists of an electron and q magnetic flux quanta attached on it, where q is an even number [10]. Within this picture, the fractional quantum Hall plateau at the Landau level filling factor $\nu = p/(qp + 1)$ corresponds to an integer number p of filled Landau levels of CFs. The CFs see an effective magnetic field $B_{\text{eff}} (= B_{\text{ext}} - qn_e\phi_0)$ in the mean field sense, where B_{ext} is a perpendicular external magnetic field, n_e is the electron density and ϕ_0 is a flux quantum. At $\nu = 1/2$, where $B_{\text{eff}} = 0$, it has been asserted that the Fermi surface of CFs with $q = 2$ is well defined, with gapless CF states [11].

When there are CFs around an antidot in a uniform external magnetic field B_{ext} , B_{eff} is non-uniform, since n_e is decreased due to the antidot potential. Near the centre of an antidot, where CFs cannot exist, CFs interact with much higher effective field B_{ext} than B_{eff} of other regions. We can model this inhomogeneous distribution of B_{eff} as a modulated magnetic quantum dot where $B_0^* = B_{\text{ext}}$, B_0 is a constant value of B_{eff} in the region far from the antidot ($|B_0^*/B_0| \ll 1$), and r_0 is chosen to enclose correct magnetic flux at the region of $B_{\text{eff}} = B_0$. The sign of B_0 is the same as B_0^* for the higher field-commensurability condition, whilst it is opposite to B_0^* for the lower field condition. For the sake of convenience, we will always take B_0 as positive. The primary commensurability peaks around $\nu = 1/2$, which are our concern, are mainly associated with localized CF states around an antidot at the Fermi energy, the number of which is limited by the period of antidot lattice. These localized states correspond to the pinned orbits in the previous works [28] which do not contribute to the conduction. Thus, studying CF states of the magnetic quantum dot for the higher field- and the lower field-resonant conditions, we can obtain a qualitative view for the asymmetric resonant peaks rather than the quantitative evaluation of the peaks. The latter

requires the calculation for the quantum transport of CFs in a whole antidot lattice. In the magnetic quantum dot, since the high enough B_0^* already prohibits CFs from moving into the dot, the antidot electrostatic potential gives only a small correction to the CF states so that it will be neglected in our work.

The single particle Schrödinger equation for a modulated magnetic quantum dot is equivalent to that in section 8.2 except for the CF effective mass m_{CF}^* and V_{eff} . We first consider the case of constant m^* . Because $B = B_0^*$ for $r < r_0$ and $B = B_0$ for $r > r_0$, using dimensionless units in the previous section, $V_{\text{eff}}(r)$ can be defined as

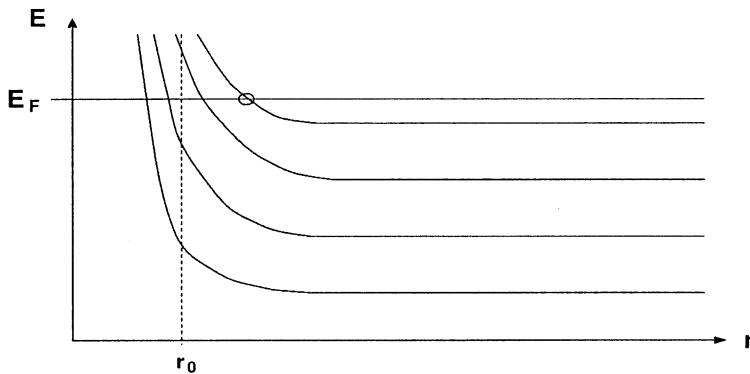
$$V_{\text{eff}}(r) = \begin{cases} \frac{m^2}{2r^2} + \frac{1}{2}\alpha^2 r^2 + \alpha m & (r < r_0) \\ \frac{m_{\text{eff}}^2}{2r^2} + \frac{1}{2}r^2 + m_{\text{eff}} & (r > r_0), \end{cases} \quad (8.5)$$

where $m_{\text{eff}} = m - s$ is the effective angular momentum, $s = s_0(1 - \alpha)$, and $\alpha = B_0^*/B_0$. From the experiments in [16], $B_0 = 0.35$ T, $B_0^* \simeq \pm 11.55$ T, and thus, $\alpha \simeq \pm 33$ for the higher field- and lower field-resonant peak conditions, respectively. The radius $r_0 (= 125$ nm) and the period $d (= 600$ nm) of the antidot lattice are 2 and 10, respectively, in the dimensionless units. The Fermi energy (E_F) of CFs is of the order of $|\alpha|$. Our main interests reside in the distribution of localized states near a single antidot (or equivalently a magnetic quantum dot) at E_F . In figure 8.4, for the resonant conditions, schematic representations² of CF energy levels of the magnetic quantum dot are shown as a function of position, where V_{eff} has a minimum value. Although there are $p (\sim 17)$ of levels below E_F , only several levels which cross E_F are presented for simplicity. The smooth variation of antidot electrostatic potential and B_{eff} near r_0 modify energy levels, but do not change the overall feature of energy dispersions, as we will see later. There exist large differences between figures 8.4(a) and (b); more localized states appear nearby the dot at E_F for the lower field-resonance than for the higher field-resonance condition.

When α is positive as in figure 8.4(a), CF energy levels monotonously increase as their locations become closer to the dot. The low lying levels cross E_F near the dot, while the higher levels cross E_F rather away from the dot. For the state marked by a circle in figure 8.4(a), which occurs when $m < -s_0\alpha$, the effective potential $V_{\text{eff}}(r)$ are shown in figure 8.5(a). This state may contribute to the conduction due to the scattering with neighbouring dots if r_b in figure 8.5(a) is greater than $d - d_0$, where $d_0 (\sim r_0)$ is the size of an antidot at E_F . This is confirmed in experiments [17] by noting that the height of the resonant peak becomes smaller as d decreases. Thus, the number of localized states near a given dot, which do not contribute to the conduction and give rise to the resistance, is less than p for the higher field-peak condition. When α is negative

² As a guide for one's eyes, the minimum values of each energy level in figure 8.4(b) are a little exaggerated, considering the effect of the smoothness of the antidot potential or B_{eff} . For a more smoothly varying B_{eff} , more energy levels have a minimum value which is smaller than E_F (see the dotted curve in figure 8.4(b)).

(a)



(b)

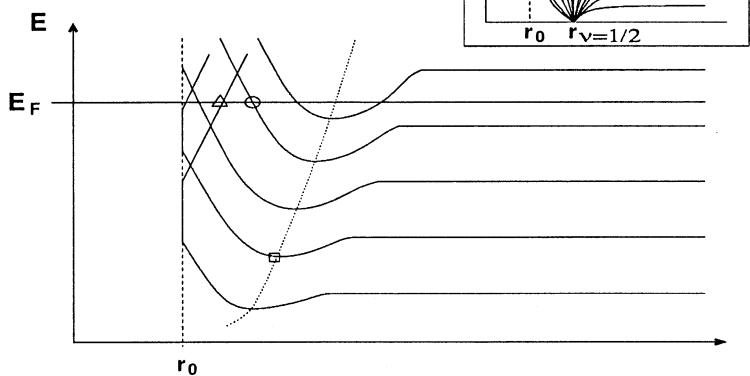


Figure 8.4. Schematic representations of CF energy levels of magnetic quantum dots as a function of position; (a) higher field-resonance ($\alpha > 0$) and (b) lower field-resonance ($\alpha < 0$). The inset in (b) shows the effect of CF mass divergence at $v = 1/2$ for sufficiently smoothly varying antidot potentials.

(see figure 8.4(b)), the energy level dispersion is very different from that in figure 8.4(a). The energy minima, shown as a dotted curve in figure 8.4(b), occur when $0 < m < |\alpha|s_0$. In this case, V_{eff} has double well structures, as shown in figure 8.5(b), with the energies lowered. This energy lowering becomes more significant for smoothly varying B_{eff} near r_0 since this type of B_{eff} lowers the barrier height in double well structures (see previous footnote). Thus, a large

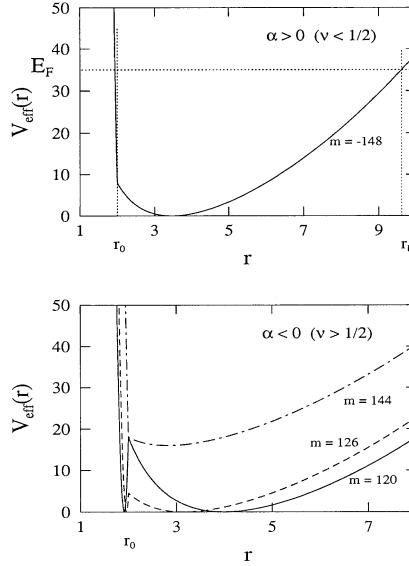


Figure 8.5. The shapes of V_{eff} in equation (8.5) for the states marked by (a) the circle in figure 8.4(a), and (b) the circle (broken), triangle (chain) and square (full) in figure 8.4(b).

number of levels, which are above E_F in the region far from the dot, can cross E_F near the dot as shown in figure 8.4(b). This striking feature is absent from the higher field-resonance condition. The positions of two minima in V_{eff} get closer to r_0 as m increases up to $|\alpha|s_0$ (see the circle in figure 8.4(b)). As m further increases, V_{eff} continues to have a single minimum at r_0 with increasing value of $V_{\text{eff}}(r_0)$ for $|\alpha|s_0 \leq m \leq (2 + |\alpha|)s_0$. For $m > (2 + |\alpha|)s_0$, the minimum position of V_{eff} becomes larger than r_0 , shown as the state marked by a triangle in figure 8.4(b) and a dashed-dot curve in figure 8.5(b). It should be noted that this state marked by a triangle always corresponds to the pinned orbits since the spatial extension of this state is fairly small due to the lifted bottom of V_{eff} (see figure 8.5(b)). This behaviour results in a larger peak height for the lower field-resonance ($\alpha < 0$) than for the higher field-resonance ($\alpha > 0$).

So far, we have neglected the variation in the effective mass of CFs. The diverging effective mass [11, 29] at $\nu = 1/2$ makes all the level minima down to the bottom of V_{eff} at the same position of zero effective field ($\nu = 1/2$), if the antidot electrostatic potential varies very smoothly [20]. Then, many levels which are above E_F in the region far from the dot go down and cross E_F near r_0 , as schematically shown in the inset of figure 8.4(b). Thus, the lower field-peak can be greatly enhanced due to the diverging effective mass at $\nu = 1/2$, while the higher field-peak is not affected by this effect since there is no region of $B_{\text{eff}} = 0$ in the latter case.

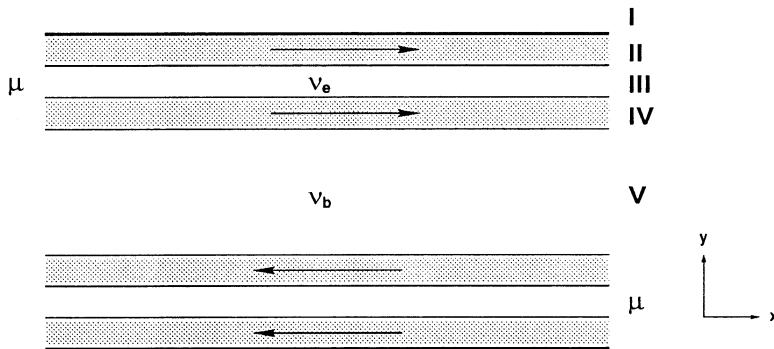


Figure 8.6. A schematic diagram for a two-dimensional conductor connected to the left and right reservoirs with the chemical potential μ . Compressible (shaded) and incompressible (white) regions are labelled. The arrows indicate additional current flows induced by $\Delta\mu$.

8.4 Composite fermion edge channels

In this section, we investigate the nature of CF edge states in the fractional quantum Hall system which consists of alternating compressible and incompressible strips [21] (see figure 8.6). For an ideal 2DEG system, which is connected to two electron reservoirs of an electron chemical potential μ , electrons are confined to move in the xy plane in the presence of an external uniform magnetic field $\vec{B} = B\hat{z}$ and a slowly varying electrostatic confinement potential $U(y)$, as shown in figure 8.6. The electron filling factor in the incompressible bulk region V is $\nu_b = p_b/(m_b p_b + 1)$, where p_b is the CF filling factor and m_b is the number of flux quanta bound to each electron. Note that m in this section is equivalent to q in the previous section. Although there may be many incompressible strips near edges, we focus on a simple case with one incompressible region III sandwiched between two compressible regions II and IV. This incompressible strip has the filling factor $\nu_e = p_e/(m_e p_e + 1) < \nu_b$, where p_e and m_e are the CF filling factor and the number of flux quanta, respectively. Because of the excitation energy gap [21, 22], the electron densities n_e in the incompressible regions III and V are uniformly distributed, while in the compressible strips, they decrease as going to the region I, where $\nu = 0$ or $p = 0$. The effective CF chemical potential in the upper edge is denoted by μ_{eff} .

In this section, we consider only the simple case³ of $p_b > 0$, $p_e > 0$, and $m_e = m_b = m$. In mean field theory, the effective electric and magnetic fields which interact with CFs are written as

$$\vec{E}_{\text{eff}} = \nabla_r U/e + \langle \vec{v} \rangle \times m\phi_0 n_e \hat{z} \quad (8.6)$$

³ The results for more general cases will be published elsewhere.

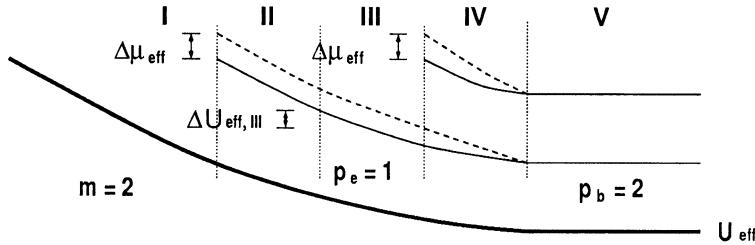


Figure 8.7. Schematic diagrams for the CF energy levels (full curves) for $\nu_b = 2/5$ and $\nu_e = 1/3$. In this case, $m_b = m_e = 2$, $p_b = 2$ and $p_e = 1$. The labelled regions are the same as those in figure 8.6. The heavy curves indicate the effective CF potentials and the broken curves represent the energy levels changed by ΔU_{eff} .

$$\vec{B}_{\text{eff}} = \vec{B} - m\phi_0 n_e \hat{z}, \quad (8.7)$$

where e is the absolute value of electron charge, $\phi_0 = h/e$ is the magnetic flux quantum, and $\langle \vec{v} \rangle$ is the average drift velocity. The second terms in equations (8.6) and (8.7) represent the fields induced by the magnetic flux bound to each CF [12]. From equation (8.6), we obtain the effective CF potential U_{eff} such as

$$U_{\text{eff}}(y) = U(y) + m\phi_0 \int^y dI, \quad (8.8)$$

where I denotes a current. In the incompressible regions, since B_{eff} is constant, the non-interacting-CF energies can be written as [11]

$$E_p = (p + \frac{1}{2})\hbar e |B_{\text{eff}}| / m_{\text{CF}}^* + U_{\text{eff}}, \quad (8.9)$$

where $p = 0, 1, \dots$. In the compressible strips with non-uniform electron densities, although we need self-consistent calculations to obtain the exact CF energies, we may guess the same expression as that of equation (8.9) for a sufficiently slowly varying potential U . The energy levels for $\nu_b = 2/5$ and $\nu_e = 1/3$ are shown in figure 8.7. If the sign of B_{eff} does not change in the compressible regions, the energy levels in the sandwiched incompressible region will be smoothly connected to those in the neighbouring regions. We address that although the exact expressions for the energy levels and U_{eff} are unknown, our following theory is independent of the exact form. In analogy with the previous theory [20], for the energy levels in figure 8.7, the direction for CF current is the same as that of electron current, consistent with recent edge-magnetoplasmon experiments [30].

If the chemical potential of the left reservoir is disturbed by $\Delta\mu$, the effective CF chemical potential μ_{eff} and the current I in each region will also be changed by $\Delta\mu_{\text{eff}}$ and ΔI , respectively. However, for a sufficiently small $\Delta\mu$, $\Delta I = 0$ in the incompressible regions, because of the uniform electron densities due

to the energy gap $\hbar e |B_{\text{eff}}| / m_{\text{CF}}^*$. Then, from equation (8.8), $\Delta U_{\text{eff},R_i}$ in an incompressible region R_i is simply written as

$$\Delta U_{\text{eff},R_i} = \phi_0 \sum_{R'=R_i}^V m \Delta I_{R'}, \quad (8.10)$$

i.e. $\Delta U_{\text{eff},I} = \phi_0 m (\Delta I_{\text{II}} + \Delta I_{\text{IV}})$ and $\Delta U_{\text{eff},\text{III}} = \phi_0 m \Delta I_{\text{IV}}$, and the energy levels E_p in this region are shifted by a constant value $\Delta U_{\text{eff},R_i}$, as shown in figure 8.7. Because the energy shift $\Delta U_{\text{eff},R_i}$ also changes the effective chemical potential, $\Delta \mu_{\text{eff}}$ satisfies the relation,

$$\Delta \mu_{\text{eff}} = \frac{e_{R_i}^*}{-e} \Delta \mu + \Delta U_{\text{eff},R_i}, \quad (8.11)$$

for $R_i \in \{\text{I, III}\}$, where $e_{R_i}^*$ is the local CF charge, i.e. $e_{\text{I}}^* = -e$ and $e_{\text{III}}^* = -e/(m_e p_e + 1)$. In this case, the first term results from the fact that $|mp + 1|$ CFs with a local charge e^* are excited in the incompressible region with $v = p/(mp + 1)$, when one electron is added to that region [11, 12]. We find that p_e energy levels intersect with μ_{eff} on the edge of region II, while $(p_b - p_e)$ levels intersect with μ_{eff} in the case of the region IV (see figure 8.7). These levels give rise to the current changes ΔI_{II} and ΔI_{IV} , respectively. The remaining p_e levels below the effective chemical potential in the region IV contribute to both ΔI_{II} and ΔI_{IV} , because of the energy shift $\Delta U_{\text{eff},\text{III}}$. As a consequence, the total current changes in the regions II and IV, which are derived from $I_p = -(e/h) \int dE_p$, where I_p is the current associated with the energy level p [2], are self-consistently related to $\Delta \mu_{\text{eff}}$ and ΔU_{eff} ,

$$\Delta I_{\text{II}} = -\frac{e}{h} (p_e \Delta \mu_{\text{eff}} - p_e \Delta U_{\text{eff},\text{III}}) \quad (8.12)$$

$$\Delta I_{\text{IV}} = -\frac{e}{h} [(p_b - p_e) \Delta \mu_{\text{eff}} + p_e \Delta U_{\text{eff},\text{III}}]. \quad (8.13)$$

Here we point out that the charge carried by each CF in the compressible regions is $-e$, because there is no excitation gap in these regions [21]. From equations (8.10)–(8.13), we derive $\Delta \mu_{\text{eff}}$ as

$$\Delta \mu_{\text{eff}} = \frac{\Delta \mu}{m_b p_b + 1}. \quad (8.14)$$

The current change ΔI_{R_c} in a compressible region R_c is also found to be

$$\Delta I_{R_c} = -\frac{e}{h} \Delta v_{R_c} \Delta \mu, \quad (8.15)$$

where Δv_{R_c} is the difference of the filling factors between the two neighbouring incompressible regions; $\Delta v_{\text{II}} = v_e$ and $\Delta v_{\text{IV}} = v_b - v_e$. The relations in equations (8.14) and (8.15) are satisfied for all possible combinations of m_b ,

m_e , p_b and p_e as well as for systems with many alternating incompressible and compressible strips near the edge⁴. From equation (8.14), we note that $\Delta\mu_{\text{eff}}$ depends only on m_b and p_b in the bulk region and is the same in all the incompressible regions, although the CF energy levels are shifted by different values. For all cases, we find that the current change ΔI_{R_c} in a compressible region R_c satisfies the relation in equation (8.15), consistent with the interacting electron picture by Beenakker [21]; ΔI_{R_c} does not depend on the CF filling factor and the number of flux quanta carried by each CF, but the filling factor difference $\Delta\nu_{R_c}$. Thus, the CF edge channels can be defined as the CF edge states in the compressible regions, and the resulting total current change caused by $\Delta\mu$ is expressed as $(-e/h)\nu_b\Delta\mu$. Furthermore, generalizing the Landauer–Büttiker formula [2], we can write from equation (8.15) the total current change such as $\Delta I = \sum_{R_c} \Delta\nu_{R_c} T_{R_c}$, assuming that a fraction T_{R_c} of the current change ΔI_{R_c} induced by one reservoir is transmitted to the other. Then, the conductance satisfies the relation $G = (e^2/h) \sum_{R_c} \Delta\nu_{R_c} T_{R_c}$, which is consistent with recent experiments [23, 24]. Our Landauer–Büttiker formula for CFs is more general than that in previous work [19, 20], because it provides a simpler and clearer description for complex systems with alternating compressible and incompressible strips. In the non-interacting-electron description, the current change is contributed from the energy levels intersecting with μ , then, ΔI_{R_c} in a compressible region R_c considered here could be written as $\Delta I_{R_c} = (-e/h)n\Delta\mu_{\text{eff}}$, where n is the number of CF energy levels intersecting with μ_{eff} . However, we address that ΔI_{R_c} is contributed from all the energy levels below μ_{eff} , as shown in equations (8.12) and (8.13). Thus, the properties of CF edge channels are different from those for non-interacting-electron edge channels. Finally, we emphasize that U_{eff} plays an important role in determining μ_{eff} and the current change within the CF edge theory, because U_{eff} and the CF energy levels depend on the distribution of edge currents which flow along compressible strips.

8.5 Conclusion

We study the edge states of electrons and composite fermions in non-uniform magnetic fields. For a magnetic quantum dot, the edge states, which are formed by inhomogeneous distribution of magnetic fields, depend critically on the number of missing flux quanta and show quite distinctive aperiodic oscillations in magnetoresistance. Using the magnetic quantum dot, we explain the origin of asymmetric commensurability peaks of composite fermions near $v = 1/2$ in the magnetoresistance measurements with antidot lattices. This asymmetric feature happens due to the non-uniform distribution of effective composite fermion magnetic field near antidots, which induce very different energy dispersions between the lower and higher field-resonance conditions. Finally, we describe

⁴ The results for more general cases will be published elsewhere.

the edge states in the fractional quantum Hall systems using composite fermions. We find that the current in a compressible region depends on the difference of the electron filling factors between the two neighbouring incompressible regions. The non-uniform effective composite fermion magnetic and electric fields are found to be important in describing edge channels.

Acknowledgments

We thank N Kim for helpful discussions. This work was supported in part by the SPRC of Jeonbuk National University, by the CTPC and CMS at KAIST, and by the Korea Ministry of Education, grant No BSRI 97-2435.

References

- [1] Willardson R K, Beer A C and Weber E R (ed) 1992 *Semiconductors and Semimetals* vol 35 (New York: Academic)
- [2] Büttiker M 1988 *Phys. Rev. B* **38** 9375
- [3] McCord M A and Awschalom D D 1990 *Appl. Phys. Lett.* **57** 2153
- [4] Leadbeater M L, Allen S J Jr, DeRosa F, Harbison J P, Sands T, Ramesh R, Florez L T and Keramidas V G 1991 *J. Appl. Phys.* **69** 4689
Krishnan K M 1992 *Appl. Phys. Lett.* **61** 2365
- [5] Bending S J, von Klitzing K and Ploog K 1990 *Phys. Rev. Lett.* **65** 1060
- [6] Leadbeater M L, Foden C L, Burke T M, Burroughes J H, Grimshaw M P, Ritchie D A, Wang L L and Pepper M 1995 *J. Phys.: Condens. Matter* **7** L307
- [7] Müller J E 1992 *Phys. Rev. Lett.* **68** 385
- [8] Peeters F M and Matulis A 1993 *Phys. Rev. B* **48** 15 166
Peeters F M and Matulis A 1994 *Phys. Rev. Lett.* **72** 1518
Peeters F M and Matulis A 1997 *Phys. Rev. B* **56** 7508
Peeters F M and Vasilopoulos P 1993 *Phys. Rev. B* **47** 1466
- [9] Xue D P and Xiao G, 1992 *Phys. Rev. B* **45** 5986
Wu X G and Ulloa S E 1992 *Solid State Commun.* **82** 945
- [10] Jain J K 1989 *Phys. Rev. Lett.* **63** 199
- [11] Halperin B I, Lee P A and Read N 1993 *Phys. Rev. B* **47** 7312
- [12] Goldhaber A S and Jain J K 1995 *Phys. Lett. A* **199** 267
- [13] Solimany L and Kramer B 1995 *Solid State Commun.* **96** 471
- [14] Sim H-S, Ahn K-H, Chang K J, Ihm G, Kim N and Lee S J 1998 *Phys. Rev. Lett.* **80** 1501
- [15] Peeters F M, Matulis A and Ibrahim I S 1996 *Physica B* **227** 131
- [16] Jain J K 1988 *Phys. Rev. Lett.* **60** 2074
- [17] Kang W, Stormer H L, Pfeiffer L N, Baldwin K W and West K W 1993 *Phys. Rev. Lett.* **71** 3850
- [18] Fleischmann R, Geisel T, Holzknecht C and Ketzmerick R 1996 *Europhys. Lett.* **36** 167
- [19] Brey L 1994 *Phys. Rev. B* **50** 11 861
Chklovskii D B 1995 *Phys. Rev. B* **51** 9895
- [20] Kirczenow G and Johnson B L 1995 *Phys. Rev. B* **51** 17 579

- Kirczenow G 1996 *Phys. Rev. B* **53** 15 767
- [21] Beenakker C W J 1990 *Phys. Rev. Lett.* **64** 216
- [22] Chang A M 1990 *Solid State Commun.* **74** 871
- [23] Chang A M and Cunningham J E 1989 *Solid State Commun.* **72** 651
- [24] Kouwenhoven L P, van Wees B J, van der Vaart N C, Harmans C J P M, Timmering C E and Foxon C T 1990 *Phys. Rev. Lett.* **64** 684
- Chang A M and Cunningham J E 1992 *Phys. Rev. Lett.* **69** 2114
- [25] Halperin B I 1982 *Phys. Rev. B* **25** 2185
- [26] van Wees B J, Kouwenhoven L P, Harmans C J P M, Williamson J G, Timmering C E, Broekaart M E I, Foxon C T and Harris J J 1989 *Phys. Rev. Lett.* **62** 2523
- Tan W-C and Inkson J C 1996 *Phys. Rev. B* **53** 6947
- [27] Chklovskii D B and Lee P A 1993 *Phys. Rev. B* **48** 18 060
- [28] Weiss D, Roukes M L, Menschig A, Grambow P, von Klitzing K and Weimann G 1991 *Phys. Rev. Lett.* **66** 2790
- Fleischmann R, Geisel T and Ketzmerick R 1992 *Phys. Rev. Lett.* **68** 1367
- [29] Kukushkin I V, Haug R J, von Klitzing K and Ploog K 1994 *Phys. Rev. Lett.* **72** 736
- Manoharan H C, Shayegan M and Klepper S J 1994 *Phys. Rev. Lett.* **73** 3270
- Du R R, Störmer H L, Tsui D C, Yeh A S, Pfeiffer L N and West K W 1994 *Phys. Rev. Lett.* **73** 3274
- Kukushkin I V, Haug R J, von Klitzing K and Ploog K 1994 *Phys. Rev. Lett.* **72** 736
- Manoharan H C, Shayegan M and Klepper S J 1994 *Phys. Rev. Lett.* **73** 3270
- Du R R, Störmer H L, Tsui D C, Yeh A S, Pfeiffer L N and West K W 1994 *Phys. Rev. Lett.* **73** 3274
- [30] Ashoori R C, Stormer H L, Pfeiffer L N, Baldwin K W and West K 1992 *Phys. Rev. B* **45** 3894

Chapter 9

Electronic states in circular and ellipsoidally deformed quantum dots

*S Tarucha^{1,2}, D G Austing², S Sasaki², L P Kouwenhoven³,
S M Reimann⁴, M Koskinen⁴ and M Manninen⁴*

¹ *Department of Physics, University of Tokyo, 7-3-1, Hongo,
Bunkyo-ku, Tokyo, 113-0033, Japan*

² *NTT Basic Research Laboratories, 3-1, Morinosato, Wakamiya,
Atsugi-shi, Kanagawa 243-0198, Japan*

³ *Department of Applied Physics, Delft University of Technology,
PO Box 5046, 2600 GA Delft, The Netherlands*

⁴ *University of Jyväskylä, Physics Department, PO Box 35,
40351, Jyväskylä, Finland*

The electronic properties of circular and ellipsoidally deformed quantum dot single electron transistors are studied by measuring Coulomb oscillations of circular and rectangular mesas. The filling of the few electron ground states in circular mesa devices at zero magnetic field gives rise to a pronounced shell structure characteristic of a two-dimensional harmonic confining potential with rotational symmetry, and is in accordance with Hund's first rule. These atomic-like features are readily disrupted in the rectangular mesa devices due to the lifting of the single-particle level degeneracies. Measurements with a magnetic field parallel to the current allow us to identify changes in the quantum numbers labelling the electronic states which are induced by the deformation. In particular, the transition from a spin-triplet in a circular dot to a spin-singlet in elliptical dots for the four-electron ground state at zero-magnetic field is investigated. The spin-singlet in the rectangular mesa is clearly identified independently from the measurement of the Zeeman effect. These observations are in good agreement with model calculations based on spin-density functional theory, and a single-particle picture is also very useful. Even for a small deformation, breaking the

circular symmetry significantly modifies the shell structure and changes the total spin of a state. Finally, we argue that the magnetic field dependence for the rectangular mesa devices suggests that the anisotropy of an elliptical dot may be much higher than that suggested by the geometry of the device mesa in which the dot is located.

9.1 Introduction

Recent advances in epitaxial growth and processing technology have enabled us to fabricate semiconductor nanostructures whose dimensions are comparable to the de Broglie wavelength of electrons. Transport measurements of these nanostructures have revealed a variety of phenomena associated with the effects of quantum mechanical confinement [1]. Conductance quantization in one-dimensional quantum point contacts, and resonant tunnelling through quantum wires and quantum boxes are such examples. These properties directly reflect the quantization of energy. In addition, charge quantization is observed for electrons tunnelling through a small dot. This dot acts as an island for the electrons. When tunnelling occurs, the charge on the island suddenly changes by the quantized amount ' e '. This leads to a change in the electrostatic potential of the dot by a unit of energy called the 'charging energy', $E_c = e^2/C$, where C is the typical capacitance of the island. The one-by-one change in the number of electrons on the island, N , gives rise to oscillations in the tunnelling conductance (Coulomb oscillations) when the gate voltage is swept. These oscillations are usually periodic when the number of electrons is 'large'. However, in a small dot holding just a few electrons, the charging energy can no longer be parametrized with a constant capacitance, and the Coulomb oscillations are significantly modified by electron-electron interactions, and quantum confinement effects. Thus, the *addition energy* needed to put an extra electron on the dot becomes strongly dependent on the number of electrons on the dot. Such a system can be regarded as an artificial atom [2].

The energy spectrum of a quantum dot has been extensively studied for small semiconductor dots, using so-called single electron transistors (SETs) with many different geometries and configurations. This is particularly true for SETs with a planar geometry defined in a two-dimensional electron gas by surface Schottky gates. The lateral constrictions forming the dot, the tunnel junctions between the dot and the reservoirs, and the plunger gate used to tune the electrostatic potential of the dot are all made by the Schottky gates. This enables a wide variation of device geometry, and a large freedom of gate operation for tuning the transmission probability of the tunnel junctions as well as the electron number in the dot. However, there are a number of limitations. For example, the geometry of the dot is not so well defined since the actual confining potential imposed by the Schottky gates can be significantly different from the gate geometry. In addition, it is almost impossible to fabricate a dot containing just a few electrons. In contrast, due to

the presence of heterostructure tunnel barriers and vertical side walls, the dot in a vertical quantum dot SET can have both good shape and high symmetry. Such a vertical SET allows N to be precisely tuned, *starting from zero*. The fabrication technology has only recently been developed [3, 4]. Transport measurement on vertical SETs reveals atom-like properties such as a two-dimensional (2D) ‘shell structure’ and filling in accordance with Hund’s first rule. The high symmetry of the vertical circular dot leads to maximal level degeneracy of the 2D single-particle states for parabolic lateral confinement. At 0 T, consecutive filling of each set of degenerate states is directly responsible for the characteristic shell structure, and Hund’s first rule accounts for the parallel filling of electrons amongst a half-filled shell of degenerate states. In this paper we first review the atom-like properties observed for the circular dot. Then, we discuss the effect of anisotropy in the lateral confinement on the shell structure and the spin states. Breaking the circular symmetry by deforming the lateral confining potential lifts the degeneracies present in the circular dot, and ‘destroys’ the distinctive shell structure for the circle, and modifies other atomic-like properties [5, 6]. In the experiment, we use rectangular dots to study the effect of anisotropy in the lateral confinement, and compare results to model calculations based on spin-density functional theory. In addition, we measure the Zeeman effect to identify the spin states in the rectangular dots.

9.2 Experimental details

We use a double barrier structure (DBS) to fabricate vertical quantum dots with a high degree of cylindrical symmetry in the lateral confinement, and with an asymmetry in the lateral confinement. The DBS consists of an undoped 13 nm $\text{In}_{0.05}\text{Ga}_{0.95}\text{As}$ well and two undoped $\text{Al}_{0.22}\text{Ga}_{0.78}\text{As}$ barriers of thickness 9.0 and 7.5 nm, and this is processed to form one circular mesa with a nominal top contact diameter, D , of $0.5 \mu\text{m}$ (W), and three rectangular mesas with top contact area $(L \times S)0.55 \times 0.4 \mu\text{m}^2$ (X), $0.65 \times 0.45 \mu\text{m}^2$ (Y), and $0.6 \times 0.4 \mu\text{m}^2$ (Z). $L(S)$ is the nominal dimension of longest (shortest) side of the top contact [3, 4]. [Figure 9.1](#) shows a schematic diagram of the circular mesa device. Above and below the DBS there is an n-doped GaAs contact (source and drain), and the circular dot is located between the two heterostructure barriers. A single Schottky gate is placed on the side of the mesa, wrapping round the dot. [Figure 9.2](#) shows typical scanning electron micrographs of a circular mesa, and a rectangular mesa taken immediately after the deposition of the gate metal. For the rectangular mesas, a simple way to classify them is to define a geometric parameter, β , to be the ratio L/S . For X, Y, and Z respectively, β is nominally 1.375, 1.44 and 1.5. Due to a slight undercut, the area of the mesas is a little less than that of the top contact. [Figure 9.2](#) also shows schematically the slabs of semiconductor between the two tunnelling barriers, and the resulting dots bounded by the shaded depletion region.

The current, I , flowing vertically through the dot is measured as a function

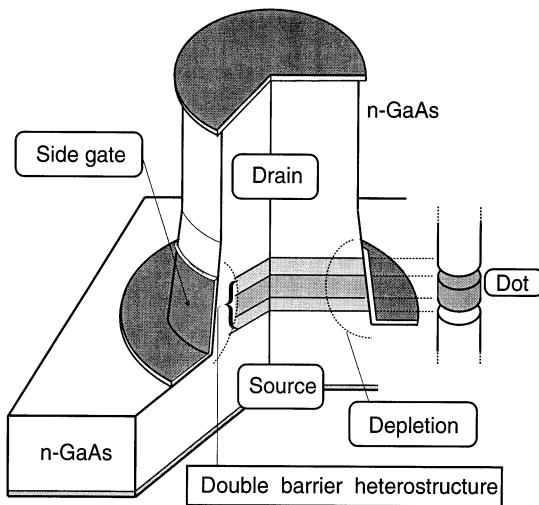


Figure 9.1. Schematic of the quantum dot in a vertical device.

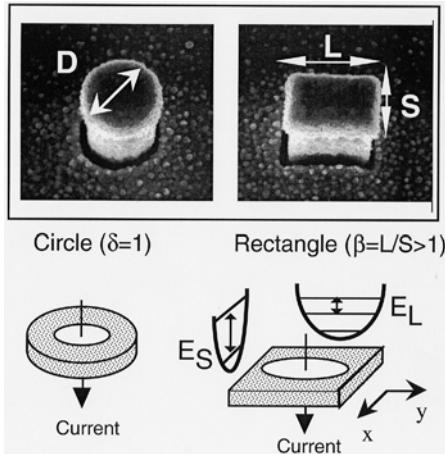


Figure 9.2. Scanning electron microscope (SEM) images showing submicrometre circular and rectangular mesas, and schematic diagrams of the resulting circular and elliptical shaped dots in the mesas.

of gate voltage, V_g , in response to a dc voltage, V_{sd} , applied between the contacts. For an arbitrarily small V_{sd} , the ground states (GSs) of an N -electron quantum dot can be investigated directly from monitoring the current. When no current flows (Coulomb blockade), the number of electrons in the dot, N , is well defined. On the other hand, when current flows the number of electrons can oscillate from N

to $N + 1$. Sweeping V_g , a series of sharp current peaks due to the charging of the dot (Coulomb oscillations) can be observed.

9.3 Electronic properties of circular quantum dots

9.3.1 Atom-like properties: shell structure and Hund's first rule

A good analogue to the three-dimensional shell in atoms [8] can be realized for an artificial atom with the shape of a circular disk. The disk-shaped quantum dots we can fabricate are formed in a laterally grated submicrometre DBS, and contain a tunable number of electrons starting from zero. If the lateral confinement has the form of a harmonic potential, the eigenenergy, $E_{n,\ell}$, for the lateral states are labelled using two orbital quantum numbers: the radial quantum number, $n (= 0, 1, 2, \dots)$, and the angular momentum quantum number, $\ell (= 0, \pm 1, \pm 2, \dots)$, and has the form:

$$E_{n,\ell} = (2n + |\ell| + 1)\hbar\omega_0, \quad (9.1)$$

where $\hbar\omega_0$ is the lateral confinement energy. Here, we neglect the Zeeman effect, so that each state is spin degenerate. $E_{n,\ell}$ eigenstates (n, ℓ) are systematically degenerate, so including spin degeneracy, states at 0 T in the first, second and third shells respectively are two-fold degenerate with energy $E_{0,0}$ (1s orbital), four-fold degenerate with energy $E_{0,1} = E_{0,-1}$ (2p orbitals), six-fold degenerate with energy $E_{0,2} = E_{0,-2}$ (3d orbitals) = $E_{1,0}$ (3s orbital). For non-interacting electrons these states are consecutively filled from the lowest, and complete filling of each set of degenerate states is attained for special electron numbers of $N = 2, 6, 12, 20, \dots$. These are the ‘magic numbers’ that characterize the shell structure. For interacting electrons the degeneracy is lifted due to the Coulomb interactions. However, when the quantum mechanical confinement energy is comparable to, or greater than the interaction energy, the above shells are still consecutively filled from the lowest, so that we can define the same series of magic numbers as for the non-interacting case. In addition, for the filling of electrons in the same shell, parallel spins are favoured by Hund’s first rule. This leads to another series of magic numbers of $N = 4, 9, 16, \dots$, corresponding to the half filling of the second, third, fourth shells, respectively.

[Figure 9.3](#) shows the current oscillations observed for a dot in a typical circular mesa device (W) [4]. A small bias of 0.15 mV is set for V_{sd} , so that only the GSs contribute to the current. The absolute values of N can be identified for each zero-current region between the peaks, starting from $N = 0$, because for $V_g < -1.6$ V no further current peaks are observed, i.e. the dot is empty. When N becomes smaller than 20, the oscillation period depends strongly on N . In contrast, Coulomb oscillations observed for large dots containing more than 100 electrons look very periodic [7]. The current peak to the left of a Coulomb blockade region with N trapped electrons thus measures the electrochemical potential, $\mu(N)$, of the N electron GS. The peak spacing labelled by ‘ N ’

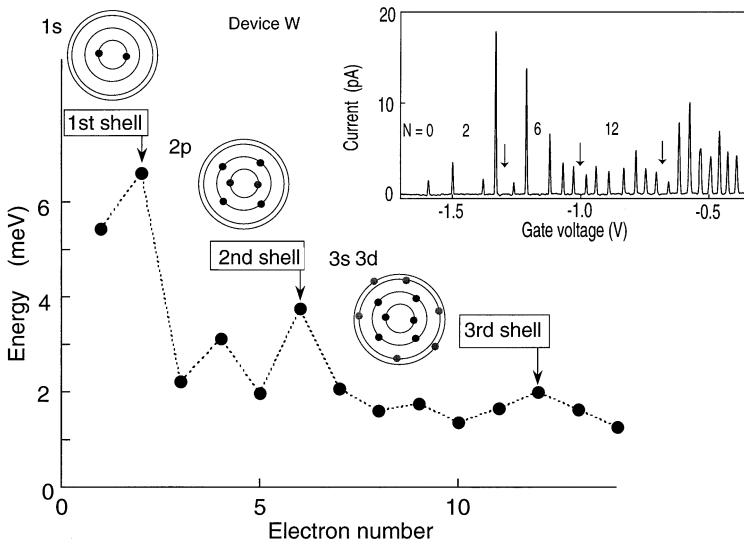


Figure 9.3. Change of the electrochemical potential, $\mu(N+1) - \mu(N)$, as a function of electron number, N [4]. Inset: Coulomb oscillations in current versus gate voltage at $B = 0$ T measured at 50 mK.

corresponds to the change of electrochemical potential, $\Delta_2(N) = \mu(N+1) - \mu(N)$. $\Delta_2(N)$, which can also be determined from measuring the widths of the so-called ‘Coulomb diamonds’, is plotted in figure 9.3. In correspondence with the pattern of spacings between the Coulomb oscillations, $\Delta_2(N)$ is unusually large for $N = 2, 6$ and 12 , and is also relatively large for $N = 4, 9$ and 16 (see arrows in figure 9.3). The values of $2, 6$ and 12 arise from the complete filling of the first, second and third shells, respectively, while those of $4, 9$ and 16 are due, respectively, to the half filling of the second, third and fourth shells with parallel spins (Hund’s first rule). The 2D shell structure is pictorially illustrated in figure 9.3.

9.3.2 Magnetic field dependence

In the presence of a magnetic field (B -field) parallel to the tunnelling current the electronic states are modified. This can be used to identify the quantum states in the dot. The eigenstates for a 2D harmonic potential are the Fock–Darwin (F–D) states [9]. The eigenenergies are expressed as:

$$E_{n,\ell} = -\frac{\ell}{2}\hbar\omega_c + \left(n + \frac{1}{2} + \frac{1}{2}|\ell|\right)\hbar\sqrt{4\omega_0^2 + \omega_c^2} \quad (9.2)$$

where $\hbar\omega_c = eB/m^*$ is the cyclotron frequency. Figure 9.4(a) shows $E_{n,\ell}$ versus B calculated for $\hbar\omega_0 = 3$ meV, which is a value deduced by a comparison with

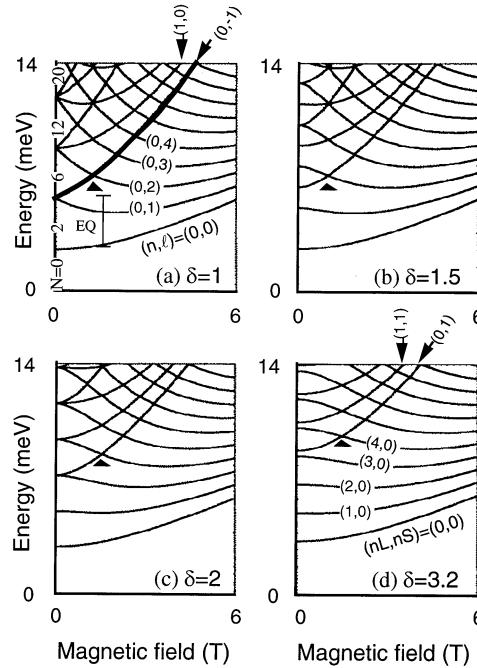


Figure 9.4. Calculated B -field dependence of the first ten single-particle energy levels for a circular dot with $\delta = 1$, and (a) for elliptical dots with $\delta = 1.5, 2$ and 3.2 , (b)–(d) [5].

the experimental data [4]. Spin is neglected so each state is two-fold degenerate. The orbital degeneracy at $B = 0$ T is lifted on increasing B , reflecting the first term in equation (9.2). As B is increased further, new crossings can occur. The last crossings occur along the bold line in figure 9.4(a). Beyond this crossing the F–D states correspond to the lowest Landau level.

Figure 9.5(a) shows the B -field dependence of the position of the current oscillations (see also figure 9.3) [4]. We take into account the interaction energy as well as the F–D states to interpret the experimental data. The current peaks generally shift in pairs with B . This pairing is due to the lifting of spin degeneracy. So from the shift of the paired peaks on increasing B , we assign quantum numbers to the respective pairs. For example, up to about 1 T, the lowest, second lowest, and third lowest pairs seem to correspond to the filling of electrons in the F–D states $(n, \ell) = (0, 0)$, $(0, -1)$, and $(0, 1)$ with antiparallel spins, respectively. Also, at higher B -field, the ‘wiggles’ or anticrossings between pairs of peaks correspond to the crossings of F–D states. For example, the anticrossing at the * label corresponds to the crossing of the F–D states $(0, -1)$ and $(0, 2)$, and the anticrossing at the ** label corresponds to the crossing of the F–D states $(0, -1)$ and $(0, 3)$ (see figure 9.4(a)). However, from close inspection to the second and

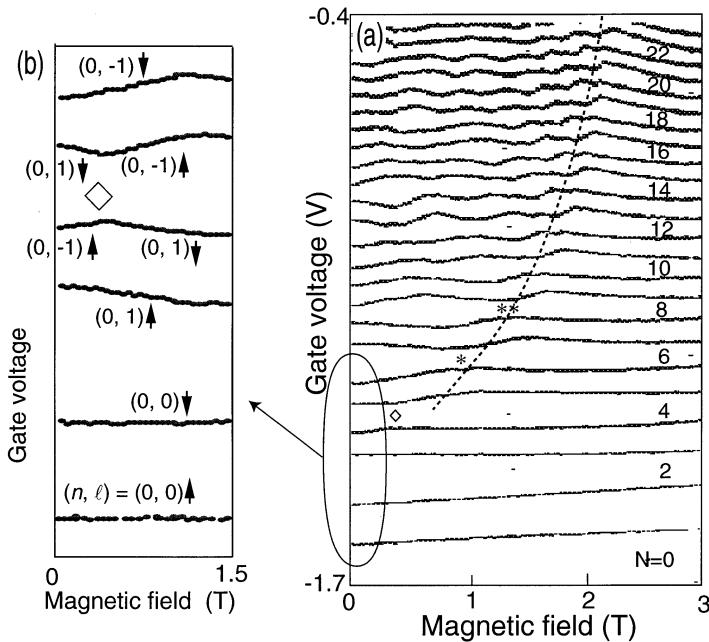


Figure 9.5. (a) B -field dependences of current peak positions measured for the device (W) whose Coulomb oscillations are shown in figure 9.3 [4]. (b) Magnified plot of the B -field dependences of the peak positions for the first six current peaks. The corresponding F-D states are indicated by (n, ℓ) quantum numbers, and by \uparrow or \downarrow , where \uparrow or \downarrow is a state with an up-spin or a down-spin, respectively.

third lowest pairs of peaks in the vicinity of $B = 0$ T, we find that the pairing is further modified in line with Hund's first rule. The details will be discussed in the next section.

The last anticrossing of each pair of peaks appears along the dotted curve in figure 9.5(a), which corresponds to the bold curve in figure 9.4(a). This dotted curve actually identifies the regime of filling factor $\nu = 2$. For $\nu < 2$, we see various other transitions associated with B -field enhanced Coulomb interactions, such as spin flip transitions between $\nu = 2$ and 1 [10], a spin single-triplet transition for $N = 2$ at $\nu = 1$ [11, 12], and the formation of the so-called maximum density droplet for $N > 2$ at $\nu = 1$ [13].

9.3.3 Spin triplet for the four-electron ground state

The pairing in the current peaks as observed in figure 9.5(a) arises from the antiparallel spin filling of orbital states. This can be modified when interaction effects determining spin filling cannot be neglected. Figure 9.5(b) shows a

magnified plot for the first six peaks evolving with magnetic field close to 0 T. The pairing of the third peak with the fifth, and the fourth peak with the sixth from 0 to 0.4 T, as opposed to the usual pairing of the third peak with the fourth, and the fifth peak with the sixth for $B > 0.4$ T, is a consequence of Hund's first rule, i.e. the $N = 4$ state is a spin-triplet so two parallel-spin electrons fill the two originally degenerate states $(n, \ell) = (0, 1)$ and $(0, -1)$ in the half-filled second shell [4, 7]. For $B > 0.4$ T, the fifth and sixth peaks, as a pair, first move up, reflecting the value of angular momentum quantum number $\ell = -1$. The spins of the added electrons are shown pictorially in figure 9.5(b). To explain why Hund's first rule gives a spin-triplet, we can introduce energy, E_{EX} , to represent the reduction in energy due to exchange between same two spin electrons in the half-filled second shell (estimated to be about 0.7 meV for W [4, 7]). The $N = 4$ triplet-state is thus lower in energy than the $N = 4$ singlet-state at 0 T by E_{EX} , and as a consequence $\Delta_2(3, 5) < \Delta_2(4)$ by about $2E_{\text{EX}}$. This effect persists as long as the splitting between states $(0, 1)$ and $(0, -1)$ is less than E_{EX} . At a critical B -field labelled by \diamond , this splitting exceeds E_{EX} , and the GS becomes a spin-singlet, i.e. there is a B -field induced triplet–singlet transition. This argument also implies that the first excited state (ES) is a spin-singlet for $B < 0.4$ T, whereas it is a spin-triplet for $B > 0.4$ T. We usually measure Coulomb oscillations to calculate only the GSs. To investigate the configurations of the GSs and ESs responsible for the triplet–singlet transition, we use an excitation spectroscopy technique [12].

Application of a bias voltage, V_{sd} , between the source and drain (typically a few meV) opens a ‘transport window’ between the Fermi energies for detecting both the N-electron GS and ESs in the dot (figure 9.6(a)). GSs and ESs lying within the transport window can contribute to current. If the gate voltage is made more positive, then the levels in the dot shift down in energy. When V_{sd} is smaller than the energy difference between the GS and lowest ES, only the GS contributes to the current, I , because the electron tunnelling into the ES is blocked. This is the usual case for measuring Coulomb oscillations, so we see just a series of current peaks as a function of gate voltage corresponding to the one-by-one change of electrons in the GSs of the dot (figure 9.3). When V_{sd} is sufficiently large, however, both the GSs and ESs can be within the transport window, and contribute to current. Electron tunnelling into the ESs can occur following the electron escape from a GS. I versus V_g therefore becomes a series of current ‘stripes’. Changes in the current level inside each stripe are measured when the ESs enter the transport window. The boundaries of the current stripe identify when the GS enters or leaves the transport windows, so the stripe has a width in energy given by eV_{sd} .

Figure 9.6(b) shows the spectrum for the $N = 3$ and 4 GSs and ESs measured at $V_{\text{sd}} = 1.6$ mV for a circular mesa device (not W). To emphasize variations in current we plot the derivative, dI/dV_g , in the plane of V_g and B . The dark curves identify the GSs and ESs. The white long-dashed curves trace the GSs, and the first and second ESs are indicated by the white dotted- and dotted-dashed curves, respectively. The $N = 3$ GS and first ES follow the B -field dependence of F–D

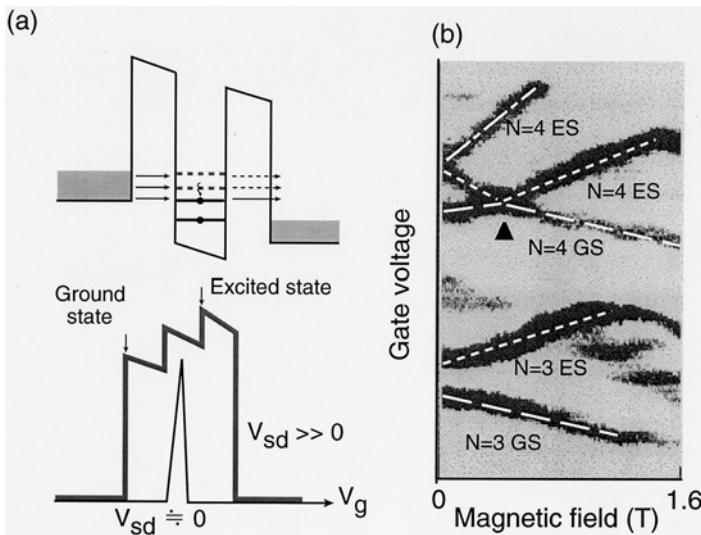


Figure 9.6. (a) Energy diagram and tunnelling current with a finite voltage is applied between the source and drain. (b) Derivative of the tunnelling current, dI/dV_g , as a function of gate voltage and magnetic field measured for the circular mesa (not W) with $V_{sd} = 1.6$ mV. Changes in the grey scale indicate changes in the current amplitude. The white long-dashed curves indicate the GSs, the first and second ESs are shown by the white dotted-, and chain curves.

states $(n, \ell) = (0, 1)$ and $(0, -1)$, respectively. On the other hand, the spectrum for $N = 4$ clearly shows the crossing of the F-D states $(0, 1)$ and $(0, -1)$ at 0.4 T, which causes the singlet-triplet transition in the GS. This observation fully supports our arguments described above on the effect of exchange to lower the energy of the triplet-state in the half-filled second shell [4, 12]. Note that the second ES for $N = 4$ is also evident, and corresponds to the state $(0, -1)$ in the F-D diagram.

9.4 Electronic properties of elliptical quantum dots

9.4.1 Deformed dots in rectangular mesa devices

For a rectangular mesa, the lateral confining potential of the dot is expected to be elliptical-like due to rounding at the corners provided the number of electrons in the dot is not too large, or too small [7, 14]. Assuming the confining potential is still perfectly parabolic, we choose to characterize the ‘ellipticity’ by a deformation parameter, $\delta = E_S/E_L$. Here, $E_S(E_L)$ is the confinement energy at 0 T along the minor (major) axis ($E_S > E_L$) as indicated in figure 9.2. The 2D states in the elliptical dot are labelled by the quantum numbers (n_L, n_S) , where

$n_L(n_S)$ is a quantum number ($= 0, 1, 2, \dots$) associated with the energy parabola along the major (minor) axis [5]. The energy of single-particle state (n_L, n_S) is $(n_L + 1/2)E_L + (n_S + 1/2)E_S$.

For a perfectly circular mesa, we trivially generalize our definition of δ so that $\delta = \beta = 1$. On the other hand, for the rectangular mesas, there is no simple correspondence between β , a ratio of lengths characteristic of the top metal contact, and δ , a ratio of energies characteristic of the dot in the mesa, and is dependent on V_g (or equivalently N). This means ‘accidental’ degeneracies will be hard to see over an extended range of N , and in any case may be lifted if the confinement potential is not completely parabolic. Nevertheless, although we are not saying that $\delta = \beta$, we assume that β is a ‘measure’ of δ , and thus one expects $\delta_Z > \delta_Y > \delta_X > \delta_W$.

Figure 9.4 shows the B -field evolution up to 6 T of the first 10 single-particle energy levels for a circular dot ($\delta = 1$) and elliptical dots ($\delta = 1.5, 2, 3.2$) calculated according to the model described by Madhav and Chakraborty [5] with a fixed confinement energy of $E_Q = 3$ meV. The confinement energies for the elliptical dots are simply derived from the relation $E_L E_S = E_Q E_Q$. For the circle and the $\delta = 3.2$ ellipse, quantum numbers (n, ℓ) and (n_L, n_S) , respectively, for some of the states are indicated.

As the deformation is gradually increased, figures 9.4(a)–(d), degeneracies of the single-particle states at 0 T are generally lifted, although accidental degeneracies can occur at certain ‘magic’ deformations, for example, (b) and (c), leading to subshell closures, provided the confining potential remains perfectly parabolic. The resulting patterns, however, are very different from that for the circular case, (a), and in practice may be hard to observe.

A weak B -field parallel to the current can also induce level degeneracies in both circular and elliptical dots when single-particle levels cross, but any shell structure at finite field is of a lower order and less apparent than that for the circle at 0 T [5]. As discussed before, ‘wiggles’ in the position of pairs of current peaks are expected at low field because of these crossings. The first lowest energy ‘wiggle’ originates from the crossing marked by \blacktriangle in figure 9.4, and this crossing moves to higher energy (N) with increasing δ . The spectra also highlight other simple points (a) the B -field lifts all degeneracies present at 0 T; (b) a B -field can always induce degeneracies at finite field when single-particle levels cross, provided the confinement potential is perfectly parabolic; and (c) as δ increases, the single-particle energy level spacing generally decreases ($\leq E_L$).

While illustrative, ultimately any modelling of real dots must go beyond a system of non-interacting electrons confined by a 2D harmonic potential, and include Coulomb interactions. Numerical diagonalization has been successfully employed to calculate basic electronic properties of dots with anisotropic confining potentials [6]. Such ‘exact’ calculations, however, are limited to $N \leq 6$. In order to study dots containing a larger number of electrons, spin-density functional theory is a powerful technique, which explicitly incorporates the electron–spin interactions, and has led to a number of interesting predictions

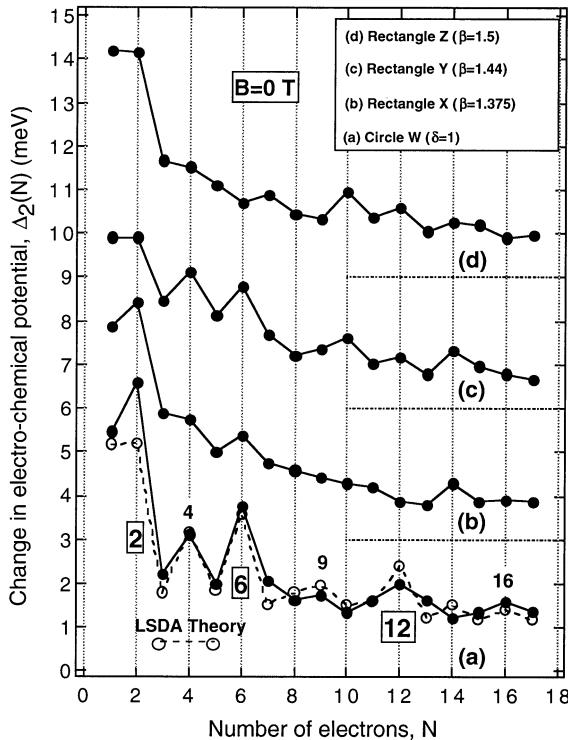


Figure 9.7. $\Delta_2(N)$, for the circular mesa W, (a) and the three rectangular mesas X, Y, and Z, (b)–(d) [7, 14]. The traces are offset vertically by 3 meV for clarity. The fit for W is given by a local spin-density approximation (LSDA) (see figure 9.10).

for the GS structure of dots [14, 15]. Both approaches predict changes in the addition energy spectra, and transitions in the spin-states as deformation is varied. An example of the latter is the breakdown of the conditions for which Hund's first rule gives a triplet-spin state for $N = 4$, and this is marked by a transition from a spin-triplet to a spin-singlet configuration at some critical deformation [6, 14].

We first discuss the energy spectra at 0 T measured for devices with different β -values. We measure the Coulomb oscillations, and derive the change in the electrochemical potential, $\Delta_2(N)$, for each device. The results are shown in figure 9.7 as a function of electron number, N , up to $N = 17$ for (a) W, (b) X, (c) Y and (d) Z at 0 T [7, 14]. The data for W is the same as that shown in figure 9.2. The result from a local spin-density approximation (LSDA) calculation is also included [14, 15]. The 2D shell structure including the effects of Hund's first rule is well reproduced by the calculation.

For circle W, as N is decreased, $\Delta_2(N)$ becomes generally larger due to the increase of the Coulomb interaction when a dot is ‘squeezed’. This is also true for

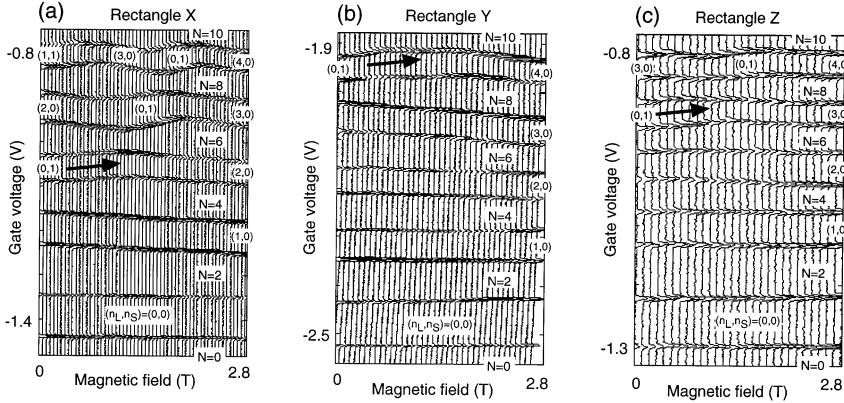


Figure 9.8. B -field dependence of the Coulomb oscillation peak positions for the mesas X, Y and Z, (a)–(c). The B -field is parallel to the current [14].

the rectangular mesas, but now there are no prominent maxima at $\Delta_2(2, 6, 12)$. The shell structure for the circular dot has now become disrupted or ‘smeared out’, and this is attributed directly to the lifting of the degeneracies of the single-particle states present in a circular dot [4, 7]. Deformation kills the shell structure for a circle, and even quite a small deformation can make a big difference [6, 14]. This is evident from traces, (b)–(d) in figure 9.7. The effect of deformation is discussed later in a comparison with model calculations.

9.4.2 Magnetic field dependence

Figure 9.8 shows the B -field dependence, for a weak field applied parallel to the current, of the Coulomb oscillation peak positions for the rectangular mesas X, Y and Z, (a)–(c) [14]. The data consists of current versus V_g traces taken at a very small bias at different B -fields. Peaks are paired, and there are no obvious deviations close to 0 T for $N = 4$ due to exchange effects. Quantum numbers (n_L, n_S) of the single-particle states are assigned, and the first upmoving pair of peaks is marked by a thick arrow. With increasing δ , from the calculated spectra in figure 9.4, both the first up-moving pair of peaks, and the lowest single-particle level crossing (identified in figure 9.4 by \blacktriangle) are simply expected to move systematically to higher N (or equivalently to higher energy) [5, 14].

For the elliptical dots, normal peak pairing, even from 0 T, occurs so the spin-state for $N = 4$ is a singlet. The exchange effect is maximal for a circular dot at $N = 4$ because the $(n\ell) = (0, 1)$ and $(0, -1)$ states are degenerate at 0 T and their wavefunctions have the same symmetry, but with deformation these states become the $(n_L, n_S) = (1, 0)$ and $(0, 1)$ states in an elliptical dot, which are split at 0 T. This energy splitting, γ , increases with δ . If $\gamma < E_{\text{EX}}$ at 0 T, exchange can still operate to lower the energy, and thus the $N = 4$ GS remains a spin-

triplet. On the other hand, if $\gamma > E_{\text{EX}}$ at 0 T, the energy gain due to exchange is not sufficiently large to compensate for the splitting, so normal pairing occurs. Thus, we can expect a triplet–singlet transition at some critical deformation. The absence of deviations to the normal peak pairing at $N = 4$ in figures 9.8(a)–(c) apparently confirms that δ is greater than 1.2–1.3 in line with calculations we present shortly (see figure 9.10) [6, 14].

A striking feature about figures 9.8(a)–(c) is the position of the first up-moving pair of peaks. For X, Y, and Z respectively, it is the third, fifth and fourth pair of peaks. As revealed by figure 9.4, in a simple single-particle picture, the first up-moving state is $(n_L, n_S) = (0, 1)$ [5, 14]. Inspection of these spectra shows that this state is, in the weak-field limit, from the bottom, the third, fourth, and fifth state respectively for $1 \leq \delta < 2$, $2 \leq \delta < 3$, $3 \leq \delta < 4$. Thus, starting from $\delta = 1$, the first up-moving pair of peaks should go from the third to fourth, fourth to the fifth, ... at certain ‘magic’ deformations as δ is increased. Given that Coulomb effects are neglected in this simple picture, and that in practice δ is expected to vary with N , it looks as if $1 < \delta < 2$ for X, $3 < \delta < 4$ for Y, $2 < \delta < 3$ for Z. Although ellipses X, Y, and Z are deformed beyond the triplet–singlet transition, we must conclude the following: (a) δ can be much higher than that suggested by the β values; and (b) the ordering given by increasing β values may not always reflect the true ordering in δ , i.e. deformation in Y seems to be stronger than in Z, so the true sequence may be W–X–Z–Y for the four mesas. The former is not so unexpected since we have no independent way of measuring δ , but the latter is perhaps more surprising.

9.4.3 Study of Zeeman effect on the spin states

From studying the addition energy spectra of the rectangular mesa devices, we find that in the presence of an asymmetric lateral confinement, the shell structure is readily disrupted, and that a spin singlet is favoured for the $N = 4$ GS at 0 T. However, neither the addition energy spectra, nor the B -field dependence of Coulomb oscillation peaks give *direct* information about the spin states, or whether the state is a singlet or triplet. For example, at $N = 4$ the change of the electrochemical is large both for the circular mesa W and rectangular mesa Y, although the $N = 4$ GS is a triplet for the former (in line with Hund’s first rule), and is a singlet for the latter, as deduced *indirectly* by applying a B -field parallel to the current. To see the effect of deformation on the spin state we measure the Zeeman effect by applying a B -field perpendicular to current [16]. The electrochemical potential for the N -electron GS $\mu(N)$, decreases or increases with B -field, depending on whether the *change* of the total spin is positive or negative, respectively, on addition of an extra electron to the $N - 1$ -electron GS. The B -field dependence of $\mu(N)$ is given by,

$$\mu(N, B) - \mu(N, B = 0 \text{ T}) (\equiv \Delta\mu_N(B)) = g(S_N - S_{N-1})\mu_B B, \quad (9.3)$$

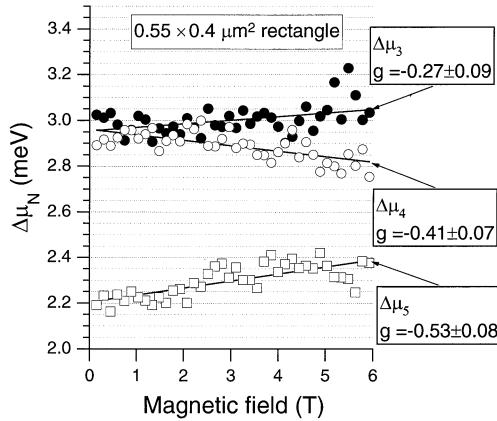


Figure 9.9. Zeeman shift of the levels, μ (3), μ (4) and μ (5), for the elliptical dot in the $0.55 \times 0.4 \mu\text{m}^2$ rectangular mesa (X) [16]. The straight lines are fits to the data, which are used to estimate the electronic g -factor.

where g is the electronic g -factor, and S_N is the total spin of the N -electron GS. Let us now consider the case of $N = 4$. When the GS of the half-filled second shell is a spin-singlet, $\Delta\mu_N(B)$ is given by,

$$\begin{aligned}\Delta\mu_3(B) &= \Delta\mu_3(0) - g\mu_B B, \\ \Delta\mu_4(B) &= \Delta\mu_4(0) - g\mu_B B,\end{aligned}$$

and

$$\Delta\mu_5(B) = \Delta\mu_5(0) - g\mu_B B. \quad (9.4)$$

On the other hand, when the $N = 4$ spin state is a triplet,

$$\begin{aligned}\Delta\mu_3(B) &= \Delta\mu_3(0), \\ \Delta\mu_4(B) &= \Delta\mu_4(0) - g\mu_B B,\end{aligned}$$

and

$$\Delta\mu_5(B) = \Delta\mu_5(0). \quad (9.5)$$

Note that a plot of $\Delta\mu_N(B)$ versus B is useful as we can also correct for the influence of background charge on the Coulomb oscillation peaks which sometimes causes the peaks to drift a little during the course of the measurement.

Figure 9.9 shows the B -field dependence of $\Delta\mu_N$ measured for $N = 3, 4$, and 5 for the elliptical dot in mesa X ($0.55 \times 0.4 \mu\text{m}^2$) [16]. The B -field up to 6 T is applied perpendicular to current. $\Delta\mu_3$ and $\Delta\mu_5$ increase with B -field, and $\Delta\mu_4$ decreases. The increase of $\Delta\mu_3$ with B -field indicates that the third and fourth electrons to enter the dot have opposite spins, i.e. the spin-state is a singlet. The decrease of $\Delta\mu_4$, and increase of $\Delta\mu_5$ indicate that the fifth and

sixth electrons also enter spin-up, and spin-down respectively into the next state. As a consequence, we attribute the relatively large $\Delta_2(4)$ for the elliptical dots (see figure 9.7(b)–(d)) to the lifting of the degeneracy of the single-particle states originally present in the circular dot. In figure 9.9, we estimate the electron g -factor by fitting a straight line to the data. The estimated g -factor ranges from -0.27 to -0.54 , and this is comparable to the value for bulk $\text{In}_x\text{Ga}_{1-x}\text{As}$ for small x [17].

9.5 Comparison to model calculations

We apply the methods of spin-density functional theory (SDFT) to model the changes to the shell structure due to the deformation of dot's lateral confinement at 0 T. To obtain the GS energies and densities for N electrons confined in an externally imposed potential, we solve the spin-dependent single-particle Kohn–Sham (KS) equations [18],

$$\left[-\frac{\hbar^2}{2m^*} \nabla_r^2 + V_{\text{eff}}^\sigma(r) \varphi_{i,\sigma}(r) - \right] = \varepsilon_{i,\sigma} \varphi_{i,\sigma}(r), \quad (9.6)$$

in a plane-wavebasis to avoid any symmetry restrictions. In equation 9.6, the index σ accounts for the spin (\uparrow or \downarrow), and $\mathbf{r} = (x, y)$. The effective mean-field potential, $V_{\text{eff}}^\sigma(r)$, contains contributions from the external harmonic confining potential, the Hartree potential of the electrons, and the functional derivative of the local exchange-correlation energy, for which we use the approximation of Tanatar and Ceperley [19] (see also [15, 20] for details). The electrostatic confinement due to the lateral depletion region imposed by the side wall and the Schottky gate is approximated by a 2D anisotropic harmonic potential with frequencies $\omega_x = \omega/\sqrt{\delta}$ and $\omega_y = \omega/\sqrt{\delta}$,

$$V_{\text{ext}}(x, y) = \frac{1}{2} m^* \omega^2 \left(\delta x^2 + \frac{1}{\delta} y^2 \right). \quad (9.7)$$

The ratio of the oscillator frequencies, $\delta = \omega_x/\omega_y$, thus defines the ratio of the major and minor axes of the ellipsoidal equipotentials. We impose the constraint, $\omega^2 = \omega_x \omega_y$, which is equivalent to conserving the area of the quantum dot with deformation [20]. The x - and y -axes are indicated in figure 9.2 with

above-defined E_S and E_L , respectively, correspond to $\hbar\omega_x$ and $\hbar\omega_y$. The strength, ω , of the external parabolic confinement leading to an average particle density, $n_0 = 1/(\pi r_s^2)$, in a circular dot is approximated by $\omega^2 = e^2/(4\pi\epsilon_0\epsilon m^* r_s^3 \sqrt{N})$ [15]. Minimizing the energy density functional by self-consistently solving the above KS equations, GS energies are obtained for different electron numbers and deformation parameters. Full technical details are given elsewhere [15, 20], and here we report only the results. We emphasize that from recent measurements, it is clear that as N increases the confinement weakens

sixth electrons also enter spin-up, and spin-down respectively into the next state. As a consequence, we attribute the relatively large $\Delta_2(4)$ for the elliptical dots (see figure 9.7(b)–(d)) to the lifting of the degeneracy of the single-particle states originally present in the circular dot. In figure 9.9, we estimate the electron g -factor by fitting a straight line to the data. The estimated g -factor ranges from -0.27 to -0.54 , and this is comparable to the value for bulk $\text{In}_x\text{Ga}_{1-x}\text{As}$ for small x [17].

9.5 Comparison to model calculations

We apply the methods of spin-density functional theory (SDFT) to model the changes to the shell structure due to the deformation of dot's lateral confinement at 0 T. To obtain the GS energies and densities for N electrons confined in an externally imposed potential, we solve the spin-dependent single-particle Kohn–Sham (KS) equations [18],

$$\left[-\frac{\hbar^2}{2m^*} \nabla_r^2 + V_{\text{eff}}^\sigma(r) \varphi_{i,\sigma}(r) - \right] = \varepsilon_{i,\sigma} \varphi_{i,\sigma}(r), \quad (9.6)$$

in a plane-wavebasis to avoid any symmetry restrictions. In equation 9.6, the index σ accounts for the spin (\uparrow or \downarrow), and $\mathbf{r} = (x, y)$. The effective mean-field potential, $V_{\text{eff}}^\sigma(r)$, contains contributions from the external harmonic confining potential, the Hartree potential of the electrons, and the functional derivative of the local exchange-correlation energy, for which we use the approximation of Tanatar and Ceperley [19] (see also [15, 20] for details). The electrostatic confinement due to the lateral depletion region imposed by the side wall and the Schottky gate is approximated by a 2D anisotropic harmonic potential with frequencies $\omega_x = \omega/\sqrt{\delta}$ and $\omega_y = \omega/\sqrt{\delta}$,

$$V_{\text{ext}}(x, y) = \frac{1}{2} m^* \omega^2 \left(\delta x^2 + \frac{1}{\delta} y^2 \right). \quad (9.7)$$

The ratio of the oscillator frequencies, $\delta = \omega_x/\omega_y$, thus defines the ratio of the major and minor axes of the ellipsoidal equipotentials. We impose the constraint, $\omega^2 = \omega_x \omega_y$, which is equivalent to conserving the area of the quantum dot with deformation [20]. The x - and y -axes are indicated in figure 9.2. With this convention, the above-defined E_S and E_L , respectively, correspond to $\hbar\omega_x$ and $\hbar\omega_y$. The strength, ω , of the external parabolic confinement leading to an average particle density, $n_0 = 1/(\pi r_s^2)$, in a circular dot is approximated by $\omega^2 = e^2/(4\pi\epsilon_0\epsilon m^* r_s^3 \sqrt{N})$ [15]. Minimizing the energy density functional by self-consistently solving the above KS equations, GS energies are obtained for different electron numbers and deformation parameters. Full technical details are given elsewhere [15, 20], and here we report only the results. We emphasize that from recent measurements, it is clear that as N increases the confinement weakens

in such a way that the particle density tends to a constant [21]. This is implicit in our model, as for any given value of r_s , the oscillator frequency ω , and the related frequencies ω_x and ω_y , decrease with increasing N . δ is also kept constant for simplicity, although δ is expected to vary with N in practice.

There are no fitting parameters in the equations, and only a suitable choice of r_s is required to generate the addition energy spectra. The value of $r_s = 1.5a_B^*$ used in the model calculations is realistic as the value estimated experimentally for a circular quantum dot is 1.3 to $1.4a_B^*$ [21]. $a_B^* = \hbar^2(4\pi\varepsilon_0\varepsilon)/m^*e^2$ is an effective atomic unit, which for GaAs is about 10.3 nm. $r_s = 1.5a_B^*$ in the model presented here corresponds to an effective confinement energy, E_Q , for $N = 1$ of about 5.7 meV. This value is consistent with the upper limit of E_Q observed in practice (about 5 meV [12]), and justifies the $E_Q = 3$ meV value as a reasonable average for calculating the simple single-particle spectra shown in figure 9.4 for the first 10 levels.

Figure 9.10 shows $\Delta_2(N)$, calculated for quantum dots with various values of δ [14]. The lowest trace is the calculation for the circular dot, and reproduces well the familiar shell structure for a 2D harmonic oscillator. The agreement with experiment is strikingly good (see also figure 9.7(a)). Deforming the confinement slightly by changing the deformation parameter to $\delta = 1.1$ (trace (b) in figure 9.10), the calculation still predicts fairly clear shell closures at $N = 2$, 6 and 12. These numbers can still be considered as ‘magic’, but the actual values of $\Delta_2(2, 6, 12)$ are noticeably suppressed, because degeneracies have been lifted [7, 14]. As the deformation increases further, the peaks for $N = 2, 6, 12$, and 20 are further suppressed. This is a simple consequence of the removal of the level ‘bunching’ with deformation. Even for the cases where ‘accidental’ subshell closures occur at certain ‘magic’ deformations, for example, $\delta = 1.5$, and 2 as seen in figure 9.4, the reduced separation between degenerate single-particle energy levels (E_L) would make any shell structure less clear to observe, and the sequence of ‘magic’ numbers would be very different (e.g. for $\delta = 2$ it would be 2, 4, 8, 12, 18, ...) compared to those for $\delta = 1$.

Also apparent is that a systematic detailed one-to-one correspondence of $\Delta_2(N)$ between traces (b)–(d) in figure 9.7 and traces (b)–(f) in figure 9.10 is impossible to make. Although the experimental data for mesa X partly resembles the theoretical data for $\delta = 1.1$ to 1.3, the data for mesas Y and Z do not seem to resemble that for $\delta > 1.3$, except perhaps for a weak tendency to oscillate between even- N and odd- N .

Theoretically, figure 9.10 shows that there are transitions in the GS spin-configurations with deformation [20]. The total spin, S , is identified by different symbols in the figure. These transitions are particularly numerous for, but are not restricted to, the even- N systems, and are clearly very sensitive to the actual value of the deformation. For example, in the case of $N = 6$, the total spin is predicted to change from $S = 0$ (i.e. a paramagnetic state) at $\delta = 1$, through an $S = 0$ spin-density wave (SDW) state [15, 20], to $S = 1$ at $\delta = 1.5$: an indication of ‘piezo-magnetic’ behaviour [20, 22], i.e. changes of the dot magnetization with

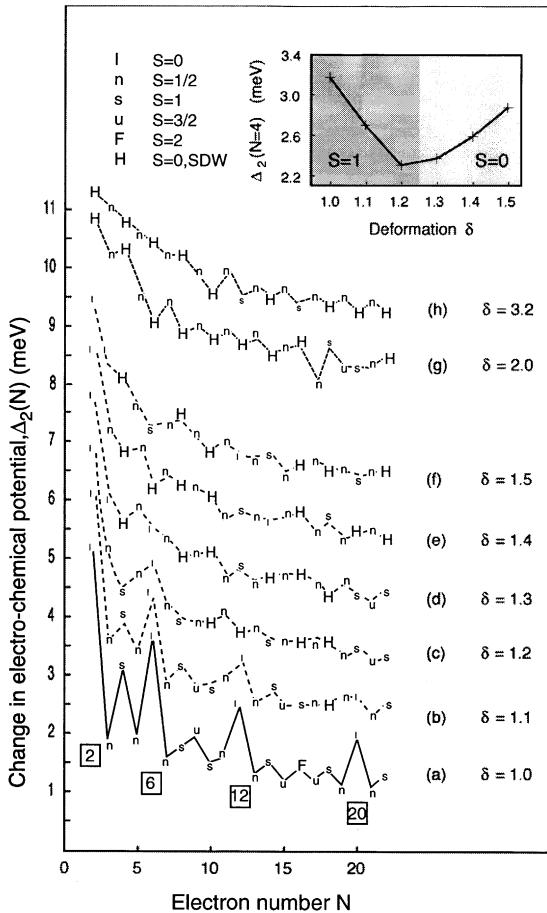


Figure 9.10. Model calculations for the change in electrochemical potential within spin-density functional theory [14]. The different traces correspond to zero, weak and moderate deformation parameters $\delta = 1.0$ to 1.5 , (a)–(f), and higher deformation parameters $\delta = 2$ and $\delta = 3.2$, (g) and (h). The traces are offset vertically by 1 meV for clarity, and there is an additional 1 meV offset between traces (f) and (g). The total spin, S , for different deformations and electron numbers are identified by different symbols as defined in the figure. The inset shows $\Delta_2(N = 4)$ versus δ .

deformation. Although experimentally we are not in a position to differentiate between an $S = 0$ ‘normal’ state and an $S = 0$ SDW state with broken spin symmetry in the internal coordinates [23]—indeed the interpretation of a SDW is still debated in the literature [24]—the SDFT calculations described here predict that the latter becomes more prevalent for even- N systems as δ increases,

particularly for small average particle densities [15, 20].

Another interesting case is what happens to the $N = 4$ GS. The inset in figure 9.10 shows $\Delta_2(N = 4)$ versus deformation up to $\delta = 1.5$. Starting with the circular dot, Hund's first rule gives a total spin of $S = 1$ for the triplet state favouring spin alignment of the two electrons in the second shell rather than a total spin of $S = 0$ for the singlet state in which the spins are paired. As the deformation is initially increased, the energy separation between the two levels (n_L, n_S) = (1, 0) and (0, 1)—the two originally degenerate levels ($n\ell$) = (0, 1) and (0, -1) in the second shell of the circular dot—increases (see (a) and (b) in figure 9.4), and so the spin-triplet state becomes progressively less favourable. $\Delta_2(4)$ continuously decreases with δ , and at a value between 1.2 and 1.3, a spin-zero state (actually predicted by the SDFT described here to be a SDW) appears. For higher values of δ , $\Delta_2(4)$ starts to increase. Other recent calculations employing numerical diagonalization for elliptical dots have also revealed a spin triplet–singlet transition at about $\delta = 1.2$ for $N = 4$ [6].

Inspection of figure 9.7 gives values of $\Delta_2(4)$ for mesas W, X, Y, and Z respectively of 3.1, 2.7, 3.1, and 2.5 meV. Whilst it is reassuring that these energies lie in the range predicted by SDFT, it is tempting to say that the trend for mesas W, X, and Y is consistent with that predicted in figure 9.10, i.e. $N = 4$ is a spin-triplet for W, and a spin-singlet for X, and Y, and attribute the apparently anomalously low value for Z to sample specific fluctuations [4].

For mesas Y and Z, the $\Delta_2(N)$ traces in figure 9.7 seem to show a weak tendency to oscillate between a slightly larger even- N value, and a slightly smaller odd- N value, and this oscillation seems clearer for Y than for Z. For the moment we do not try to account for the clarity of this oscillation in dots Y and Z, but try to explain the origin of the oscillation, although we are now being forced to entertain the idea that δ for Y and Z may be much larger than 1.5. Starting from the over simple single-particle picture with a fixed confinement energy, and then including a constant interaction which is the same for even- N and odd- N , a larger even- N value is expected, because only Δ_2 (even- N) can contain a finite contribution due to the single-particle energy level spacing. A slightly more advanced model, which is more realistic in principle, would be to have a constant interaction for odd- N (electron added to an $S = 1/2$ state already containing one electron) that is stronger than the constant interaction for even- N (electron added to an empty state). If the former is larger than the latter plus the single-particle spacing (more likely in practice as N increases), a weak tendency to oscillate between smaller even- N and larger odd- N could occur. This pattern is what the SDFT calculations predict in figure 9.10 for $\delta = 2$ and $\delta = 3.2$. The fact that $\Delta_2(N)$ for Y and Z is often a little larger for even- N than odd- N should not be taken to mean that the constant interaction model is more accurate. Rather the Coulomb interactions may not be so strong in practice, due to screening by the leads for example, as those in our model—and model that also does not include the self-consistent calculation of the electrostatic confining potential. Indeed, in the SDFT calculations of Lee *et al* [25], the electrostatic confining potential is much stronger

(e.g. $E_S = 20$ meV, $E_L = 10$ meV), and they find that $\Delta_2(N)$ is generally a little larger for even- N than for odd- N .

9.6 Conclusions

We have experimentally and theoretically investigated the electronic properties of circular and ellipsoidally deformed quantum dots containing a few electrons. Associated with the cylindrical symmetry and the parabolicity of the lateral confinement potential, atom-like properties such as a clear shell structure, and filling in accordance with Hund's first rule are all observed at zero-magnetic field for the circular dot. These atom-like features, as identified in the addition energy spectrum, are readily disrupted in the elliptical dots even for a small anisotropy breaking circular symmetry. Application of a magnetic field parallel to the current leads to a systematic modification of the energy spectrum for the circular dot, reflecting the Fock–Darwin diagram and normal antiparallel spin filling of the same orbital states. At zero-magnetic field, parallel spin filling is only favoured in line with Hund's first rule when each shell is half-filled. For the second shell, this is well identified from measurement of excitation spectra, i.e. the $N = 4$ ground state is a spin-triplet. In contrast, for the elliptical dots, the $N = 4$ ground state is a spin-singlet due to the lifting of the single-particle level degeneracies. This spin-singlet is confirmed by measuring the Zeeman effect in an elliptical dot. These observations on the effects of deformation are in good agreement with recent theories, as well demonstrated here by the application of spin-density functional theory at zero-magnetic field. Interestingly, this theory predicts the existence of $S = 0$ SDW states in strongly deformed dots. This has not yet been clarified in the present set of experiments, although the magnetic field dependence strongly suggests that the anisotropy of an elliptical dot can be significantly higher than that expected from the geometry of the mesa in which the dot is located. To investigate the possible existence of SDW states, *in situ* manipulation of the lateral potential geometry of a quantum dot is highly desirable, and this may be achieved by fully exploiting a multiple-gated vertical single electron transistor we have recently developed [26].

Acknowledgments

We would like to acknowledge T Honda, R van der Hage, J Janssen, T Oosterkamp, H Tamura, Y Tokura and T Uesugi for their help and useful discussions. This work is partly supported by the Specially Promoted Research, grant-in-aid for scientific research from the Ministry of Education, Science and Culture in Japan, the NEDO joint research program (NTDP-98), the Dutch Organization for Fundamental Research on Matter (FOM) and the Academy of Finland, and the TMR program of the European Community under contract ERBFMBICT972405.

References

- [1] For reviews, Altshuler B L, Lee P A and Webb R A (ed) 1991 *Mesoscopic Phenomena in Solids* (Amsterdam: Elsevier)
Meirav U and Foxman E B 1996 *Semicond. Sci. Technol.* **11** 255
Sohn L L, Kouwenhoven L P and Schoen G (ed) 1997 *Proc. NATO Advance Study Institute on Mesoscopic Electron Transport (Kluwer Series E345)* (Dordrecht: Kluwer)
- [2] Reed M 1993 *Sci. Am.* **268** 118
Kastner M A 1993 *Phys. Today* **46** 24
Ashoori R C 1996 *Nature* **379** 413
- [3] Austing D G, Honda T and Tarucha S 1996 *Semicond. Sci. Technol.* **11** 388
- [4] Tarucha S, Austing D G, Honda T, van der Hage R J and Kouwenhoven L P 1996 *Phys. Rev. Lett.* **77** 3613
- [5] Madhav A V and Chakraborty T 1994 *Phys. Rev. B* **49** 8163
- [6] Ezaki T, Sugimoto Y, Mori N and Hamaguchi C 1998 *Semicond. Sci. Technol.* **13** A1
- [7] Tarucha S, Austing D G, Honda T, van der Hage R J and Kouwenhoven L P 1997 *Japan. J. Appl. Phys.* **36** 3917
- [8] Alonso M and Finn E J 1968 *Quantum and Statistical Physics* (Reading, MA: Addison-Wesley)
- [9] Fock V 1928 *Z. Phys.* **47** 446
Darwin C G 1930 *Proc. Camb. Phil. Soc.* **27** 86
- [10] McEuen P L, Foxman E B, Kinaret J, Meirav U, Kastner M A, Wingreen N S and Wind S J 1992 *Phys. Rev. B* **45** 11419
McEuen P L, Wingreen N S, Foxman E B, Kinaret J, Meirav U, Kastner M A, Meier Y and Wind S J 1993 *Physica B* **189** 70
For our disk-shaped dot see Ausing D G, Tokura Y, Honda T, Tarucha S, Danoesastro M, Janssen J, Oosterkamp T H and Kouwenhoven L P 1999 *Japan. J. Appl. Phys.* **38** 372
- [11] First proposal is given by Wagner M, Merkt U, Chaplik A V 1992 *Phys. Rev. B* **45** 1951
For experimental evidence see Su B, Goldman V J and Cunningham J E 1992 *Phys. Rev. B* **46** 7644
Ashoori R C, Stormer H L, Weiner J S, Pfeiffer L N, Baldwin K W and West K W 1993 *Phys. Rev. Lett.* **71** 613
- [12] Kouwenhoven L P, Oosterkamp T H, Danoesastro M W S, Eto M, Austing D G, Honda T and Tarucha S 1997 *Science* **278** 1788
- [13] MacDonald A H 1993 *Aust. J. Phys.* **46** 345
For our disk-shaped dot see Oosterkamp T H, Janssen J W, Kouwenhoven L P, Austing D G, Honda T and Tarucha S 1999 *Phys. Rev. Lett.* **82** 2931
- [14] Austing D G, Sasaki S, Tarucha S, Reimann S M, Koskinen M and Manninen M 1999 *Phys. Rev. B* submitted
- [15] Koskinen M, Manninen M and Reimann S M 1997 *Phys. Rev. Lett.* **79** 1389
- [16] Sasaki S, Austing D G and Tarucha S 1998 *Physica B* **256–258** 157
- [17] Wimbauer Th., Oettinger K, Efros Al. L, Meyer B K and Brugger H 1994 *Phys. Rev. B* **50** 8889
- [18] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
- [19] Tanatar B and Ceperley D M 1989 *Phys. Rev. B* **39** 5005

- [20] Reimann S M, Koskinen M, Kolehmainen J, Manninen M, Austing D G and Tarucha S 1999 *Eur. J. Phys.* D to appear
- [21] Austing D G, Tokura Y, Honda T, Tarucha S, Danoesastro M, Janssen J, Oosterkamp T H and Kouwenhoven L P 1999 *Japan. J. Appl. Phys.* **38** 372
- [22] Reimann S M, Koskinen M and Manninen M 1999 *Phys. Rev. B* **59** 1613
- [23] Ring P and Schuck P 1980 *The Nuclear Many-Body Problem* (New York: Springer)
- [24] Hirose K and Wingreen N S 1999 *Phys. Rev. B* **59** 4604
- [25] Lee I-H, Rao V, Martin R M and Leburton J-P 1998 *Phys. Rev. B* **57** 9035
- [26] Austing D G, Honda T and Tarucha S 1998 *Physica E* **2** 583
Austing D G, Honda T and Tarucha S 1997 *Semicond. Sci. Technol.* **12** 631

PART 4

OPTICAL PROPERTIES

Chapter 10

Electron–hole and exciton systems in low dimensions

Tetsuo Ogawa

Department of Physics, Tohoku University, Aoba-ku, Sendai 980-8578, Japan and ‘Structure and Transformation’, PRESTO, Japan Science and Technology Corporation, Sendai 980-8578, Japan

E-mail: ogawa@cmpt01.phys.tohoku.ac.jp

Optical responses of low-dimensional semiconductors are reviewed from a theoretical viewpoint, stressing the electron–hole Coulomb interaction and the excitonic effects. Recent progress of the theoretical studies is highlighted and important problems to be solved are introduced. Dimensionality of the Wannier exciton in particular, is discussed in terms of its linear and nonlinear responses. Theories for a few exciton systems, high-density degenerate electron–hole systems, and a Mahan exciton in a one-dimensional metal are also given.

10.1 Introduction

Stimulated by recent progress in fabrication techniques for low-dimensional semiconductors, geometrical confinement effects on their transport and optical properties have been attracting much attention. In particular, the optical characteristics of such materials, which reflect higher excited states and the relaxation effects, are of special importance not only in terms of fundamental condensed-matter physics [1] but also photonic device applications. In optical responses of semiconductors or insulators, an electron and a hole excited respectively in a conduction and a valence band, or their composite particle called an exciton, play central roles. In low-dimensional structures, the Coulomb correlation effects or the excitonic effects become more prominent than in bulk structures, leading to peculiar optical characteristics combined with the

geometrical confinement effects. In this paper, optical responses of low-dimensional semiconductors are reviewed from a theoretical viewpoint, stressing the electron–hole Coulomb interaction and the excitonic effects. Recent progress of the theoretical studies is highlighted and important problems to be solved are introduced.

This paper is organized as follows. In section 10.2, fundamental remarks on low-dimensional electron–hole systems are summarized. Section 10.3 is devoted to the single-exciton problems including linear and nonlinear optical responses of the Wannier exciton in low dimensions. Nonlinear optical quantities (e.g., two-photon absorption rates) are stressed in terms of a probe for determining the exciton's dimensionality. Dielectric confinement effects are also introduced there. In section 10.4, a few exciton problems are discussed. Theories of $\chi^{(3)}$ excitonic nonlinearity are introduced in terms of the fermionic or bosonic picture. High-density degenerate electron–hole systems and a Mahan exciton in a one-dimensional metal are also studied in section 10.5 as examples of the excitonic many-body problems.

10.2 Remarks on low-dimensional excitons

When we define an exciton as a bound state of an electron in a conduction band and a hole in a valence band¹, there are two limiting types of exciton: the Wannier exciton and the Frenkel exciton. Generally speaking, the wavefunction of the electron–hole relative motion of the Wannier exciton is more sensitively affected by a spatial geometry than that of the Frenkel excitons. Thus we will review theoretical problems related to the electron–hole relative motion of a Wannier exciton confined in one [2] or two dimensions.

The Wannier exciton has three degrees of freedom: a centre-of-mass motion, an electron–hole relative motion, and a spin configuration, where the latter two are internal degrees of freedom. In perfect rigid crystals in d dimensions ($d = 1, 2, 3$), the centre-of-mass motion is well described by a plane wave with a d -dimensional wavevector \mathbf{K} . Since only the centre-of-mass state of $\mathbf{K} \sim 0$ contributes to the optical responses, the electron–hole relative motion and the subband (sublevel) structures mainly determine the optical properties of low-dimensional exciton systems. In the following, we consider a quasi-two-dimensional ($d \simeq 2$) and a quasi-one-dimensional ($d \simeq 1$) semiconductors² with a dipole-allowed direct band gap $E_{\text{gap}}^{(d)}$ ³, which consists of a single conduction

¹ Although an exciton in the narrow sense means only the lowest-energy (1s) state in electron–hole bound states, there exist many bound states in general. In the broad sense, scattering (unbound) states of an electron and a hole are also treated as excitons.

² Dimensionality of the electron–hole relative motion will be defined afterwards. In this paper, the ‘quasi-two-dimensional’ (‘quasi-one-dimensional’) case with a finite confinement width of the system is denoted as ‘ $d \simeq 2$ ’ (‘ $d \simeq 1$ ’). The purely two-dimensional (one-dimensional) case is written as $d = 2$ ($d = 1$).

³ The upward energy shift due to the subband formation is included in the definition of $E_{\text{gap}}^{(d)}$.

band and a single valence band [3].

The relation between the number of excitons⁴ and the optical responses is a long-standing problem in exciton physics. Under a weak excitation condition, only an electron and a hole are created, which form a bound state due to the Coulomb attraction. This is just a two-body problem, which can be solved in principle, for example, with the use of the effective-mass approximation. However, in the case of stronger excitation where many (more than one) electrons and holes are excited in semiconductors, many-body effects should be taken into account: for example, formation of an excitonic molecule as a four-fermion problem or the Bose–Einstein condensation of excitons as a $(N + N)$ -fermion problem. In any case, the interparticle Coulomb interaction plays an essential role in exciton problems.

There are several lengthscales characterizing the electron–hole system: the de Broglie wavelength of Bloch electrons λ_e and λ_h , the exciton effective Bohr radius a_B^* for single exciton systems in $d = 3$, the Fermi wavelength $2\pi/k_F$ for high-density degenerate electron–hole systems, the coherence/localization length of the centre-of-mass motion in unperfect crystals, and so on. The dimensionality of exciton systems depends on the ratio of the geometrical-confinement length L_\perp to these lengthscales in which we are interested⁵. We should always be concerned about which degree of freedom is confined geometrically.

Here we note the general characteristics of low-dimensional systems. (a) The low-dimensional systems are sensitive to lattice distortions and fluctuations, which lead to the Peierls instability, the self-trapped states, and the polaron formation. Soliton excitations are also possible in one dimension. (b) Quantum fluctuation plays an important role at low temperature, which suppresses any long-range orderings. Then the mean-field approximation becomes invalid. (c) Electronic correlation becomes of more importance due to effective reduction of the kinetic energy. (d) The electronic structures are easily affected by the randomness. An electron or an exciton suffers from the Anderson localization in the presence of infinitesimally weak randomness. All these features should be taken into account in general in the study of low-dimensional systems.

10.3 Single-exciton problem

To simplify the problem, we employ the effective-mass approximation and the envelope function approximation. Exchange and spin–orbit interactions are neglected here, so only the spin–singlet exciton is considered. We take the \mathbf{r}_\parallel axis along the unconfined directions and the \mathbf{r}_\perp axis along the confined directions. In $d \simeq 2$ systems, $\mathbf{r}_\parallel = (x, y)$ and $\mathbf{r}_\perp = z$, while for $d \simeq 1$ systems, $\mathbf{r}_\parallel = x$ and

⁴ We should note that the number of excitons is not always a well-defined quantity. Fluctuation of the exciton number is often important.

⁵ An energy scale describing the geometrical confinement corresponds to the subband separation, being proportional approximately to L_\perp^{-2} .

$\mathbf{r}_\perp = (y, z)$. A normalization length along the unconfined directions is L_\parallel . Since we are most interested in low dimensionality of the electron–hole relative motion of a Wannier exciton, $a < L_\perp < a_B^*$ is assumed⁶ with the lattice constant a .

The d -dimensional exciton envelope function for $\mathbf{K} \sim 0$ may be written as

$$L_\parallel^{-d/2} \Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}}) \phi_{\{\alpha\}}(\mathbf{r}_{e\perp}) \phi_{\{\beta\}}^*(\mathbf{r}_{h\perp}), \quad (10.1)$$

where $\mathbf{r}_{\text{eh}} \equiv \mathbf{r}_{e\parallel} - \mathbf{r}_{h\parallel}$ is the electron–hole relative coordinate, $\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}$ represents the electron–hole relative motion along the unconfined (\mathbf{r}_{eh}) directions, which is specified by a quantum number(s) v , and $\phi_{\{\alpha\}}$ ($\phi_{\{\beta\}}$) denotes a subband envelope function in the conduction (valence) band, which are determined by the shape of the confinement potential in quantum structures. When the confinement is strong, $\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}$ satisfies

$$\left[-\frac{\hbar^2}{2\mu} \Delta_d + V^{(d)}(\mathbf{r}_{\text{eh}}) \right] \Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}}) = E_v^{(d)} \Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}}), \quad (10.2)$$

where μ is the electron–hole reduced mass, Δ_d d -dimensional Laplacian, $E_v^{(d)}$ an exciton energy measured from $E_{\text{gap}}^{(d)}$, $V^{(d)}(\mathbf{r}_{\text{eh}})$ the effective d -dimensional Coulomb potential,

$$V^{(d)}(\mathbf{r}_{\text{eh}}) = \int V^{(3)}(\mathbf{r}_e, \mathbf{r}_h) |\phi_{\{\alpha\}}(\mathbf{r}_{e\perp})|^2 |\phi_{\{\beta\}}(\mathbf{r}_{h\perp})|^2 \prod d\mathbf{r}_{e\perp} d\mathbf{r}_{h\perp}, \quad (10.3)$$

with the Coulomb potential $V^{(3)}(\mathbf{r}_e, \mathbf{r}_h) = -(e^2/\epsilon) |\mathbf{r}_e - \mathbf{r}_h|^{-1}$ in $d = 3$. The exciton has discrete spectra for $E_v^{(d)} < 0$ and continuous spectra for $E_v^{(d)} > 0$. For $d = 2$ with $L_\perp \rightarrow 0$, single-exciton problems have been solved exactly [4], showing that the exciton binding energy $|E_0^{(2)}|$ is four times larger than that in $d = 3$ (correspondingly, the Bohr radius becomes half). Thus the two dimensionality enhances the excitonic effects. In the $d \simeq 2$ case ($L_\perp > 0$), variational approximation has been employed to discuss the excitonic effects [5].

In $d = 1$ ($L_\perp \rightarrow 0$), the Coulomb attractive interaction is given by $V^{(1)}(\mathbf{r}_{\text{eh}}) = -(e^2/\epsilon) |x_e - x_h|^{-1}$. Under this situation the exciton bound states have pathological features [6]: (i) the divergence of the binding energy of the lowest state ($|E_0^{(1)}| \rightarrow \infty$), (ii) the $\sqrt{\delta(\mathbf{r}_{\text{eh}})}$ -like wavefunction of the lowest state, and (iii) the violation of the non-degeneracy theorem for other bound states. These unphysical anomalies result from the divergence of the Coulomb potential at the origin ($\mathbf{r}_{\text{eh}} = 0$). Since in one dimension an electron and a hole cannot move without touching each other, the divergence at the origin becomes more crucial than at higher dimensions.

In order to avoid the divergence and to obtain analytical solutions, a regularized potential [7, 8] is often employed: $V^{(1)}(\mathbf{r}_{\text{eh}}; x_0) \equiv -(e^2/\epsilon) (|\mathbf{r}_{\text{eh}}| +$

⁶ In the $d = 2$ limit, the exciton Bohr radius becomes $\frac{1}{2}a_B^*$. Therefore, strictly speaking, $L_\perp < \frac{1}{2}a_B^*$ is required to realize the $d \simeq 1$ situation.

$x_0)^{-1}$ with a cutoff $x_0 \geq 0$. The cutoff x_0 is found to be proportional to the lateral size of the wire (e.g., $x_0 \simeq 0.18L_\perp$ for a square cross section). Then we have relative wavefunctions of the bound states as [9]

$$\Psi_{\nu,\{\alpha\}\{\beta\}}^{(1)}(\tilde{\mathbf{r}}_{\text{eh}}) = N_\nu \tilde{\mathbf{r}}_{\text{eh}} e^{-\tilde{r}_{\text{eh}}/2} \Gamma(1 + \nu) [F(1 - \nu, 2; \tilde{\mathbf{r}}_{\text{eh}}) - G(1 - \nu, 2; \tilde{\mathbf{r}}_{\text{eh}})], \quad (10.4)$$

where $\tilde{\mathbf{r}}_{\text{eh}} = 2(|\mathbf{r}_{\text{eh}}| + x_0)/(va_B^*)$, N_ν is the normalization constant, $\Gamma(z)$ is the gamma function, $F(\alpha, \gamma; \tilde{\mathbf{r}}_{\text{eh}})$ and $G(\alpha, \gamma; \tilde{\mathbf{r}}_{\text{eh}})$ are the confluent hypergeometric functions. Here ν specifies the energy eigenvalues ($E_\nu^{(1)} = -E_R^*/\nu^2$), is not necessarily an integer and is determined by additional constraints [9]:

$$\Psi_{\nu,\{\alpha\}\{\beta\}}^{(1)}(\tilde{\mathbf{r}}_{\text{eh}} = 2x_0/va_B^*) = 0 \quad (\text{even parity state}), \quad (10.5)$$

$$\frac{d\Psi_{\nu,\{\alpha\}\{\beta\}}^{(1)}(\tilde{\mathbf{r}}_{\text{eh}})}{d\tilde{\mathbf{r}}_{\text{eh}}} \Big|_{\tilde{\mathbf{r}}_{\text{eh}}=2x_0/va_B^*} = 0 \quad (\text{odd parity state}). \quad (10.6)$$

E_R^* is the effective Rydberg. As $x_0 \rightarrow 0$ (i.e. $L_\perp \rightarrow 0$), the energy of the lowest state $E_0^{(1)}$ becomes negatively infinite, indicating the divergence of the binding energy. This is a peculiar feature of the $d = 1$ exciton. When $x_0 = 0$, other bound states become doubly degenerate for odd and even parities at finite energies, $-E_R^*(n - 1)^{-2}$ with $n = 2, 3, 4, \dots$. These features suggest that the attractive force between an electron and a hole in $d \simeq 1$ systems is more effective than in $d \simeq 2$ systems [10].

Because the above results are based on the continuum model with the effective-mass and the envelope-function approximations, the ultraviolet cutoff or the finiteness of the lattice constant is not taken into account. Nevertheless, enhancement of the excitonic effects is obvious in lower dimensions.

10.3.1 One-photon absorption process

The transition probability between the ground state and an exction state is reflected in the one-photon absorption (OPA) spectrum. The OPA coefficient of the d -dimensional exciton is proportional to

$$W_{\text{OPA}}^{(d)}(\omega) \propto |\langle c | \hat{\mathbf{e}} \cdot \mathbf{p} | v \rangle|^2 \sum_{\{\alpha\}} \sum_{\{\beta\}} |\langle \phi_{\{\alpha\}} | \phi_{\{\beta\}} \rangle|^2 \sum_{\nu} |\Psi_{\nu,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}} = 0)|^2 \times \delta(\hbar\omega - E_{\text{gap}}^{(d)} - E_\nu^{(d)}), \quad (10.7)$$

where $\langle c | \hat{\mathbf{e}} \cdot \mathbf{p} | v \rangle$ is the interband-transition dipole matrix element with polarization unit vector $\hat{\mathbf{e}}$ and momentum operator \mathbf{p} , and $|c\rangle$ ($|v\rangle$) is the band wavefunction. Here we assume for $d \simeq 1$ that $\phi_{\{\alpha\}}$ and $\phi_{\{\beta\}}$ have decoupled forms such as $\phi_{\{\alpha\}} \simeq \prod_{i=y,z} \phi_{\alpha_i}(i)$. From this, the subband selection rules become $\alpha_z - \beta_z = \text{even} \simeq 0$ for $d \simeq 2$, and $\alpha_y - \beta_y = \text{even} \simeq 0$ and $\alpha_z - \beta_z = \text{even} \simeq 0$ for $d \simeq 1$, as summarized in table 10.1.

Table 10.1. The subband selection rules in the one-photon-absorption (OPA) process for arbitrary polarization direction $\hat{\varepsilon}$ of the incident light.

Exciton dimension d	Arbitrary $\hat{\varepsilon}$
$d \simeq 2$	$\alpha_z - \beta_z = \text{even} \simeq 0$
$d \simeq 1$	$\alpha_y - \beta_y = \text{even} \simeq 0$ and $\alpha_z - \beta_z = \text{even} \simeq 0$

The one-photon transition is allowed only for the exciton bound states with an even parity, $|\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}} = 0)| \neq 0$ (e.g., the 1s state). The bound states with an odd parity ($|\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}} = 0)| = 0$, e.g., the 2p state) are not reflected in OPA at all. The oscillator strength ($f_{v,\{\alpha\}\{\beta\}}^{(d)} \propto |\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}} = 0)|^2$) of the lowest-energy bound state (always even parity) becomes anomalously large. The oscillator strength of other bound states with an even parity vanishes completely in the limit of $x_0 \rightarrow 0$. Thus, in $d \simeq 1$ the OPA of *only* the lowest exciton state is extremely (divergingly) strong in comparison with the other bound states.

Above the band-gap energy ($\hbar\omega \geq E_{\text{gap}}^{(d)}$), a difference in the interband OPAs in $d \simeq 2$ and $d \simeq 1$ excitons is more clear⁷. For the $d \simeq 2$ systems, the interband absorption is enhanced due to the excitonic effects, while it is reduced for the $d \simeq 1$ exciton systems. To see this, the Sommerfeld factor, which is defined as the ratio of the interband OPA rate with excitonic effects to that without excitonic effects, has been evaluated for $d = 2$ and $d \simeq 1$. The Sommerfeld factor for $d = 2$ is

$$S^{(2)}(\hbar\omega) = \frac{e^{\pi\alpha_2}}{\cosh(\pi\alpha_2)} \geq 1, \quad (10.8)$$

where $\alpha_2 = \alpha_2(\hbar\omega) = [2\mu(a_B^*)^2(\hbar\omega - E_{\text{gap}}^{(2)})]^{-1/2}$. The OPA rate at the band edge ($\hbar\omega = E_{\text{gap}}^{(2)}$) is twice as large as that without the excitonic effect, i.e. $S^{(2)}(\hbar\omega = E_{\text{gap}}^{(2)}) = 2$. The Sommerfeld factor for $d = 2$ is also called the Coulomb ‘enhancement’ factors because of $S^{(2)}(\hbar\omega) \geq 1$. On the other hand, the Sommerfeld factor for $d \simeq 1$ is

$$S^{(1)}(\hbar\omega; x_0) = \frac{e^{\pi\alpha_1}}{8} \frac{|D_0^{(2)}W_0^{(1)} - D_0^{(1)}W_0^{(2)}|^2}{|D_0^{(1)}|^2 + |D_0^{(2)}|^2} \leq 1, \quad (10.9)$$

where $\alpha_1 = \alpha_1(\hbar\omega) \equiv [2\mu(a_B^*)^2(\hbar\omega - E_{\text{gap}}^{(1)})]^{-1/2}$. Here $W_0^{(j)} \equiv W^{(j)}(\tilde{\mathbf{r}}_{\text{eh}} = 2ikx_0)$, $D_0^{(j)} \equiv dW^{(j)}(\tilde{\mathbf{r}}_{\text{eh}})/d\tilde{\mathbf{r}}_{\text{eh}}|_{\tilde{\mathbf{r}}_{\text{eh}}=2ikx_0}$, and

$$W^{(j)}(\mathbf{r}_{\text{eh}}) \equiv \tilde{\mathbf{r}}_{\text{eh}} e^{-\tilde{\mathbf{r}}_{\text{eh}}/2} \Gamma(1 \pm i\alpha)[F(1 + i\alpha, 2; \tilde{\mathbf{r}}_{\text{eh}}) \pm G(1 + i\alpha, 2; \tilde{\mathbf{r}}_{\text{eh}})], \quad (10.10)$$

⁷ The OPA strength is proportional to the joint density of states (JDOS) when no electron-hole Coulomb interaction is taken into account.

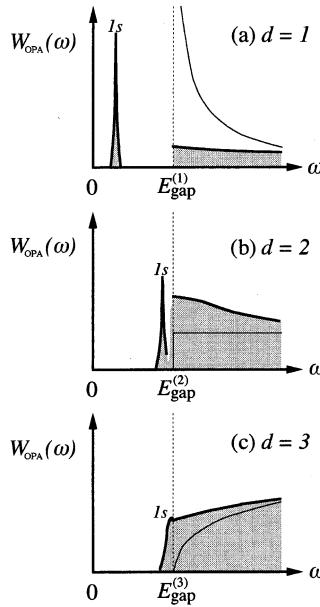


Figure 10.1. Schematic drawings of the one-photon-absorption (OPA) spectra $W_{\text{OPA}}^{(d)}$ for (a) $d \simeq 1$, (b) $d \simeq 2$ and (c) $d = 3$ semiconductors as a function of the frequency ω of an incident light field. Only the 1s exciton bound state is drawn below the band gap $E_{\text{gap}}^{(d)}$. Thin curves mean the OPA spectra without the excitonic effects, which are proportional to the joint density of states (JDOS). (After [10]).

for $j = 1$ (2) corresponding to the + (−) sign of the right-hand side. Here the scattering state of the $d \simeq 1$ exciton of energy $E_k^{(1)} = \hbar\omega > 0$ is characterized by the wavenumber k through $E_k^{(1)} = \hbar^2 k^2 / 2\mu$. The Sommerfeld factor in $d \simeq 1$ is less than unity, which means that the electron–hole Coulomb attraction suppresses the OPA of an allowed interband transition. At $\hbar\omega = E_{\text{gap}}^{(1)}$, i.e. $\alpha_1 \rightarrow +\infty$, the Sommerfeld factor vanishes, $S^{(1)}(\hbar\omega = E_{\text{gap}}^{(1)}; x_0) = 0$, which cancels the divergence due to the joint density of states (JDOS) of $d = 1$ systems. More surprisingly, in the limit of $x_0 \rightarrow 0$, no interband absorption takes place, i.e. $S^{(1)}(\hbar\omega; x_0 = 0) = 0$ for $\forall \hbar\omega \geq E_{\text{gap}}^{(1)}$. Figure 10.1 shows the schematic shapes of the OPA spectra for $d \simeq 1, 2, 3$, indicating the prominent 1s exciton peak in $d \simeq 1$ [12]⁸.

⁸ Low-dimensional excitonic effects have been discussed in relation to luminescence processes of silicon nanocrystals (see, e.g., [11]).

10.3.2 Two-photon absorption process

In the OPA process, the polarization dependence appears only in the interband matrix element, $|\langle c|\hat{e} \cdot \mathbf{p}|v\rangle|$. Therefore, OPA anisotropy for the polarization directions reflects only that of the Bloch (band) electrons not of the excitons nor of the subband structures. On the other hand, the two-photon absorption (TPA) process reflects directly the dimensionality of excitons, i.e. squeezing of the wavefunction of the exciton's relative motion itself in a low-dimensional geometry. We shall now discuss the polarization anisotropy of the TPA.

Consider the situation where the photon energies $\hbar\omega_1$ and $\hbar\omega_2$ of incident light beams are both close to $\frac{1}{2}E_{\text{gap}}^{(d)}$ and these two beams have a common polarization vector \hat{e} . The complete summation over all the intermediate states is performed by approximating off-resonant energy denominators as a constant [13]. Contrary to the OPA case, the TPA probability $W_{\text{TPA}}^{(d)}$ for d -dimensional exciton systems has different forms depending on the polarization directions [14, 15].

When $\hat{e} \parallel \mathbf{r}_{\parallel}$, the TPA rate is proportional to $W_{\text{TPA}}^{(d)}(\omega_1 + \omega_2; \hat{e} \parallel \hat{\mathbf{r}}_{\parallel}) \propto |\langle c|\hat{e} \cdot \mathbf{p}|v\rangle|^2 \mu_{\parallel}^{-2} G^{(d)}(\hat{e} \parallel \mathbf{r}_{\parallel})$, where

$$G^{(d)}(\hat{e} \parallel \mathbf{r}_{\parallel}) = (L_{\text{norm}})^{d-3} \sum_{\{\alpha\}} \sum_{\{\beta\}} |\langle \phi_{\{\alpha\}} | \phi_{\{\beta\}} \rangle|^2 \sum_v \left| \frac{\partial}{\partial \mathbf{r}_{\text{eh}}} \Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}}) \right|_{\mathbf{r}_{\text{eh}}=0}^2 \times \delta(\hbar\omega_1 + \hbar\omega_2 - E_v^{(d)} - E_{\text{gap}}^{(d)}), \quad (10.11)$$

with L_{norm} the normalization length along \mathbf{r}_{\perp} direction and μ_{\parallel} the electron-hole reduced mass for the motion along \mathbf{r}_{\parallel} direction. Here we find the subband selection rule: $\alpha_z - \beta_z = \text{even} \simeq 0$ for $d \simeq 2$ and $\alpha_i - \beta_i = \text{even} \simeq 0$ ($i = y$ or z) for $d \simeq 1$. These are identical with those in the OPA case. Only the odd-parity states are allowed for the TPA process of $\hat{e} \parallel \mathbf{r}_{\parallel}$ because the derivative of the exciton wavefunction at the origin is non-zero only for odd-parity states, which are forbidden for the OPA process. Figure 10.2(a) shows the envelope-function part $G^{(2)}$ for $\hat{e} \parallel \mathbf{r}_{\parallel}$. The case of $d \simeq 1$ was given in [15].

On the other hand, when $\hat{e} \parallel \mathbf{r}_{\perp}$, the TPA coefficient is proportional to $W_{\text{TPA}}^{(d)}(\omega_1 + \omega_2; \hat{e} \parallel \mathbf{r}_{\perp}) \propto |\langle c|\hat{e} \cdot \mathbf{p}|v\rangle|^2 \mu_{\zeta}^{-2} G^{(d)}(\hat{e} \parallel \mathbf{r}_{\perp})$, with

$$G^{(d)}(\hat{e} \parallel \mathbf{r}_{\perp}) = (L_{\text{norm}})^{d-3} \sum_{\alpha_{\zeta}} \sum_{\beta_{\zeta}} \left| \left\langle \phi_{\alpha_{\zeta}} \left| \frac{\partial}{\partial x_{\zeta}} \right| \phi_{\beta_{\zeta}} \right\rangle \right|^2 \times \sum_{\alpha_{\zeta'}} \sum_{\beta_{\zeta'}} |\langle \phi_{\alpha_{\zeta'}} | \phi_{\beta_{\zeta'}} \rangle|^2 \sum_v |\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}}=0)|^2 \times \delta(\hbar\omega_1 + \hbar\omega_2 - E_{\text{gap}}^{(d)} - E_v^{(d)}), \quad (10.12)$$

where μ_{ζ} is the electron-hole reduced mass for the motion along the confinement direction. For $d \simeq 1$, ζ is an index specifying an component of $\mathbf{r}_{\perp} = (y, z)$ parallel to the polarization direction \hat{e} , while ζ' indicates the rest component in

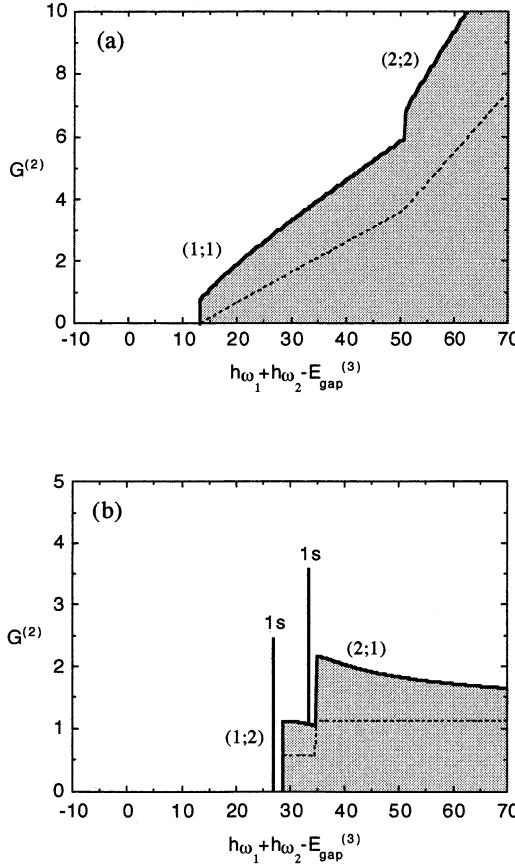


Figure 10.2. The envelope function part $G^{(2)}$ of the two-photon-absorption (TPA) spectrum of a $d \simeq 2$ semiconductor for (a) $\hat{\epsilon} \parallel \hat{r}_\parallel$ and (b) $\hat{\epsilon} \parallel r_\perp = z$. Subband indices mean (α_z, β_z) . The quasi-continuum region just below the TPA edge is not drawn here. In (a), odd-parity (e.g., 2p) exciton states are allowed for the TPA transition, but are invisible in this scale. Dotted curves mean $G^{(2)}$ without the excitonic effects. The actual TPA spectra include also $|\langle c|\hat{\epsilon} \cdot \mathbf{p}|v\rangle|^2$. (After [15]).

\mathbf{r}_\perp . The subband selection rule for $d \simeq 2$ is $\alpha_z - \beta_z = \text{odd} \simeq \pm 1$, while for $d \simeq 1$, $\alpha_\zeta - \beta_\zeta = \text{odd} \simeq \pm 1$ for the polarization direction and $\alpha_{\zeta'} - \beta_{\zeta'} = \text{even} \simeq 0$ for the other confinement direction. Only the even-parity states are allowed for the TPA process of $\hat{\epsilon} \parallel \mathbf{r}_\perp$ because $\Psi_{v,\{\alpha\}\{\beta\}}^{(d)}(\mathbf{r}_{\text{eh}} = 0)$ is non-zero only for even-parity excitons, which are allowed for the OPA process. Figure 10.2(b) shows the envelope-function part $G^{(2)}$ for $\hat{\epsilon} \parallel \mathbf{r}_\perp = z$. Thus we find that not only even-parity states but also odd-parity ones can be detected by TPA with appropriate control of the polarization direction. These anisotropic TPA characteristics result

Table 10.2. The subband selection rules in the two-photon-absorption (TPA) process for a polarization direction $\hat{\varepsilon}$ of the incident lights. For $d \simeq 1$, ζ is an index specifying an component of $r_{\perp} = (y, z)$ parallel to the polarization direction $\hat{\varepsilon}$, while ζ' indicates the rest component in r_{\perp} .

Exciton dimension d	$\hat{\varepsilon} \parallel r_{\parallel}$	$\hat{\varepsilon} \parallel r_{\perp}$
$d \simeq 2$	$\alpha_z - \beta_z = \text{even} \simeq 0$	$\alpha_z - \beta_z = \text{odd} \simeq \pm 1$
$d \simeq 1$	$\alpha_y - \beta_y = \text{even} \simeq 0$ and $\alpha_z - \beta_z = \text{even} \simeq 0$	$\alpha_{\zeta} - \beta_{\zeta} = \text{odd} \simeq \pm 1$ and $\alpha_{\zeta'} - \beta_{\zeta'} = \text{even} \simeq 0$

directly from the low-dimensional nature of the electron–hole relative motion. The subband selection rule is also anisotropic depending on the polarization direction, as summarized in table 10.2.

In the regions of continuous spectra ($\hbar\omega_1 + \hbar\omega_2 \geq E_{\text{gap}}^{(d)}$), the $d \simeq 1$ exciton effects enhance the TPA for $\hat{\varepsilon} \parallel \hat{x}$ (the unconfinement direction), whereas they reduce the TPA for $\hat{\varepsilon} \parallel \hat{x}_{\zeta}$ (a confinement direction), which is in striking contrast to the higher-dimensional cases, where the TPA is always enhanced by excitonic effects. As a result, in the case of the TPA in one dimension, one can observe both the reduction (for $\hat{\varepsilon} \parallel \hat{x}$) and enhancement (for $\hat{\varepsilon} \parallel \hat{x}_{\zeta}$) in the same sample simply by rotating the polarization direction. In an intermediate regime between $d \simeq 2$ and $d \simeq 1$, the dimensional crossover effects appear only in the TPA process (not in the OPA) [16]. These theories can explain experimental results [17, 18].

10.3.3 Dielectric confinement effect

In determining the effective Coulomb potential in less than two dimensions, we need to include the dielectric image-charge effect arising from the difference in the dielectric constant between the material (ϵ) and the surroundings (ϵ_1). This is a confinement effect of the Coulomb interaction. If the surrounding materials have a smaller dielectric constant and a larger energy gap than the relevant material, the electron–hole Coulomb attraction in the relevant material works very effectively through the surroundings with reduced screening. As a result, the exciton binding energy is expected to increase much more than in a normal low-dimensional system without this effect.

We shall consider a $d \simeq 2$ system with the well width L_{\perp} ($|z| \leq \frac{1}{2}L_{\perp}$) (its dielectric constant is ϵ) [19, 20]. The barrier regions have the dielectric constants, ϵ_1 ($\epsilon > \epsilon_1$). The electrostatic interaction between two point charges in the $d \simeq 2$ system is given by

$$V_{\text{image}}^{(2)}(\mathbf{r}_{\text{eh}}) = -\frac{e^2}{\epsilon} \int d^2k \frac{e^{i\mathbf{k} \cdot \mathbf{r}_{\text{eh}}}}{k \sinh(kL_{\perp} + 2\eta)}$$

$$\times \cosh \left[k \left(\frac{L_\perp}{2} - z_e \right) + \eta \right] \cosh \left[k \left(\frac{L_\perp}{2} + z_h \right) + \eta \right]. \quad (10.13)$$

Here $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2}$, $\eta = \frac{1}{2} \ln[(\epsilon + \epsilon_1)/(\epsilon - \epsilon_1)]$, $\mathbf{r}_{eh} = (x_{eh}, y_{eh}) = (x_e - x_h, y_e - y_h)$. (i) When the electron is close to the hole ($r^2 \equiv |\mathbf{r}_{eh}|^2 + z_{eh}^2 \ll L_\perp^2$), the interaction coincides with that in bulk. (ii) When the electron is far from the hole ($|\mathbf{r}_{eh}| \gg (\epsilon/\epsilon_1)L_\perp$), it is determined by ϵ_1 only, i.e. $V_{\text{image}}^{(2)} \rightarrow -(e^2/\epsilon_1)|\mathbf{r}_{eh}|^{-1}$. (iii) When $\epsilon \gg \epsilon_1$, $V_{\text{image}}^{(2)} \rightarrow -2e^2/(\epsilon L_\perp) \{\ln[(\epsilon/\epsilon_1)(L_\perp/|\mathbf{r}_{eh}|)] - \gamma\}$ in the region of $L_\perp \ll |\mathbf{r}_{eh}| \ll (\epsilon/\epsilon_1)L_\perp$. Here $\gamma = 0.577$.

In a square wire ($d \simeq 1$) structure ($|y| \leq \frac{1}{2}L_\perp$, $|z| \leq \frac{1}{2}L_\perp$), the electron-hole interaction including the dielectric image-charge effect is derived as [9]

$$U_{\text{image}}^{(1)}(\mathbf{r}_e, \mathbf{r}_h) = -\frac{e^2}{\epsilon} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} q^{|m|+|n|} [x_{eh}^2 + y_m^2 + z_n^2]^{-1/2}, \quad (10.14)$$

where $q \equiv (\epsilon - \epsilon_1)/(\epsilon + \epsilon_1)$, $y_m \equiv y_e - mL_\perp - (-1)^m y_h$, and $z_n \equiv z_e - nL_\perp - (-1)^n z_h$. After averaging over y_e , z_e , y_h and z_h on the cross section of the wire, we have

$$V_{\text{image}}^{(1)}(x_{eh}) = -4E_R^* \frac{a_B^*}{L_\perp} \int_0^\infty dk_y \int_0^\infty dk_z F(k_y) F(k_z) \times \frac{\exp \left[-\pi x_{eh} \sqrt{k_y^2 + k_z^2}/L_\perp \right]}{\sqrt{k_y^2 + k_z^2}}, \quad (10.15)$$

where

$$F(k) = \frac{1 + 2q(1 - q^2) \cos(\pi k) - q^4}{1 - 2q^2 \cos(2\pi k) + q^4} f(k). \quad (10.16)$$

Here $f(k)$ is given using $j_0(x) \equiv \sin x/x$ as

$$f(k) = j_0^2 \left(\frac{\pi}{2} k \right) + j_0 \left(\frac{\pi}{2} (k-2) \right) j_0 \left(\frac{\pi}{2} k \right) + j_0 \left(\frac{\pi}{2} (k+2) \right) j_0 \left(\frac{\pi}{2} k \right) + \frac{1}{4} j_0^2 \left(\frac{\pi}{2} (k-2) \right) + \frac{1}{4} j_0^2 \left(\frac{\pi}{2} (k+2) \right) + \frac{1}{2} j_0 \left(\frac{\pi}{2} (k-2) \right) j_0 \left(\frac{\pi}{2} (k+2) \right). \quad (10.17)$$

The calculated potential with the dielectric image-charge effects $V_{\text{image}}^{(1)}(x_{eh})$ can be well fitted by $V^{(1)}(x_{eh}; x_0)$. As the ratio ϵ/ϵ_1 increases, the attractive potential becomes more long-ranged and is enhanced for all regions of x_{eh} over the bare Coulomb potential $V^{(1)}(x_{eh}) = -(e^2/\epsilon)|x_{eh}|^{-1}$, as shown in figure 10.3. Therefore, the peculiar features of the $d \simeq 1$ excitons would be more pronounced by the dielectric image-charge effect.

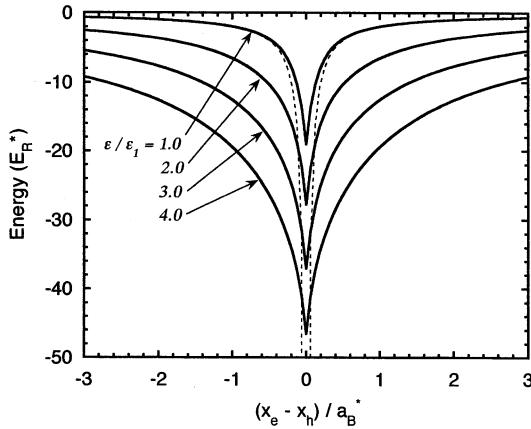


Figure 10.3. Effective one-dimensional Coulomb potential of $d \simeq 1$ quantum wire with a square cross section. The confinement length is assumed to be $L_{\perp} = \frac{1}{2}a_B^*$. Dielectric constants of the well region (barrier region) is ϵ (ϵ_1). Their ratio are $\epsilon/\epsilon_1 = 1.0, 2.0, 3.0$, and 4.0 . The dotted curves are the bare Coulomb potential (corresponding to $L_{\perp} \rightarrow 0$) for $\epsilon/\epsilon_1 = 1.0$. (After [9]).

10.4 A few exciton problems

The two-exciton excited state in semiconductors is related to the excitonic molecule, which has been investigated for a long time [21]. Recently, it was found that the $\chi^{(3)}$ process is sensitive to details of the two-exciton state. Therefore, a few exciton problems are important also in the study of nonlinear optical responses. In this section, we shall overview theories of an excitonic molecule in one dimension, and the recent status of the study of excitonic nonlinearities in semiconductors is reviewed.

10.4.1 An excitonic molecule in one dimension

When two electrons in the valence band are excited up to the conduction band, the system has four excited particles, two electrons and two holes, which can form an excitonic molecule (EM) in an analogy of the formation of a hydrogen molecule. In $d = 3$, the binding energy of the excitonic molecule was shown to be always positive [22]. Here we confine ourselves to the excitonic molecule in a $d \simeq 1$ semiconductor.

First, we introduce a theory of an excitonic molecule in $d \simeq 1$ with the Heitler-London approximation [8]. Here two Wannier excitons are assumed to exist in a continuum model of a $d \simeq 1$ semiconductor. The one-dimensional

effective Hamiltonian is

$$\hat{H}_{\text{EM}}^{(1)} = -\frac{\partial^2}{\partial \xi_1^2} - \frac{\partial^2}{\partial \xi_2^2} + \frac{2\mu}{m_h^*} \left(\frac{\partial}{\partial \xi_1} - \frac{\partial}{\partial \xi_2} \right) \frac{\partial}{\partial x_{\text{hh}}} - \frac{2\mu}{m_h^*} \frac{\partial^2}{\partial x_{\text{hh}}^2} - \frac{\mu}{2(m_e^* + m_h^*)} \frac{\partial^2}{\partial X^2} + 2U^{(1)}(\xi_1, \xi_2, x_{\text{hh}}), \quad (10.18)$$

where m_e^* (m_h^*) is the effective mass of an electron (a hole), the coordinates are defined as $\xi_1 \equiv x_{e1} - x_{h1}$, $\xi_2 \equiv x_{e2} - x_{h2}$, $x_{\text{hh}} \equiv x_{h1} - x_{h2}$, and $X \equiv [2(m_e^* + m_h^*)]^{-1}[m_e^*(x_{e1} + x_{e2}) + m_h^*(x_{h1} + x_{h2})]$, and the potential $U^{(1)}$ is given as

$$U^{(1)}(\xi_1, \xi_2, x_{\text{hh}}) = V^{(1)}(\xi_1 - \xi_2 + x_{\text{hh}}; x_0) + V^{(1)}(x_{\text{hh}}; x_0) - V^{(1)}(\xi_1; x_0) - V^{(1)}(\xi_2; x_0) - V^{(1)}(\xi_1 + x_{\text{hh}}; x_0) - V^{(1)}(\xi_2 - x_{\text{hh}}; x_0). \quad (10.19)$$

Here $V^{(1)}(x; x_0)$ is the regularized Coulomb potential in $d \simeq 1$. Since the polarization effect (deformation of the electron–hole relative wavefunction of a exciton due to the other exciton) is neglected, problem of the excitonic molecule is reduced to that of the hole–hole relative motion. The binding energy of the excitonic molecule in $d \simeq 1$, $E_{\text{EM}}^{(1)}$, is given by solving the eigenvalue problem for the hole–hole relative wavefunction $\psi_{\text{hh}}^{(1)}$, i.e.

$$\left[-\frac{2\mu}{m_h^*} \frac{\partial^2}{\partial x_{\text{hh}}^2} + V_{\text{hole}}^{(1)}(x_{\text{hh}}) \right] \psi_{\text{hh}}^{(1)} = -E_{\text{EM}}^{(1)} \psi_{\text{hh}}^{(1)}. \quad (10.20)$$

Here $V_{\text{hole}}^{(1)}(x_{\text{hh}})$ is the effective hole potential [8]:

$$V_{\text{hole}}^{(1)}(x_{\text{hh}}) = -2E_0^{(1)} + \int_{-\infty}^{\infty} d\xi_1 \int_{-\infty}^{\infty} d\xi_2 \Phi_{\text{corr}}^{(1)}(\xi_1, \xi_2, x_{\text{hh}}) \hat{H}_{\text{EM}}^{(1)} \Phi_{\text{corr}}^{(1)}(\xi_1, \xi_2, x_{\text{hh}}), \quad (10.21)$$

where $E_0^{(1)}$ is the ground-state exciton energy in $d \simeq 1$, and an eigenfunction of $\hat{H}_{\text{EM}}^{(1)}$ is assumed to be $\psi^{(1)}(\xi_1, \xi_2, x_{\text{hh}}) = \phi_{\text{hh}}^{(1)}(|x_{\text{hh}}|) \Phi_{\text{corr}}^{(1)}(\xi_1, \xi_2, x_{\text{hh}})$. Here $\Phi_{\text{corr}}^{(1)}$ describes the exciton correlations,

$$\Phi_{\text{corr}}^{(1)}(\xi_1, \xi_2, x_{\text{hh}}) = N(x_{\text{hh}}) [\Psi_v^{(1)}(\xi_1) \Psi_v^{(1)}(\xi_2) + \Psi_v^{(1)}(\xi_1 + x_{\text{hh}}) \Psi_v^{(1)}(\xi_2 - x_{\text{hh}})]$$

with the normalization function $N(x_{\text{hh}})$, where $\Psi_v^{(1)}$ is the relative wavefunction of an exciton in $d \simeq 1$ given in equation (10.4). According to numerical calculation, $E_{\text{EM}}^{(1)}$ is more than five times larger than $E_{\text{EM}}^{(3)} \simeq 0.04E_R^*$ for bulk GaAs. Thus, an excitonic molecule becomes more stable in lower-dimensional systems, similar to the exciton case.

Conventionally, an excitonic molecule is regarded as a bound state of two (Wannier) excitons. More rigorously, however, we should solve a four-fermion

problem taking into full account the correlation effect on the lengthscale of a lattice constant. In other words, a usual bosonic picture of the excitonic molecule cannot capture the binding mechanism quantitatively. We need to look into the way in which two electrons and two holes are reshuffled. In addition, in lower dimensions, the electronic correlation effects lead to the crossover between the Frenkel-type EM and the Wannier-type EM. In order to treat consistently this crossover, both a lattice model and the long-range interaction are required for the description of an EM. We solved the four-fermion model of an EM by exact diagonalization on a finite $d = 1$ lattice [23]. The EM Hamiltonian employed there is

$$\begin{aligned} \hat{\mathcal{H}}_{\text{EM}}^{(1)} = & -t_e \sum_{i\sigma} (\hat{a}_{i+1,\sigma}^\dagger \hat{a}_{i\sigma} + \text{h.c.}) - t_h \sum_{i\sigma} (\hat{d}_{i+1,\sigma}^\dagger \hat{d}_{i\sigma} + \text{h.c.}) + E_0 \sum_{i\sigma} \hat{d}_{i\sigma}^\dagger \hat{d}_{i\sigma} \\ & - \sum_{ij\sigma\sigma'} U_{ij} \hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma} \hat{d}_{j\sigma'}^\dagger \hat{d}_{j\sigma'} + \sum_{ij\sigma\sigma'} V_{ij} (\hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma} \hat{a}_{j\sigma'}^\dagger \hat{a}_{j\sigma'} + \hat{d}_{i\sigma}^\dagger \hat{d}_{i\sigma} \hat{d}_{j\sigma'}^\dagger \hat{d}_{j\sigma'}), \\ & - V_{\text{dipole}} \sum_{i\sigma} (\hat{a}_{i+1,\sigma}^\dagger \hat{d}_{i+1,-\sigma}^\dagger \hat{d}_{i,-\sigma} \hat{a}_{i\sigma} + \text{h.c.}), \end{aligned} \quad (10.22)$$

where $\hat{a}_{i\sigma}^\dagger$ ($\hat{d}_{i\sigma}^\dagger$) creates an electron in the conduction band (a hole in the valence band) with spin σ at site i , t_e (t_h) is the transfer energy of electrons (holes), and E_0 is the energy difference between the two orbitals from which conduction and valence bands arise. The attractive electron-hole interaction between the sites i and j is denoted by U_{ij} , while V_{ij} represents the repulsive electron-electron and hole-hole interactions. The last term describes the transfer of an electron-hole pair from site to site. We assumed that $U_{ij} = U_0$ for $i = j$, $U_{ij} = U_1/|i - j|$ for $i \neq j$ and $V_{ij} = V_0$ for $i = j$, $V_{ij} = V_1/|i - j|$ for $i \neq j$, where physically $U_0 > U_1 \geq 0$, $V_0 > V_1 \geq 0$, and $V_1 = U_1$.

The numerical diagonalization of the Hamiltonian has been carried out for a finite $d = 1$ system with the periodic boundary condition. We found that the Heitler-London scheme gives only about half the exact result for the whole range of the transfer energies. This discrepancy indicates that the correlation effect is of significance in $d = 1$ systems. The Heitler-London approximation is expected to give reasonable results only when the electron-hole relative motion is spatially extended. We stress, however, that in $d = 1$ systems the correlated motion of electrons and holes over the lengthscale of a lattice constant is essential even in the weak-coupling case. Moreover, continuum models for an EM holds only in the weak-coupling regime, where the hole-hole correlation is peaked at a finite distance larger than a lattice constant, while it decays faster for the strong-coupling regime where an EM is more compact in space and its size is of an atomic order. In $d = 1$, the coupling among particles becomes essentially stronger in comparison with that in higher dimensions. Therefore, the lattice model is suitable for description of the elementary excitations in one dimension such as a $d = 1$ EM in the whole range of the electron-hole coupling constant. Other physical quantities of the $d = 1$ EM are given in [23].

In some $d \simeq 1$ materials, a novel quasi-particle called an ‘excitonic n -string’ [24] can be formed by n electrons and n holes with small integer n . But in general, polyexcitons of $n > 2$ become unstable in higher dimensions. ‘Charged excitons’ composed of two electrons and one hole (or one electron and two holes) are also predicted and observed [25]. Condensation of charge-transfer excitons is also attracting attention in terms of the photoinduced phase transitions [26, 27].

10.4.2 Optical nonlinearity due to excitons

Nonlinear optical responses have been studied for a long time in a relation also to the photonic device application and the laser operation. In particular, the $\chi^{(3)}$ process is of special importance because it is the lowest-order nonlinear effect in ordinary materials. When the energy of the incident light is tuned at the 1s exciton level, the nonlinearity results from some characteristics of the 1s excitons. Recently, relations between the $\chi^{(3)}$ nonlinearity and the interaction between the 1s excitons are investigated theoretically [28]. Aims of theoretical studies on the optical nonlinearity in semiconductors is to clarify spatiotemporal dynamics of excitons (or electrons and holes) in photoexcited semiconductors and to predict temporal or spectral variations of induced nonlinear polarizations. For these purposes, there are two types of approach: the fermion-picture treatment where electrons and holes play their individual roles and the boson-picture treatment in which a pair of an electron and a hole is regarded as a composite boson.

10.4.2.1 Fermion-picture treatment

In the fermionic treatment, we start with two types of fermions considering separately electrons excited in a conduction band and holes in a valence band. There populations and the induced polarization (interband dipole moment) are traced with the Heisenberg equations of motion. A typical example of this treatment is the semiconductor Bloch equations (SEBs) [29, 30] in the momentum space. When we confine ourselves to the interband transitions neglecting intraband transitions, semiconductors are described as an ensemble of two-level atoms with inhomogeneously broadening. In analogy with the optical Bloch equations describing coherent optical transient phenomena, the equations of motion for three macroscopically-averaged quantities are

$$\begin{aligned} \frac{dn_{\mathbf{k}}^e(t)}{dt} &= -2 \operatorname{Im}[\Omega_{\mathbf{k}} p_{\mathbf{k}}^*] - \Gamma_{\mathbf{k}}^e n_{\mathbf{k}}^e, \\ \frac{dn_{\mathbf{k}}^h(t)}{dt} &= -2 \operatorname{Im}[\Omega_{\mathbf{k}} p_{\mathbf{k}}^*] - \Gamma_{\mathbf{k}}^h n_{\mathbf{k}}^h, \\ \frac{dp_{\mathbf{k}}(t)}{dt} &= -i[\tilde{\epsilon}_{\mathbf{k}}^e + \tilde{\epsilon}_{\mathbf{k}}^h] p_{\mathbf{k}} - i(n_{\mathbf{k}}^e + n_{\mathbf{k}}^h - 1)\Omega_{\mathbf{k}} - \gamma_{\mathbf{k}} p_{\mathbf{k}}, \end{aligned} \quad (10.23)$$

where $n_{\mathbf{k}}^e(t) \equiv \langle \sum_{\sigma} \hat{a}_{\mathbf{k}\sigma}^\dagger \hat{a}_{\mathbf{k}\sigma} \rangle$ [$n_{\mathbf{k}}^h(t) \equiv \langle \sum_{\sigma} \hat{d}_{\mathbf{k}\sigma}^\dagger \hat{d}_{\mathbf{k}\sigma} \rangle$] is the Heisenberg representation of the mean density of the conduction electrons [valence holes]

with momentum \mathbf{k} at time t and the interband polarization $p_{\mathbf{k}}(t) \equiv \langle \sum_{\sigma} \hat{a}_{\mathbf{k}\sigma} \hat{d}_{\mathbf{k}\sigma} \rangle$. Here $\hat{a}_{\mathbf{k}\sigma}^{\dagger}$ ($\hat{d}_{\mathbf{k}\sigma}^{\dagger}$) creates an electron in the conduction band (a hole in the valence band) with momentum \mathbf{k} and spin σ , $\tilde{\epsilon}_{\mathbf{k}}^v \equiv \epsilon_{\mathbf{k}}^v - \sum_{\mathbf{q}} V_{|\mathbf{k}-\mathbf{q}|} n_{\mathbf{q}}^v$ is the single-particle energy renormalized by the Coulomb interaction $V_{\mathbf{q}}$ between an electron ($v = e$) and a hole ($v = h$), $\Omega_{\mathbf{k}} \equiv d_{cv} E(t) + \sum_{\mathbf{q} \neq \mathbf{k}} V_{|\mathbf{k}-\mathbf{q}|} p_{\mathbf{q}}$ is the generalized Rabi frequency for the incident light field $E(t)$, d_{cv} is the interband dipole matrix element, and $\Gamma_{\mathbf{k}}^v, \gamma_{\mathbf{k}}$ are the phenomenological relaxation constants.

In order to obtain a closed form of coupled equations, the Hartree–Fock (mean-field) approximation was employed, where the four (or higher) operator terms in the Heisenberg equations of motion are split into products of two-operator terms, the densities and the interband polarization. This is valid only for the case of high-density and homogeneous excitations. However, since the single exciton state of the zero centre-of-mass momentum $\mathbf{K} = 0$ can be described by the SBEs, it has been frequently used to analyse experimental results in bulk semiconductors even for the weak-excitation case. Nevertheless, we have to note that the contribution from an excitonic molecule (two-exciton states) is completely neglected in the SBEs. This is a fatal shortcoming of this formulation for the description of the $\chi^{(3)}$ process close to the 1s exciton resonance [31, 32]⁹. To overcome this, other fermion-picture theories, for example, the real-space density matrix theory taking into account higher-order correlations [33] or the Green function formalism are proposed and progressed.

10.4.2.2 Boson-picture treatment

When we consider the excitonic $\chi^{(3)}$ nonlinearity, we need to consider only the subspace containing up to two-electron excited states in the intermediate-state summation for perturbation calculations. In such a rather weak excitation regime, the screening effect is still weak, then an electron and a hole tend to make a pair (an exciton bound state) with each other. In the boson-picture treatment, a *pair* of an electron and a hole is considered as a basis, and it is treated as a (quasi-)boson.

Kuwata-Gonokami *et al* carried out a degenerate four-wave mixing experiment near the 1s exciton resonance with a GaAs quantum well in a high- Q cavity under circularly-polarized light irradiation. To explain their experimental results, they proposed an elastic-scattering model of two cavity polaritons [34]. The cavity polariton is a hybridized state of a 1s exciton and a photon in the cavity, and is regarded as a (quasi-)boson. Such a phenomenological model is called the weakly interacting boson (WIB) model. In the WIB model, we first assume there is a pure-boson describing a quasi-particle in an excited state of the material, whose operators are written as $\hat{B}_{\mathbf{k}\sigma}$ and $\hat{B}_{\mathbf{k}\sigma}^{\dagger}$ with an internal degree of freedom (helicity or total angular momentum) $\sigma = \pm 1$. They are assumed to satisfy the pure-boson commutation relation: $[\hat{B}_{\mathbf{k}\sigma}, \hat{B}_{\mathbf{k}'\sigma'}^{\dagger}] = \delta_{\mathbf{k}\mathbf{k}'} \delta_{\sigma\sigma'}$.

⁹ The excitation-induced dephasing was considered in the semiconductor Bloch equations to treat the $\chi^{(3)}$ process (see Lindberg *et al* [32]).

Usually, this pure-boson describes a 1s exciton in an approximate sense. Only three types of interactions (whose momentum dependence is neglected) between bosons and photons are required: (a) an attractive interaction W between bosons with different total angular momentum, (b) a repulsive interaction R between bosons with an identical total angular momentum, and (c) the phase-space filling effect G resulting from the Pauli's exclusion principle. The WIB Hamiltonian is $\hat{\mathcal{H}}^{\text{WIB}} = \hat{\mathcal{H}}_{\text{linear}}^{\text{WIB}} + \hat{\mathcal{H}}_{\text{nonlinear}}^{\text{WIB}}$ with

$$\hat{\mathcal{H}}_{\text{linear}}^{\text{WIB}} = \sum_{\mathbf{k}\sigma} \omega_{\mathbf{k}} \hat{c}_{\mathbf{k}\sigma}^\dagger \hat{c}_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma} \Omega_{\mathbf{k}} \hat{B}_{\mathbf{k}\sigma}^\dagger \hat{B}_{\mathbf{k}\sigma} + g \sum_{\mathbf{k}\sigma} (\hat{B}_{\mathbf{k}\sigma}^\dagger \hat{c}_{\mathbf{k}\sigma} + \text{h.c.}), \quad (10.24)$$

$$\begin{aligned} \hat{\mathcal{H}}_{\text{nonlinear}}^{\text{WIB}} = & W \sum_{\mathbf{k}\mathbf{k}' \neq 0} \sum_{\mathbf{q}} \hat{B}_{\mathbf{k}+\mathbf{q},+1}^\dagger \hat{B}_{\mathbf{k}'-\mathbf{q},-1}^\dagger \hat{B}_{\mathbf{k}',-1} \hat{B}_{\mathbf{k},+1} \\ & + R \sum_{\mathbf{k}\mathbf{k}' \neq 0} \sum_{\mathbf{q}} \hat{B}_{\mathbf{k}+\mathbf{q},\sigma}^\dagger \hat{B}_{\mathbf{k}'-\mathbf{q},\sigma}^\dagger \hat{B}_{\mathbf{k}'\sigma} \hat{B}_{\mathbf{k}\sigma} \\ & - G \sum_{\mathbf{k}\mathbf{k}' \neq 0} \sum_{\mathbf{q}} (\hat{B}_{\mathbf{k}+\mathbf{q},\sigma}^\dagger \hat{B}_{\mathbf{k}'-\mathbf{q},\sigma}^\dagger \hat{B}_{\mathbf{k}'\sigma} \hat{c}_{\mathbf{k}\sigma} + \text{h.c.}), \end{aligned} \quad (10.25)$$

where $\hat{c}_{\mathbf{k}\sigma}^\dagger$ and $\hat{c}_{\mathbf{k}\sigma}$ are the operators of a photon of momentum \mathbf{k} and angular momentum (circular polarization) $\sigma = \pm 1$, $\omega_{\mathbf{k}}$ ($\Omega_{\mathbf{k}}$) is the energy dispersion of a photon (a boson). Kuwata-Gonokami calculated the matrix elements of the cavity-polariton scatterings and found that this simple phenomenology can reproduce all the experimental results, in particular, the polarization dependence. Ratios among W , R , and G can be determined by comparing with the experiments. This shows that the bosonic treatments are thought as one of the effective tools for the description of excitonic nonlinearities.

We here note that the 1s Wannier exciton is not a pure-boson but a quasi-boson. Its creation operator is defined through $\hat{b}_{\mathbf{K}\mu\nu}^\dagger \equiv \sum_{\mathbf{k}} f^{(2)}(\mathbf{k}) \hat{a}_{\alpha\mathbf{K}+\mathbf{k},\mu}^\dagger \hat{d}_{\beta\mathbf{K}-\mathbf{k},\nu}^\dagger$, where μ (ν) is a quantum number of the total angular momentum of an electron (a hole), $\alpha = m_e^*/(m_e^* + m_h^*)$, $\beta = m_h^*/(m_e^* + m_h^*)$, and $f^{(2)}(\mathbf{k}) = (\sqrt{2\pi}/L) (1 + k^2/4)^{-3/2}$ describes the Fourier transform of the electron-hole relative wavefunction in $d = 2$. This exciton operator satisfies the commutation relation:

$$[\hat{b}_{\mathbf{K}_1\mu_1\nu_1}, \hat{b}_{\mathbf{K}_2\mu_2\nu_2}^\dagger] = \delta_{\mathbf{K}_1\mathbf{K}_2} \delta_{\mu_1\mu_2} \delta_{\nu_1\nu_2} - \hat{\xi}(\mathbf{K}_1\mu_1\nu_1 | \mathbf{K}_2\mu_2\nu_2), \quad (10.26)$$

where

$$\begin{aligned} \hat{\xi}(\mathbf{K}_1\mu_1\nu_1 | \mathbf{K}_2\mu_2\nu_2) = & \delta_{\nu_1\nu_2} \sum_{\mathbf{k}} f^{(2)}(\mathbf{k}) f^{(2)}(\alpha\mathbf{K}_2 - \alpha\mathbf{K}_1 + \mathbf{k}) \\ & \times \hat{d}_{\beta\mathbf{K}_2 - \beta\mathbf{K}_1 + \alpha\mathbf{K}_2 + \mathbf{k}, \mu_2}^\dagger \hat{a}_{\alpha\mathbf{K}_1 + \mathbf{k}, \mu_1} \\ & + \delta_{\mu_1\mu_2} \sum_{\mathbf{k}} f^{(2)}(\mathbf{k}) f^{(2)}(\alpha\mathbf{K}_1 - \alpha\mathbf{K}_2 + \mathbf{k}) \\ & \times \hat{d}_{\alpha\mathbf{K}_2 - \alpha\mathbf{K}_1 + \beta\mathbf{K}_1 - \mathbf{k}, \nu_2}^\dagger \hat{d}_{\beta\mathbf{K}_1 - \mathbf{k}, \nu_1}. \end{aligned} \quad (10.27)$$

The second term $\hat{\xi}$ on the right-hand side describes the deviation from the pure-boson commutation relation. Note here that two single-exciton states are orthogonal: $\langle 0 | \hat{b}_{\mathbf{K}_1 \mu_1 \nu_1} \hat{b}_{\mathbf{K}_2 \mu_2 \nu_2}^\dagger | 0 \rangle = \delta_{\mathbf{K}_1 \mathbf{K}_2} \delta_{\mu_1 \mu_2} \delta_{\nu_1 \nu_2}$, while that of two two-exciton states is no longer orthogonal:

$$\begin{aligned} & \langle 0 | \hat{b}_{\mathbf{K}_4 \mu_4 \nu_4} \hat{b}_{\mathbf{K}_3 \mu_3 \nu_3} \hat{b}_{\mathbf{K}_1 \mu_1 \nu_1}^\dagger \hat{b}_{\mathbf{K}_2 \mu_2 \nu_2}^\dagger | 0 \rangle \\ &= \delta_{\mathbf{K}_1 \mathbf{K}_3} \delta_{\mu_1 \mu_3} \delta_{\nu_1 \nu_3} \delta_{\mathbf{K}_2 \mathbf{K}_4} \delta_{\mu_2 \mu_4} \delta_{\nu_2 \nu_4} + \delta_{\mathbf{K}_1 \mathbf{K}_4} \delta_{\mu_1 \mu_4} \delta_{\nu_1 \nu_4} \delta_{\mathbf{K}_2 \mathbf{K}_3} \delta_{\mu_2 \mu_3} \delta_{\nu_2 \nu_3} \\ &\quad - \delta_{\mathbf{K}_1 + \mathbf{K}_2, \mathbf{K}_3 + \mathbf{K}_4} (\delta_{\mu_1 \mu_3} \delta_{\nu_1 \nu_4} \delta_{\mu_2 \mu_4} \delta_{\nu_2 \nu_3} + \delta_{\mu_1 \mu_4} \delta_{\nu_1 \nu_3} \delta_{\mu_2 \mu_3} \delta_{\nu_2 \nu_4}) \\ &\quad \times F(\alpha(\mathbf{K}_1 - \mathbf{K}_3), \beta(\mathbf{K}_2 - \mathbf{K}_3)), \end{aligned} \quad (10.28)$$

where $|0\rangle$ is the ground state and $F(\mathbf{K}, \mathbf{K}') \equiv \sum_{\mathbf{k}} f^{(2)}(\mathbf{K} + \mathbf{k}) f^{(2)}(\mathbf{K}' + \mathbf{k}) f^{(2)}(\mathbf{k}) f^{(2)}(\mathbf{K} + \mathbf{K}' + \mathbf{k})$. The quasi-boson nature and the non-orthogonality of two two-exciton states are key points for formulating new boson-picture theories of a few exciton systems.

Before the above phenomenological WIB model, there were many theoretical trials to describe many-exciton systems in terms of bosons in the study of the exciton Bose-Einstein condensation [35]. Recently, theoreticians are interested in how to bosonize a few exciton systems with the spin degrees of freedom, how to evaluate interactions among the bosons and photons [36], and how to cope with the deviation of the pure-boson commutation relations. Inoue *et al* approximated the 1s Wannier exciton in $d = 2$ as a pure-boson to derive the attractive (W) and repulsive (R) interactions between the bosons depending on the total angular momentum [37]. This means that the phenomenological cavity-polariton model gets a basis from a microscopic viewpoint. Simultaneously, the Usui transformation [38] for quasi-bosons [39], the Marumori-mapping technique [40], and an exact mapping to a pure-boson space [41] are recently investigated [42]. We here note that the validity and limitations of the boson-picture treatment should always be recognized; for example, the decay processes and the detuning effects are hard to be incorporated there.

Here we make a comment on the bosonization. In the ‘exciton bosonization’, as shown in figure 10.4(a), two fermions in different bands, an electron in a conduction band and a hole in a valence band across the energy gap, are simultaneously mapped to a boson. In the low-density limit of excitons, the bosonic excitons satisfy the pure-boson commutation relation, while boson approximation becomes worse in the high-density limit, as shown in equation (10.26). On the other hand, as will be shown in the following sections, another bosonization is frequently used to describe low-energy excitations near the Fermi level in degenerate electron systems [43, 44]. As shown in figure 10.4(b), a pair of a particle and a vacancy across the Fermi level is treated as a long-wavelength boson, i.e. two fermions in the same band are mapped to a boson. In other words, the charge and spin density fluctuations of an electron gas (or a hole gas) are quantized to be bosons. In particular, the low-energy bosons near the Fermi level in $d = 1$ satisfy the pure-boson commutation relation, thus they are called the Tomonaga bosons [45, 46]. With the use of this property, one

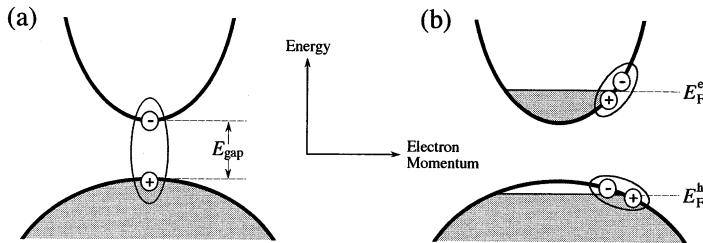


Figure 10.4. Concepts of the bosonizations for two fermions. (a) Exciton bosonization. A conduction electron (– sign) and a valence hole (+ sign) are combined to be a composite boson. Actually, superposition of several momentum states is carried out. (b) Bosonization near the Fermi level E_F in degenerate electron–hole systems. An electron (– sign) above E_F and a vacancy (+ sign) below E_F are a pairwise excitation described as a long-wavelength boson.

of exactly-solvable quantum models, the Tomonaga–Luttinger (TL) model has been proposed, in which interacting fermions can be described as independent bosons [45].

10.5 Excitonic many-body problems

We consider two topics in one-dimensional systems as examples of excitonic many-body problems: (i) high-density electron–hole system in $d = 1$ under irradiation by an intense light pulse and (ii) a $d = 1$ exciton (the Mahan exciton) in degenerate (doped) semiconductors or in metals. What will happen as increasing the number of excitons in semiconductors is a long-standing problem, in which both the quantum many-body effects and the non-equilibrium characters should be considered [47, 48]. Since the mean-field approximation is invalid in low dimensions as mentioned earlier, here we shall make use of the exactly-solvable models to avoid ambiguities due to some approximations.

10.5.1 Exciton Bose–Einstein condensation and density waves in one dimension

With irradiation of an intense laser pulse to undoped semiconducting wires, we can explore the many-body effects in a $d = 1$ high-density electron–hole system, whose quasi-equilibrium states are treated as a two-band TL liquid (properties of the TL liquid will be explained afterwards). Densities of electrons (in the conduction band) and holes (in the valence band) are assumed to be equal to each other, both of which are high enough so that the Fermi points are well defined in both bands. We further assume that this system is in quasi-equilibrium within a radiative lifetime (rigid Fermi-sea picture) [49]. We here concentrate on the

$d = 1$ high-density electron-hole system with the same Fermi wavevector k_F for both the conduction and valence bands.

In this system, low-energy excitations near each quasi-Fermi level are important for quantum orderings and their dynamical responses. Then the bosonization method is used to treat the collective excitations of the degenerate electrons and holes. To this end, we linearize the dispersions near the Fermi points with the Fermi velocities of the conduction and valence bands, $v_F^e \propto (m_e^*)^{-1}$ and $v_F^h \propto (m_h^*)^{-1}$. Each band consists of two branches corresponding to the right-moving ($j = 1$) and the left-moving ($j = 2$) particles. The interparticle interaction matrix element is assumed to be spin-independent and parameterized to the nine interaction matrix elements: g_i^e , g_i^h , and g_i^{eh} for $i = 1, 2, 4$. We study here the forward-scattering model, which includes essentially only intrabranch interactions, g_2 's (g_2^e , g_2^h , and g_2^{eh}) and is solved exactly.

One advantage of the bosonization method comes from the fact that electron (or hole) field operators can be written in terms of the phase fields, or equivalently, boson creation and annihilation operators, for example,

$$\begin{aligned} \hat{\psi}_{1\sigma}^e(x) &= (2\pi\alpha)^{-1/2} e^{ik_F x} \\ &\times \exp \left[\frac{i}{2} \{ \hat{\theta}_e(x) + \hat{\theta}_e^-(x) + \sigma [\hat{\phi}_e(x) + \hat{\phi}_e^-(x)] \} + i\hat{\varphi}_{1\sigma}^e \right], \end{aligned} \quad (10.29)$$

$$\begin{aligned} \hat{\psi}_{2\sigma}^e(x) &= (2\pi\alpha)^{-1/2} e^{-ik_F x} \\ &\times \exp \left[-\frac{i}{2} \{ \hat{\theta}_e(x) - \hat{\theta}_e^-(x) + \sigma [\hat{\phi}_e(x) - \hat{\phi}_e^-(x)] \} + i\hat{\varphi}_{2\sigma}^e \right], \end{aligned} \quad (10.30)$$

for the electron fields of branch $j = 1$ and 2, respectively, where α is a cutoff and $\hat{\varphi}_{j\sigma}^e$ is necessary for ensuring the anticommutation relation of $\hat{\psi}_{j\sigma}^e$ with different j and σ . Using the phase variables, we obtain a phase Hamiltonian of the two-band TL liquid: $\hat{\mathcal{H}} \equiv \hat{\mathcal{H}}^{\text{charge}} + \hat{\mathcal{H}}^{\text{spin}}$, where

$$\begin{aligned} \hat{\mathcal{H}}^{\text{charge}} &= \int dx \{ A_e \hat{P}_e^2(x) + C_e [\nabla \hat{\theta}_e(x)]^2 + A_h \hat{P}_h^2(x) + C_h [\nabla \hat{\theta}_h(x)]^2 \\ &+ 2C_{eh} \nabla \hat{\theta}_e(x) \nabla \hat{\theta}_h(x) \}, \end{aligned} \quad (10.31)$$

$$\begin{aligned} \hat{\mathcal{H}}^{\text{spin}} &= \int dx \{ B_e \hat{\Pi}_e^2(x) + D_e [\nabla \hat{\phi}_e(x)]^2 + B_h \hat{\Pi}_h^2(x) + D_h [\nabla \hat{\phi}_h(x)]^2 \}. \end{aligned} \quad (10.32)$$

Here $\hat{\theta}_e(x)$ [$\hat{\theta}_h(x)$] is the charge phase of the conduction (valence) electrons, whose conjugate momentum is $\hat{P}_e(x) \equiv -(4\pi)^{-1} \partial \hat{\theta}_e^-(x) / \partial x$ [$\hat{P}_h(x) \equiv -(4\pi)^{-1} \partial \hat{\theta}_h^-(x) / \partial x$], while $\hat{\phi}_e(x)$ [$\hat{\phi}_h(x)$] is the spin phase, whose conjugate momentum is $\hat{\Pi}_e(x) \equiv -(4\pi)^{-1} \partial \hat{\phi}_e^-(x) / \partial x$ [$\hat{\Pi}_h(x) \equiv -(4\pi)^{-1} \partial \hat{\phi}_h^-(x) / \partial x$]. Coefficients are given by $A_v \equiv 2\pi(\bar{v}_F^\nu - g_2^\nu)$, $B_v \equiv 2\pi\bar{v}_F^\nu$, $C_v \equiv (\bar{v}_F^\nu + g_2^\nu)/8\pi$, $D_v \equiv \bar{v}_F^\nu/8\pi$, and $C_{eh} \equiv g_2^{eh}/4\pi$, where $\nu = e$ or h and $\bar{v}_F^\nu \equiv v_F^\nu + g_4^\nu$ is the normalized Fermi velocity. The charge part of the forward-scattering

Hamiltonian, $\hat{\mathcal{H}}^{\text{charge}}$, is diagonalized by the unitary transformation: $(\hat{\theta}_e, \hat{\theta}_h) \rightarrow (\hat{\theta}_1, \hat{\theta}_2)$ via

$$\begin{bmatrix} \hat{\theta}_e \\ \hat{\theta}_h \end{bmatrix} = \begin{bmatrix} \sqrt{A_e} \cos \Theta & -\sqrt{A_e} \sin \Theta \\ \sqrt{A_h} \sin \Theta & \sqrt{A_h} \cos \Theta \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}, \quad (10.33)$$

with Θ being given by $\tan 2\Theta \equiv 2C_{eh}\sqrt{A_e A_h}/(A_e C_e - A_h C_h)$. Consequently, the diagonalized forward-scattering Hamiltonian becomes $\hat{\mathcal{H}}^{\text{forward}} = \int dx \sum_{i=1,2} \{\hat{P}_i^2(x) + [v_{\text{charge}}^{(i)} \nabla \hat{\theta}_i(x)]^2\} + \hat{\mathcal{H}}^{\text{spin}}$, where \hat{P}_i is the conjugate momentum for $\hat{\theta}_i$ ($\hat{\mathcal{H}}^{\text{spin}}$ is already diagonalized). We find that the motions of four phases, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\phi}_e$, and $\hat{\phi}_h$, are described by massless acoustic modes, which have linear dispersions in their excitation spectra. The velocities of the transformed charge phases, $\hat{\theta}_1$ and $\hat{\theta}_2$, are given by

$$v_{\text{charge}}^{(i)} = \{2(A_e C_e + A_h C_h) \pm 2[(A_e C_e - A_h C_h)^2 + 4C_{eh}^2 A_e A_h]^{1/2}\}^{1/2}, \quad (10.34)$$

respectively, where $i = 1$ (2) corresponds to the $+$ ($-$) sign of the right-hand side. On the other hand, velocities of the spin phases, $\hat{\phi}_e$ and $\hat{\phi}_h$, are v_F^e and v_F^h , respectively: $v_{\text{spin}}^v = v_F^v$.

In $d = 1$ quantum many-particle systems, the competition among various Fermi-surface instabilities occurs through quantum fluctuations even at zero temperature. There are sixteen possible order parameters, $\hat{\mathcal{O}}(x)$'s. All the correlation functions, $\langle \hat{\mathcal{O}}(x, \tau) \hat{\mathcal{O}}^\dagger(0, 0) \rangle$, behave like $[\max(x, \tau)]^{-\eta}$ for large x and large imaginary time τ . The exponent η for each order parameter is given in [50]. The phase diagram of the quasi-thermal-equilibrium state was given, which is divided into following four regions, as shown in figure 10.5.

- (a) *Exciton Bose–Einstein-condensate phase.* When the repulsive interaction between conduction electrons and valence ones is strong (positive and large g_2^{eh}), i.e. the strong attraction between electrons and holes, the exciton Bose–Einstein condensate (exciton BEC) at $K = 0$ predominates. Spin-singlet and triplet exciton Bose condensation are degenerate in our spin-independent forward-scattering model. The exciton density wave at $K = \pm 2k_F$ cannot prevail in the whole plane.
- (b) *Density-wave phase.* When the repulsion between the electrons within each band is strong (positive and large g_2^e or g_2^h), the charge-density wave (CDW) and the spin-density wave (SDW) predominate. The valence-band electrons (i.e. holes) with heavier mass have the stronger tendency toward the ordering. When the velocity ratio $|v_F^h/v_F^e|$ is decreased from unity, the CDW and SDW region extends and invades the exciton BEC region.
- (c) *Ordinary superconductivity phase.* The attractive interactions, i.e. negative g_2 's, yield conventional superconductivity (SC). The negative g_2^h results in intraband pairing with zero total momentum $K = 0$. The valence-band electrons with heavier mass have the stronger tendency toward the ordering.
- (d) *Unconventional superconductivity phase.* On the other hand, the negative g_2^{eh} causes interband Cooper pairing at $K = \pm 2k_F$. This unconventional

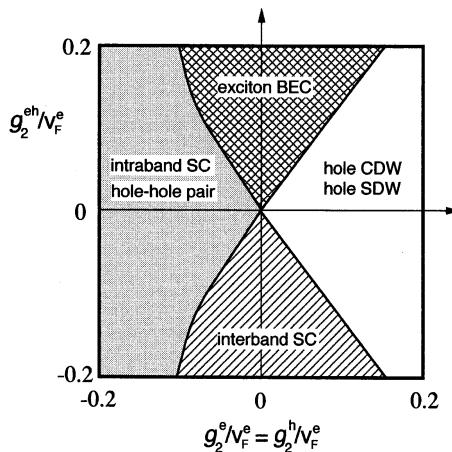


Figure 10.5. A quasi-equilibrium phase diagram of one-dimensional high-density electron-hole systems in the low-temperature limit for $v_F^e/v_F^h = 2$. The upper-right region ($g_2^e/v_F^e = g_2^h/v_F^h > 0$ and $g_2^{eh}/v_F^e > 0$) corresponds to the normal situation. CDW (SDW) stands for the charge-density wave (spin-density wave), BEC the Bose-Einstein condensate, and SC means the superconductivity. No relaxation effect is taken into account here. (After [50]).

superconductivity is a peculiar feature of the $d = 1$ system. Singlet and triplet SC's are degenerate in both the above cases.

We shall discuss the optical absorption spectrum of the degenerate electron-hole system in $d = 1$ [51]. The absorption spectrum, $W_{\text{OPA}}(\omega)$, is related to the Fourier transform of the correlation function, $\langle \hat{P}(t)\hat{P}^\dagger(0) \rangle$, where the polarization \hat{P} (namely the dipole operator) is given by $\hat{P}(t) = |M| \int dx \sum_\sigma \hat{\psi}_\sigma^{e\dagger}(x, t) \hat{\psi}_\sigma^h(x, t) e^{iE_0 t} + \text{h.c.}$, where $\hat{\psi}_\sigma^\nu(x, t) = \exp[i\hat{H}t] \hat{\psi}_\sigma^\nu(x) \exp[-i\hat{H}t]$ is the Heisenberg representation of the field operator ($\hat{\psi}_\sigma^\nu = \hat{\psi}_{1\sigma}^\nu + \hat{\psi}_{2\sigma}^\nu$), M is an interband matrix element (assumed to be constant), and E_0 is the energy of the absorption and emission edge in the single-electron picture. The correlation function of the dipole operator contains the correlation function $C_{\text{SEBC}}(t) \equiv \langle \hat{O}_{\text{SEBC}}(x, t) \hat{O}_{\text{SEBC}}^\dagger(0, 0) \rangle$ of the order parameter describing the spin-singlet exciton BEC, $\hat{O}_{\text{SEBC}}(x) \equiv \sum_{j=1,2} \sum_\sigma \hat{\psi}_{j\sigma}^{e\dagger}(x) \hat{\psi}_{j\sigma}^h(x)$. For large t , the correlation function behaves like $C_{\text{SEBC}}(t) \sim t^{-\eta_{\text{SEBC}}}$. This means that the optical spectrum shows the power-law singularity like $W_{\text{OPA}}(\omega) \sim (\hbar\omega - E_F)^\beta \Theta(\hbar\omega - E_F)$ in the vicinity of E_F , where E_F is the absorption-edge energy and β is the critical exponent:

$\beta = \eta_{\text{SEBC}} - 2 = \beta^{\text{ex}} + \beta^{\text{c-oc}} + \beta^{\text{v-oc}}$, where

$$\beta^{\text{ex}} = -\frac{\pi[v_{\text{charge}}^{(1)} - v_{\text{charge}}^{(2)}]}{2\sqrt{A_e A_h}} \left[1 + \frac{A_e A_h}{4\pi^2 v_{\text{charge}}^{(1)} v_{\text{charge}}^{(2)}} \right] \sin 2\Theta, \quad (10.35)$$

$$\beta^{\text{c-oc}} = \frac{A_h s_1}{8\pi v_{\text{charge}}^{(1)} v_{\text{charge}}^{(2)}} + \frac{\pi s_2}{2A_h} - \frac{1}{2}, \quad (10.36)$$

$$\beta^{\text{v-oc}} = \frac{A_e s_2}{8\pi v_{\text{charge}}^{(1)} v_{\text{charge}}^{(2)}} + \frac{\pi s_1}{2A_e} - \frac{1}{2}. \quad (10.37)$$

Here $s_i \equiv \frac{1}{2}[v_{\text{charge}}^{(1)} + v_{\text{charge}}^{(2)} \pm (v_{\text{charge}}^{(1)} - v_{\text{charge}}^{(2)}) \cos 2\Theta]$ with $i = 1$ (2) corresponding to the + (−) sign. This power-law anomaly in optical spectra near an edge energy is one of the Fermi-edge singularities (FESs).

We stress that the FES exponent, β , is determined by the correlation exponent, η_{SEBC} , of the order parameter describing the singlet-exciton BEC. Roughly speaking, the divergent edge spectrum (corresponding to $\beta < 0$) can be observed when the electron–hole interaction is attractive ($g_2^{\text{eh}} > 0$ and $g_4^{\text{eh}} > 0$) and rather weak electron–electron and hole–hole correlations. When the excitons lie completely at the condensation state at zero temperature, an optical spectrum is expected to show a δ -function-like peak if we can ignore the quantum fluctuations. Actually, however, quantum fluctuation can never be neglected in $d = 1$ systems and it tends to destroy the exciton condensation. Consequently, δ -function-like peak spectrum vanishes and is replaced by the power-law peak appears, which is really the FES and is a remnant of the exciton BEC. In this sense, the FES results from a fluctuating condensed state of many $d = 1$ excitons. Effects of the electron–hole backward scattering and the randomness were discussed in [50, 51].

10.5.2 Optical responses of the Tomonaga–Luttinger liquid: the Mahan exciton

When interactions among particles are of a short-range type and when the backward, umklapp and spin–flip scatterings can be neglected, critical properties of quantum $d = 1$ electrons are universally described as the TL liquid. The main aim of this section is to show that the optical responses of the TL liquid are also useful tools for investigating the Mahan exciton and the critical properties of $d = 1$ degenerate electrons.

Several optical transition probabilities reflect the final-state interaction due to an optically-generated hole potential. This interaction brings about two kinds of intrinsic dynamical effects in degenerate electron systems: an orthogonality catastrophe and a many-body excitonic correlation [52]. The former suppresses the transition but the latter enhances it. This is a characteristic point in a contrast to those in semiconductors. The most striking feature in optical responses of the TL liquids is the power-law singularity, which is characterized by some

critical exponents. We pay attention to the valence-band photoemission spectrum, the core-level photoemission spectrum, and the OPA spectrum. These optical transition processes yield the power-law anomaly in the edge spectra with different critical exponents [53, 54].

The $d = 1$ degenerate electron systems near the Fermi level are characterized by two (right- and left-moving) branches of electrons. When the interparticle interaction in one dimension is of the short-range type, the Coulomb scattering strengths are represented by constant parameters: g_2 , g_4 , and g_1 with conventional notation [46], where g_2 and g_4 correspond to the strengths of inter- and intra-branch forward scatterings, and g_1 , the backward-scattering strength. Then the backward scattering is neglected because $|g_2| \simeq |g_4| \gg |g_1|$. We further classify the interaction parameters in terms of the mutual spin orientation; the scattering between parallel-spin (\parallel) particles and that between antiparallel-spin (\perp) ones are labelled separately as $g_{i\parallel}$ and $g_{i\perp}$, respectively, for $i = 2, 4$. For convenience, we define the parameters, K_ρ and K_σ , for characterizing the interaction, i.e. $K_\nu \equiv [(v_F + g_{4\nu} - g_{2\nu})/(v_F + g_{4\nu} + g_{2\nu})]^{1/2}$, for $\nu = \rho$ (charge degree of freedom) or σ (spin) with $g_{i\rho} \equiv (g_{i\parallel} + g_{i\perp})/2$, $g_{i\sigma} \equiv (g_{i\parallel} - g_{i\perp})/2$, and v_F is the spin-independent Fermi velocity.

According to usual bosonization, we linearize the conduction-band dispersion near the Fermi points: $\epsilon(k) \simeq \sum_j \epsilon_j(k)$ with $\epsilon_j(k) \equiv \pm v_F(k \mp k_F)$, where k_F is the Fermi wavenumber and $j = 1$ ($j = 2$) corresponds to the right-(left-) moving branch. In this case, the $d = 1$ electron systems are described by the following Hamiltonian: $\hat{H}^e = \hat{H}_0^e + \hat{H}_{\text{int}}^e$, where

$$\hat{H}_0^e = \frac{2\pi v_F}{L} \sum_s \sum_{p>0} [\hat{\varrho}_{1s}(p)\hat{\varrho}_{1s}(-p) + \hat{\varrho}_{2s}(-p)\hat{\varrho}_{2s}(p)], \quad (10.38)$$

$$\begin{aligned} \hat{H}_{\text{int}}^e = & \frac{\pi}{L} \sum_{s,s'} (g_{2\parallel}\delta_{s,s'} + g_{2\perp}\delta_{s,-s'}) \sum_{p>0} [\hat{\varrho}_{1s}(p)\hat{\varrho}_{2s'}(-p) + \hat{\varrho}_{1s}(-p)\hat{\varrho}_{2s'}(p)] \\ & + \frac{\pi}{L} \sum_{s,s'} (g_{4\parallel}\delta_{s,s'} + g_{4\perp}\delta_{s,-s'}) \sum_{p>0} [\hat{\varrho}_{1s}(p)\hat{\varrho}_{1s'}(-p) + \hat{\varrho}_{2s}(-p)\hat{\varrho}_{2s'}(p)], \end{aligned} \quad (10.39)$$

where $\hat{\varrho}_{js}(p) \equiv \sum_k \hat{a}_{j,k+p,s}^\dagger \hat{a}_{jks}$ is the spin-dependent density-fluctuation operator for spin $s = \uparrow, \downarrow = \pm 1$ in the j th branch, which obeys boson commutation relations. The electronic Hamiltonian, \hat{H}^e , is described also as a sum of two diagonal parts describing the charge- and spin-density fluctuations: $\hat{H}^e = \sum_{\nu=\rho,\sigma} \sum_{p>0} v_\nu p (\hat{\alpha}_{\nu p}^\dagger \hat{\alpha}_{\nu p} + \hat{\beta}_{\nu p}^\dagger \hat{\beta}_{\nu p})$. Here $\hat{\alpha}_{\rho p}$ and $\hat{\beta}_{\rho p}$ ($\hat{\alpha}_{\sigma p}$ and $\hat{\beta}_{\sigma p}$) are boson operators for the charge (spin) fluctuation, which are related to the density-fluctuation operators via $\hat{\alpha}_{\nu p}^\dagger = (\pi/Lp)^{1/2} \sum_{s=\pm 1} [\hat{\varrho}_{1s}(p) \cosh \gamma_{\nu p} \pm \hat{\varrho}_{2s}(p) \sinh \gamma_{\nu p}]$ and $\hat{\beta}_{\nu p}^\dagger = (\pi/Lp)^{1/2} \sum_{s=\pm 1} [\hat{\varrho}_{2s}(p) \cosh \gamma_{\nu p} \pm \hat{\varrho}_{1s}(p) \sinh \gamma_{\nu p}]$ for $\nu = \rho, \sigma$ with $\cosh(2\gamma_{\nu p}) = v_F(1 + g_{2\nu})/v_\nu$ and $\sinh(2\gamma_{\nu p}) = v_F g_{2\nu}/v_\nu$. Both the charge and spin density fluctuations have linear gapless dispersions with

constant velocities:

$$v_\nu = \frac{2(v_F + g_{4\nu})}{K_\nu + K_\nu^{-1}} = [(v_F + g_{4\nu})^2 - g_{2\nu}^2]^{1/2}, \quad (10.40)$$

for $\nu = \rho, \sigma$.

Peculiar features of the TL liquid are found in its dynamical properties, which are characterized by anomalous power laws in correlation functions. We shall study the single-electron Green function: $G_{js}^e(x, t) = -i\Theta(t)\langle \hat{\psi}_{js}(x, t)\hat{\psi}_{js}^\dagger(0, 0) \rangle$, where $\Theta(t)$ is the unit-step function and $\hat{\psi}_{js}(x, t)$ is the Heisenberg representation of the field operator $\hat{\psi}_{js}(x)$ for an electron in the j th branch with spin s . Average is made by the Fermi vacuum at zero temperature. Using the diagonalized Hamiltonian, we have

$$G_{js}^e(x, t) = \frac{e^{\pm ik_F x}}{2\pi} \lim_{\delta \rightarrow 0} \frac{\Lambda + i(v_F t \mp x)}{\delta + i(v_F t \mp x)} \times \prod_{\nu=\rho,\sigma} \frac{1}{\sqrt{\Lambda + i(v_\nu t \mp x)}} \left[\frac{\Lambda^2}{(\Lambda + iv_\nu t)^2 + x^2} \right]^{\gamma_{TL}^\nu}, \quad (10.41)$$

where $j = 1$ (2) corresponds to upper (lower) sign of the right-hand side, Λ and δ are cutoffs, and the exponent is $\gamma_{TL}^\nu = (1/8)(K_\nu + K_\nu^{-1} - 2) \geq 0$. Equation (10.41) gives a universal behaviour of the Green function, which is independent of detailed cutoff forms. For $t = 0$, the single-electron Green function decays as $G_{js}^e(x, 0) \sim x^{-\beta_{TL}-1}$ with the exponent $\beta_{TL} \equiv 2(\gamma_{TL}^\rho + \gamma_{TL}^\sigma) \geq 0$, which appears in all single-electron properties; for example, the momentum distribution function $n(k) \sim 1/2 - C|k - k_F|^\beta_{TL}\text{sign}(k - k_F)$ (here C is a positive constant) and the single-particle density of states $N(\omega) \sim |\omega|^{\beta_{TL}}$. The critical exponent β_{TL} is rewritten as

$$\beta_{TL} = \frac{1}{4}(K_\rho + K_\rho^{-1} + K_\sigma + K_\sigma^{-1}) - 1 = \frac{1}{2} \sum_{\nu=\rho,\sigma} \left[1 - \left(\frac{g_{2\nu}}{v_F + g_{4\nu}} \right)^2 \right]^{-1/2} - 1 \geq 0. \quad (10.42)$$

Before the optical transition occurs, we assume that the partially-filled band has N electrons forming the TL liquid and there is no positive hole in any levels and bands. Hereafter this initial state is called the N -electron 0-hole state, denoted as the $(N, 0)$ state. In this section, we consider three types of final states after optical transitions, as shown in [figure 10.6](#).

- (a) *From $(N, 0)$ to $(N - 1, 0)$.* After an electron in the N -electron TL system is excited to be emitted as a photoelectron, the final state is the $(N - 1, 0)$ state, where no optical hole is generated and the number of electrons only decreases by one.
- (b) *From $(N, 0)$ to $(N, 1)$.* When an electron in a core (valence) level is excited by a larger-energy photon to be detected as a photoelectron, the final state

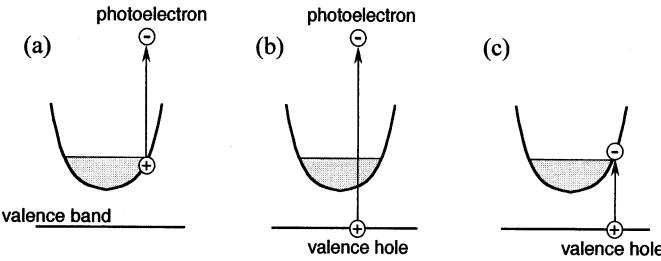


Figure 10.6. Three types of optical processes reflecting several critical exponents which characterize the Tomonaga–Luttinger liquid: (a) photoemission, (b) core-level photoemission and (c) one-photon absorption. In these figures, no dispersion of the valence band (hole localization) is assumed. The orthogonality catastrophe appears in (b) and (c). The Mahan exciton effects take place only in (c).

is the $(N, 1)$, where the number of electrons in the TL liquid remains unchanged and a hole is abruptly generated. Here the Auger processes are neglected.

- (c) *From $(N, 0)$ to $(N + 1, 1)$.* When an electron in a core level is excited by a photon to an empty state of the partially-filled band, the final state is the $(N + 1, 1)$, where an electron is added to the TL liquid band and a hole is suddenly generated, i.e. the numbers of both electron and hole change simultaneously by one.

10.5.2.1 The valence-band photoemission: $(N, 0) \rightarrow (N - 1, 0)$

The valence-band photoemission spectroscopy measures the energy distribution of the photoelectrons ejected from a partially-filled band by absorbing photons of a fixed frequency (figure 10.6(a)). The valence-band photoemission spectrum $P_{\text{PES}}(\epsilon)$ as a function of the photoelectron energy ϵ (measured from the Fermi energy E_F) reflects directly the density of states $N(\omega)$. Then $P_{\text{PES}}(\epsilon) \sim \epsilon^{\beta_{\text{TL}}} \Theta(\epsilon)$. The critical exponent β_{TL} is given by equation (10.42). The valence-band photoemission spectrum can be a direct evidence for the absence of a discontinuous jump at $k = k_F$ in the momentum distribution of the TL liquid. Thus non-zero β_{TL} is an evidence of the TL liquid.

The valence-band photoemission spectrum, shown in figure 10.7(a), is the Fourier transform of the time-domain single-electron Green function $G_{js}^e(t) = -i\Theta(t)\langle e^{it\hat{\mathcal{H}}^e/\hbar}\hat{\psi}_{js}e^{-it\hat{\mathcal{H}}^e/\hbar}\hat{\psi}_{js}^\dagger\rangle \sim t^{-\beta_{\text{TL}}-1}$, that is, $P_{\text{PES}}(\epsilon) \sim (E_F - \epsilon)^{\beta_{\text{TL}}}$. Thus, the valence-band photoemission spectroscopy is an effective tool for investigating the TL liquid properties, particularly its single-electron properties.

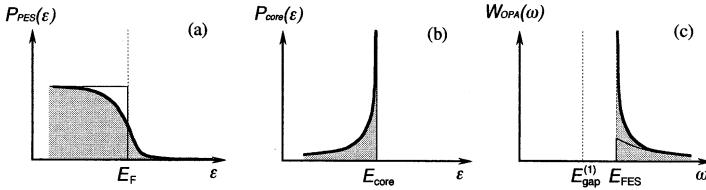


Figure 10.7. Schematic spectra corresponding to three optical processes in figure 10.6. $\beta_{\text{FES}} < 0$ is assumed in (c). The thin curves mean the spectra with no interaction, leading to (a) a step-like shape, (b) a δ -function shape and (c) a JDOS shape.

10.5.2.2 The core-level photoemission: $(N, 0) \rightarrow (N, 1)$

We mention here the orthogonality theorem. When dynamical and local perturbations are applied to degenerate electron systems, low-energy excitations near the Fermi energy come into play and give rise to an infrared divergence [52]. The overlap integral between the total electronic wavefunctions, $|\Psi_i\rangle$ and $|\Psi_f\rangle$, of the Fermi sea without and with a local potential becomes zero as the system size goes to infinity, i.e. these two states are orthogonal: $|\langle\Psi_f|\Psi_i\rangle|^2 = n_F^{-\beta_{\text{oc}}/2}$. Here n_F is the number of (*s*-wave) electrons and β_{oc} is the critical exponent for the orthogonality catastrophe. To detect phenomena related to the orthogonality property, a sudden appearance of an external local potential is necessary. This can be done in the core-level photoemission spectroscopy.

The core-level photoemission spectroscopy measures the energy distribution of the photoelectrons ejected from a core level in a metal by absorbing photons of a fixed frequency (figure 10.6(b)). If the existence of conduction electrons is neglected, only a sharp line will be observed. When the photon energy is high enough for the photoelectron to leave the metal instantly without being affected by a final-state interaction, the sharp line broadens with a long tail on its lower-energy side. The core-level photoemission spectrum, $P_{\text{core}}(\epsilon)$, the probability of finding an ejected electron at an energy ϵ lower than the main peak, is a direct measure of the probability that the conduction electrons will be excited with an excitation energy ϵ due to the hole potential.

When an optical hole with spin $-s$ (an excited electron with spin s) is created, the electron–hole Coulomb attraction abruptly appears. Here the hole is assumed to be localized. The Hamiltonian describing the electron–hole correlation is given by $\hat{\mathcal{H}}_{\text{int}}^h(s) = \hat{\mathcal{H}}_\rho^h + \hat{\mathcal{H}}_\sigma^h(s)$ with [55]

$$\mathcal{H}_\rho^h = -\frac{g_{2\rho}^h}{2} \left(\frac{\pi K_\rho}{L} \right)^{1/2} \sum_{p>0} \sqrt{p} (\hat{\alpha}_{\rho p}^\dagger + \hat{\alpha}_{\rho p} + \hat{\beta}_{\rho p}^\dagger + \hat{\beta}_{\rho p}), \quad (10.43)$$

$$\mathcal{H}_\sigma^h(s) = -\frac{g_{2\sigma}^h}{2} \left(\frac{\pi K_\sigma}{L} \right)^{1/2} s \sum_{p>0} \sqrt{p} (\hat{\alpha}_{\sigma p}^\dagger + \hat{\alpha}_{\sigma p} + \hat{\beta}_{\sigma p}^\dagger + \hat{\beta}_{\sigma p}), \quad (10.44)$$

where $g_{2\rho}^h = (g_{2\parallel}^h + g_{2\perp}^h)/2$ and $g_{2\sigma}^h = (g_{2\parallel}^h - g_{2\perp}^h)/2$ denote the strength of the electron-hole forward scattering. Using this electron-hole interaction, we shall calculate the single-hole Green function of the TL liquid in time domain; $G_s^h(t) \equiv -i\Theta(t)e^{-iE_0 t/\hbar}\langle e^{it\hat{\mathcal{H}}^e/\hbar}e^{-it[\hat{\mathcal{H}}^e + \hat{\mathcal{H}}_{int}^h(s)]/\hbar}\rangle \sim t^{-\beta_{oc}}$, where E_0 is the energy separation between the core level and the Fermi level. Here the critical exponent β_{oc} is independent of the hole spin and is given as

$$\begin{aligned}\beta_{oc} &= \frac{1}{64} \sum_{\nu=\rho,\sigma} \left(\frac{g_{2\nu}^h}{v_F + g_{4\nu}} \right)^2 (K_\nu + K_\nu^{-1})^4 \{1 \mp [1 - 4(K_\nu + K_\nu^{-1})^{-2}]^{1/2}\} \\ &= \frac{1}{4} \sum_{\nu=\rho,\sigma} \left(\frac{g_{2\nu}^h}{v_F + g_{4\nu}} \right)^2 \left(1 - \frac{g_{2\nu}}{v_F + g_{4\nu}} \right)^{-1/2} \left(1 + \frac{g_{2\nu}}{v_F + g_{4\nu}} \right)^{-3/2} \geq 0,\end{aligned}\quad (10.45)$$

where $K_\nu < 1$ ($K_\nu > 1$) corresponds to $-$ (+) sign of equation (10.45).

This is reflected in an asymptotic behaviour of the core-level photoemission spectrum near the main peak, as shown in figure 10.7(b), which is given by the Fourier transform of $G_s^h(t)$, i.e. $P_{core}(\epsilon) \sim (E_{core} - \epsilon)^{\beta_{oc}-1}\Theta(E_{core} - \epsilon)$, showing the power-law divergence near $\epsilon \sim 0$ because of $0 \leq \beta_{oc} < 1$. Appearance of the power-law singularity instead of the δ -function peak means that the quasi-particle picture is invalid in the TL liquid, resulting from simultaneous excitation of many-particle vacancy pairs near the Fermi level.

10.5.2.3 The one-photon absorption: $(N, 0) \rightarrow (N + 1, 1)$

We shall consider the OPA process (figure 10.6(c)). The absorption spectrum $W_{OPA}(\omega)$ also exhibits the power-law singularity near the Fermi edge with an exponent β_{FES} . The critical exponent β_{FES} is directly related to a power-law exponent of the current-current correlation function $C_s(t)$ at zero temperature: $C_s(t) = \langle e^{it\hat{\mathcal{H}}^e/\hbar} \hat{\psi}_s e^{-it[\hat{\mathcal{H}}^e + \hat{\mathcal{H}}_{int}^h(s)]/\hbar} \hat{\psi}_s^\dagger \rangle = \langle \hat{\mathcal{J}}_s^\dagger(t) \hat{\mathcal{J}}_s(0) \rangle$ with $\hat{\psi}_s \equiv \hat{\psi}_{1s}(0) + \hat{\psi}_{2s}(0)$. Here the current operator is $\hat{\mathcal{J}}_s(t) \equiv e^{it\hat{\mathcal{H}}^e/\hbar} \sum_{j=1,2} \hat{U}_s \hat{\psi}_j^\dagger e^{-it\hat{\mathcal{H}}^e/\hbar}$, where the unitary operator \hat{U}_s is defined through $\hat{U}_s^\dagger \hat{\mathcal{H}}^e \hat{U}_s = \hat{\mathcal{H}}^e + \hat{\mathcal{H}}_{int}^h(s)$. For large t , $C_s(t)$ decays as a power law: $C_s(t) \sim t^{-\beta_{FES}-1}$. Then the OPA spectrum, which is the Fourier transform of $C_s(t)$, shows the power-law anomaly, i.e. $W_{OPA}(\omega) \sim (\hbar\omega - E_{FES})^{\beta_{FES}} \Theta(\hbar\omega - E_{FES})$, as shown in figure 10.7(c), where E_{FES} is an energy of the Fermi absorption edge.

The FES exponent consists of three parts as $\beta_{FES} = \beta_{TL} + \beta_{oc} + \beta_{ex}$ [54,55], where β_{TL} and β_{oc} are positive (corresponding to *suppression* of the transition probability) and given in equations (10.42) and (10.45), respectively, but β_{ex} is negative, indicating the *enhancement* of the transition probability, i.e.

$$\beta_{ex} = -\frac{1}{4} \sum_{\nu=\rho,\sigma} \frac{g_{2\nu}^h}{v_F + g_{4\nu}} (K_\nu + K_\nu^{-1})$$

Table 10.3. Comparison among the critical exponents of power-law singularities in several optical spectra of a $d = 1$ metal. Corresponding correlation functions are also given. Here $\hat{\psi}$ is the field operator of an electron, $\hat{\mathcal{H}}^e$ is the electronic Hamiltonian, and $\hat{\mathcal{H}}_{\text{int}}^h$ means the electron–core hole interaction.

Optical spectra	Critical exponents	Corresponding correlation functions
Valence-band photo-emission $P_{\text{PES}}(\epsilon)$ $(N, 0) \rightarrow (N - 1, 0)$	β_{TL}	Single-electron Green function $G^e(t) \propto \langle e^{it\hat{\mathcal{H}}^e/\hbar} \hat{\psi} e^{-it\hat{\mathcal{H}}^e/\hbar} \hat{\psi}^\dagger \rangle$ $\sim t^{-\beta_{\text{TL}} - 1}$
Core-level photo-emission $P_{\text{core}}(\epsilon)$ $(N, 0) \rightarrow (N, 1)$	$\beta_{\text{oc}} - 1$	Single-hole Green function $G^h(t) \propto \langle e^{it\hat{\mathcal{H}}^e/\hbar} e^{-it[\hat{\mathcal{H}}^e + \hat{\mathcal{H}}_{\text{int}}^h]/\hbar} \rangle$ $\sim t^{-\beta_{\text{oc}}}$
One-photon absorption $W_{\text{OPA}}(\omega)$ $(N, 0) \rightarrow (N + 1, 1)$	$\beta_{\text{TL}} + \beta_{\text{oc}} + \beta_{\text{ex}}$	Current–current correlation function $C(t) \propto \langle e^{it\hat{\mathcal{H}}^e/\hbar} \hat{\psi} e^{-it[\hat{\mathcal{H}}^e + \hat{\mathcal{H}}_{\text{int}}^h]/\hbar} \hat{\psi}^\dagger \rangle$ $\sim t^{-\beta_{\text{TL}} - \beta_{\text{oc}} - \beta_{\text{ex}} - 1}$

$$= -\frac{1}{2} \sum_{\nu=\rho,\sigma} \frac{g_{2\nu}^h}{v_F + g_{4\nu}} \left[1 - \left(\frac{g_{2\nu}}{v_F + g_{4\nu}} \right)^2 \right]^{-1/2} \leq 0, \quad (10.46)$$

which is an effect due to the Mahan exciton. Thus, the FES exponent in OPA spectra contains all information on the TL-liquid character ($\beta_{\text{TL}} \geq 0$), the orthogonality catastrophe ($0 \leq \beta_{\text{oc}} < 1$), and the many-body excitonic correlations ($\beta_{\text{ex}} \leq 0$). Therefore, the FES exponent can be either negative or positive depending on the balance between the positive parts $\beta_{\text{TL}} + \beta_{\text{oc}}$ and the negative part β_{ex} . The negative (positive) β_{FES} results in the power-law divergence (convergence) of the edge spectra. The electron–hole attraction ($g_{2\parallel}^h > 0$ and $g_{2\perp}^h > 0$) is necessary to obtain the divergent edge. Here we note that the FES exponent of the TL liquid is independent of the hole motion [54]. These results are summarized in table 10.3.

The FES has been observed clearly in several experiments using modulation-doped semiconductor wires [56]. Magnetic-field effects [57] and the long-range Coulomb interactions [58] were studied extensively, but nonlinear responses of the Mahan exciton remain to be investigated.

10.6 Conclusions and prospect

We have overviewed several exciton-related problems from theoretical viewpoints. In particular, we stress universal (material-independent) features of

low-dimensional exciton systems. One may notice the crucial importance of the dimensionality of materials in optical responses of meso- and microscopic systems. Geometrical confinement of excitons and the dielectric image-charge effects as well as the single-electron band structures are keys for designing novel materials with new optical functions. Moreover, many-body effects, electron-electron correlations, the Fermi-surface effects, and the quantum fluctuations also play important roles in the optical responses of low-dimensional materials.

Thus far, ‘natural’ low-dimensional materials, for example, organic semiconducting chains, organic synthetic metals, or layered compounds, are typical low dimensional materials. While recently, ‘artificial’ low-dimensional systems, for example, inorganic semiconductor compounds with controlled geometry and carrier doping, can be obtained due to rapid progress of crystal-growth and nanometre-scale fabrication techniques. Here we remark on the characteristics and differences between them.

- (i) *Natural (mainly organic) low-dimensional materials.* The low dimensionality is an inherent property of these materials, which are composed of low-dimensional molecular structures. Constituting elements of these materials are often soft and anisotropic molecules. Therefore, (quasi-)particles moves on a soft lattice with a large fluctuation, which results from low-dimensional structural instability. Conjugate π -electron polymers, delocalized σ -electron polymers, halogen-bridged mixed valence complex, and charge-transfer complex are typical examples. Such systems exhibit peculiar phenomena resulting from a strong electron (or exciton)–lattice interaction, which phenomena cannot be expected in a single-particle band picture. For example, the Peierls transition (metal–insulator transition) is a spontaneous lattice distortion coupled to the density wave formation. An electron or a hole in an excited state induces a local lattice distortion to relax its total energy, and then solitons, polarons, self-trapped excitons, or charge-transfer excitons can be formed, which are nonlinear elementary excitations peculiar to the $d \simeq 1$ electron-lattice systems.
- (ii) *Artificial (mainly inorganic) low-dimensional materials.* The low dimensionality is introduced ‘by hand’ into bulk semiconductor compounds with the use of the microfabrication technology. Constituting elements of these materials are often hard-core atoms, then (quasi-)particles moves in a rigid and distortionless perfect crystal. Band characteristics and fundamental optical properties can be controlled by changing the ratio of components of compounds and the carrier doping density. Several state-of-the-art technologies are available particularly in the III–V and IV groups. In these materials, single-electron band structures are well known. In turn, the band picture can describe well their characteristics from a viewpoint of a single-body problem. A particle-lattice coupling is often weak in such materials. Silicon in the IV group is intermediate between the above two categories. Recent development of chemical synthesis of Si polymers also enables us to

control the dimensionality of the material [10, 59].

All the exciton problems attracting interests recently are related to quantum many-body problems and to non-equilibrium dynamics far from the thermal equilibrium. Not only an electron, a hole, and the mutual Coulomb interaction but also their stage and some supporting players play interesting roles to yield novel properties. For example, the exciton-lattice interactions lead to the self-trapping phenomena [60], the randomness can induce the exciton weak localization [61], or the exciton-photon interaction results in the exciton squeezing. These effects and their dimensionality open new field in photophysics.

Although the theories introduced here do not always include all the actual details, we believe that these findings hold universally in electron-hole systems and will be of great significance in the interpretation of experimental results. In addition, this article will offer a guiding principle in the materials design of new low-dimensional materials. In the future, the following problems will be important in the study of the optical responses of low-dimensional exciton systems.

- (a) *Dimensional crossover.* In the study on universal features of low-dimensional systems, theoreticians usually assume an *ideal* dimension. However, the dimensional crossover between d and $d \pm 1$ is also important, which plays a crucial role in an optical anisotropy in the TPA process, as discussed in [16]. Interlayer or interchain couplings should be payed more attention. Actually, a quantum-dot lattice structure of semiconductors, a semiconductor wire array, and a multilayer semiconductor in a magnetic field are recently studied extensively in terms of their transport and nonlinear optical properties.
- (b) *Dimensionality of the photon field.* When we study the dimensionality of materials, the confinement of photon fields should be simultaneously considered. This radiation confinement is closely related to the cavity QED (quantum electrodynamics) for controlling the spontaneous emission. Similar to the control of the electron Bloch states by crystal periodicity, the photonic band structure can also be artificially controlled in dielectric structures. Moreover, photon localization problems are also concerned with the dimensionality of the photon field.
- (c) *Quantum fluctuations in material and photon fields.* In low-dimensional systems, their quantum fluctuations play crucial roles and then cannot be neglected. When the radiation field is quantized, quantum fluctuation of the photons is an interesting quantity as well as that of electrons, excitons, and lattices. However, the relation between fluctuations of photons and of materials has not been understood in detail. The transfer dynamics is also an open problem.

Acknowledgments

The author is grateful to the Korea Science and Engineering Foundation and the Japan Society for the Promotion of Science. A part of this work has been supported by CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (JST) and by a grant-in-aid for scientific research from the Ministry of Education, Science, Sports and Culture of Japan.

References

- [1] Ogawa T and Kanemitsu Y 1995 *Optical Properties of Low-Dimensional Materials* vol 1 (Singapore: World Scientific)
- Ogawa T and Kanemitsu Y 1998 *Optical Properties of Low-Dimensional Materials* vol 2 (Singapore: World Scientific)
- [2] Ogawa T 1997 *Nonlinear Opt.* **18** 181
- [3] This paper does not treat quasi-zero-dimensional ($d \simeq 0$) semiconductors, 'quantum dots'. See, e.g.,
Bányai L and Koch S W 1993 *Semiconductor Quantum Dots* (Singapore: World Scientific)
Jacak L, Hawrylak P and Wójs A 1998 *Quantum Dots* (Berlin: Springer)
- [4] Shinada M and Sugano S 1966 *J. Phys. Soc. Japan* **21** 1936
- [5] Shinozuka Y and Matsuura M 1983 *Phys. Rev. B* **28** 4878
Shinozuka Y and Matsuura M 1984 *Phys. Rev. B* **29** 3717(E)
- [6] Loudon R 1959 *Am. J. Phys.* **27** 649
Elliott R J and Loudon R 1959 *J. Phys. Chem. Solids* **8** 382
Elliott R J and Loudon R 1960 *J. Phys. Chem. Solids* **15** 196
Abe S 1989 *J. Phys. Soc. Japan* **58** 62
- [7] Ogawa T and Takagahara T 1991 *Phys. Rev. B* **43** 14 325
- [8] Bányai L, Galbraith I, Ell C and Haug H 1987 *Phys. Rev. B* **36** 6099
- [9] Ogawa T and Takagahara T 1991 *Phys. Rev. B* **44** 8138
Ogawa T and Takagahara T 1992 *Surf. Sci.* **263** 506
- [10] Ogawa T 1994 *Light Emission from Novel Silicon Materials* (Tokyo: Japanese Physics Society) p 120
- [11] Ohno T, Shiraishi K and Ogawa T 1992 *Phys. Rev. Lett.* **69** 2400
Kanemitsu Y, Ogawa T, Shiraishi K and Takeda K 1993 *Phys. Rev. B* **48** 4883
- [12] Ando H, Oohashi H and Kanbe H 1991 *J. Appl. Phys.* **70** 7024
- [13] Mahan G D 1968 *Phys. Rev.* **170** 825
- [14] Shimizu A 1989 *Phys. Rev. B* **40** 1403
- [15] Shimizu A, Ogawa T and Sakaki H 1992 *Phys. Rev. B* **45** 11 338
Shimizu A, Ogawa T and Sakaki H 1992 *Surf. Sci.* **263** 512
- [16] Ogawa T and Shimizu A 1993 *Phys. Rev. B* **48** 4910
- [17] Cingolani R, Lepore M, Tommasi R, Catalano I M, Lage H, Heitmann D, Ploog K, Shimizu A, Sakaki H and Ogawa T 1992 *Phys. Rev. Lett.* **69** 1276
Cingolani R, Lepore M, Tommasi R, Catalano I M, Lage H, Heitmann D, Ploog K, Shimizu A, Sakaki H and Ogawa T 1993 *Superlatt. Microstruct.* **13** 71
- [18] Rinaldi R *et al* 1994 *Phys. Rev. Lett.* **73** 2899

- [19] Keldysh L V 1979 *JETP Lett.* **29** 658
Keldysh L V 1988 *Superlatt. Microstruct.* **4** 637
- [20] Hanamura E, Nagaosa N, Kumagai M and Takagahara T 1988 *Mater. Sci. Eng.* B **1** 255
- [21] Kleinman D A 1983 *Phys. Rev. B* **28** 871
- [22] Akimoto O and Hanamura E 1972 *J. Phys. Soc. Japan* **33** 1537
- [23] Ishida K, Aoki H and Ogawa T 1995 *Phys. Rev. B* **52** 8980
- [24] Kuwata-Gonokami M *et al* 1994 *Nature* **367** 47
- [25] Stébé B, Munsch G, Stauffer L, Dujardin F and Murat J 1997 *Phys. Rev. B* **56** 12 454
Kheng K, Cox R T, Merle d'Aubigne Y, Bassani F, Saminadayar K and Tatarenko S 1993 *Phys. Rev. Lett.* **71** 1752
- [26] Nasu K (ed) 1997 *Relaxations of Excited States and Photo-Induced Phase Transitions* (Berlin: Springer)
- [27] Koshino K and Ogawa T 1998 *Phys. Rev. B* **58** 14 804
Koshino K and Ogawa T 1998 *J. Phys. Soc. Japan* **67** 2174
- [28] Östreich Th, Schönhammer K and Sham L J 1995 *Phys. Rev. Lett.* **74** 4698
Combescot M and Combescot R 1988 *Phys. Rev. Lett.* **61** 117
- [29] Haug H and Koch S W 1994 *Quantum Theory of the Optical and Electronic Properties of Semiconductors* 3rd edn (Singapore: World Scientific)
- [30] Lindberg M and Koch S W 1988 *Phys. Rev. B* **38** 3342
- [31] Ivanov A L, Haug H and Keldysh L V 1998 *Phys. Rep.* **296** 237
- [32] Axt V M, Bartels G and Stahl A 1996 *Phys. Rev. Lett.* **76** 2543
Lindberg M, Hu Y Z, Binder R and Koch S W 1994 *Phys. Rev. B* **50** 18 060
- [33] Axt V M and Stahl A 1994 *Z. Phys. B* **93** 195
Axt V M and Mukamel S 1998 *Rev. Mod. Phys.* **70** 145
Pereira M F and Henneberger K 1998 *Phys. Rev. B* **58** 2064
- [34] Kuwata-Gonokami M, Inouye S, Suzuura H, Shirane M, Shimano R, Someya T and Sakaki H 1997 *Phys. Rev. Lett.* **79** 1341
- [35] Hanamura E and Haug H 1977 *Phys. Rep.* **33** 209
- [36] Hiroshima T 1989 *Phys. Rev. B* **40** 3862
- [37] Inoue J, Brandes T and Shimizu A 1998 *J. Phys. Soc. Japan* **67** 3384
- [38] Usui T 1960 *Prog. Theor. Phys.* **23** 787
- [39] Hawton M and Nelson D 1998 *Phys. Rev. B* **57** 4000
- [40] Marumori T, Yamamura M and Tokunaga A 1964 *Prog. Theor. Phys.* **31** 1009
- [41] Kishimoto T and Tamura T 1983 *Phys. Rev. C* **27** 341
- [42] Iachello F and Arima A 1987 *The Interacting Boson Model* (Cambridge: Cambridge University Press)
Talmi I 1993 *Simple Models of Complex Nuclei* (Harwood Academic)
- [43] Kopietz P 1997 *Bosonization of Interacting Fermions in Arbitrary Dimensions* (Berlin: Springer)
- [44] Gogolin A O, Nersesyan A A and Tsvelik A M 1998 *Bosonization and Strongly Correlated Systems* (Cambridge: Cambridge University Press)
- [45] Mahan G D 1990 *Many-Particle Physics* (New York: Plenum)
- [46] Tomonaga S 1950 *Prog. Theor. Phys.* **5** 544
Luttinger J M 1963 *J. Math. Phys.* **4** 1154
Haldane F D M 1980 *Phys. Rev. Lett.* **45** 1358
Solyom J 1979 *Adv. Phys.* **28** 201
- [47] Nozieres P and Schmitt-Rink S 1985 *J. Low. Temp. Phys.* **59** 195

- [48] Ivanov A L and Haug H 1993 *Phys. Rev. Lett.* **71** 3182
- [49] Tanguy C and Combescot M 1992 *Phys. Rev. Lett.* **68** 1935
Foing J-P, Hulin D, Joffre M, Jackson M K, Oudar J-L, Tanguy C and Combescot M 1992 *Phys. Rev. Lett.* **68** 110
- [50] Nagaosa N and Ogawa T 1993 *Solid State Commun.* **88** 295
- [51] Ogawa T 1995 *Phys. Status Solidi b* **188** 83
- [52] Kondo J and Yoshimori A (ed) 1988 *Fermi Surface Effects* (Berlin: Springer)
- [53] Ohtaka K and Tanabe Y 1990 *Rev. Mod. Phys.* **62** 929
- [54] Ogawa T, Furusaki A and Nagaosa N 1992 *Phys. Rev. Lett.* **68** 3638
- [55] Ogawa T 1997 *Mater. Sci. Eng. B* **48** 131
Ogawa T 1998 *Physica B* **249** 185
- [56] Calleja J M, Goni A R, Dennis B S, Weiner J S, Pinczuk A, Schmitt-Rink S, Pfeiffer L N, West K W, Muller J F and Ruckenstein A E 1991 *Solid State Commun.* **79** 911
For a $d \simeq 2$ system, see Skolnick M S, Rorison J M, Nash K J, Mowbray D J, Tapster P R, Bass S J and Pitt A D 1987 *Phys. Rev. Lett.* **58** 2130
- [57] Ogawa T and Otani H 1995 *J. Phys. Soc. Japan* **64** 3664
Otani H. and Ogawa T 1996 *Phys. Rev. B* **53** 4684
- [58] Otani H. and Ogawa T 1996 *Phys. Rev. B* **54** 4540
- [59] Kanemitsu Y. and Ogawa T 1996 *Surf. Sci. Lett.* **3** 1163
- [60] Sumi H. and Sumi A 1994 *J. Phys. Soc. Japan* **63** 637
- [61] Hanamura E 1996 *Phys. Rev. B* **54** 11219

Chapter 11

Size quantization and electron dynamics in nanometre-size semiconductors

Takenari Goto

Department of Physics, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan

Size quantization of carriers in ultrathin microcrystals of PbI_2 is stated over the wide range of thicknesses. We discuss the exciton confinement, the electron–hole confinement and breakdown of the effective mass approximation. In addition, we show shrinkage of the exciton wavefunction of three and four monolayer crystals.

Second, the temporal behaviour nonlinear optical properties and luminescences of CuCl nanocrystals are measured, and the lifetimes of the electron–hole pair, the exciton and the bi-exciton are estimated to be 1.6 ps, 65 ps and 380 ps, respectively.

Third, we show spectral diffusion and on-off behaviour in a luminescence spectrum of a single CdSe quantum dot on a CdSe quantum well and the quicker and smaller spectral changes in comparison with a single nanocrystal surrounded by organic material are found in the present quantum dot. These facts suggest that the electron or a hole can migrate more easily in the CdSe quantum well and a lifetime of trapped carriers is shorter.

11.1 Introduction

There are three types of nanometre-size semiconductors having zero-, one- and two-dimensional structures. Electrons, holes and/or excitons are confined in these semiconductors resulting in size quantization of their momenta and energies. In many semiconductors, these size quantizations have been studied. In a weak confinement regime, a translational motion of an exciton is size quantized. One typical example is a CuCl dot where the nanocrystal size is larger than the exciton Bohr radius [1]. In a strong confinement regime, motions of electrons and holes

are individually quantized. The typical examples are seen in CdSe dots [1] and GaAs wells [2], where the spatial size is smaller than the exciton Bohr radius [2]. On the other hand, in the very small structures of atomic size, the effective mass approximation breaks down and the electronic band structure may be changed. Hence, it is important to show an example in the whole range of the crystal size. There is a typical example in the quantum well where the well width is varied from the atomic distance to the range much longer than the exciton Bohr radius, in the layer type semiconductor PbI₂. As the crystal structure is of layer type, an ultrathin crystal can be grown. In section 11.2, the size confinements are discussed in PbI₂ ultrathin nanocrystals embedded in polymer, and in section 11.3, the size confinement of the internal motion of the exciton is also discussed.

Next, optical nonlinearity is very important in the semiconductor nanocrystal quantum dots, because it becomes very large compared to the bulk. Such a large optical nonlinearity is useful for an application to fast response memory materials. Here, we report an electronic mechanism of the nonlinear optics in a CuCl dot. The dynamics of this mechanism is a very exciting problem, which is studied by pump-probe and luminescence measurements using a very strong light from a subpicosecond pulse laser.

Third, photoinduced phenomena have been extensively studied by many investigators. Namely, there are a persistent hole burning [3], a photodarkening [4] and a photobrightening [5] in nanocrystals with different sizes, and a spectral diffusion and an on-off behaviour in a single nanocrystal [6]. These phenomena are believed to be associated with Auger ionization of a photogenerated electron (hole) [7] or a chemical reaction near the boundary of the nanocrystal [8]. The mechanism of the photoinduced phenomena, however, has not yet been clarified.

Here, we introduce recent results on a spectral diffusion and an intensity fluctuation of luminescence in CdSe quantum dots grown on the CdSe well.

11.2 Two-dimensional size confinement in ultrathin PbI₂ nanocrystals [9]

A PbI₂ crystal has a layer type structure and the layer thickness is 0.7 nm. When the PbI₂ nanocrystal is grown in ethylene-methacrylic acid copolymer, its shape is platelet-like and the layer number of the nanocrystal is more than two. The lateral size is 6 nm or more in the two monolayer crystals, and is about 100 nm in the nine monolayer crystals from TEM observation.

In figure 11.1, is shown one example of TEM photographs of the polymer film containing PbI₂ nanocrystals with two, three and four monolayers. The polymer film is 100 nm thick and the TEM photograph was taken in the polymer film cooled at 100 K to avoid the temperature rise by the electron beam. The c-axis is parallel to the photograph. Many crystals having different numbers of monolayers are embedded in the polymer film.

In the bulk crystal, the bandgap locates at an A point of the Brillouin zone,

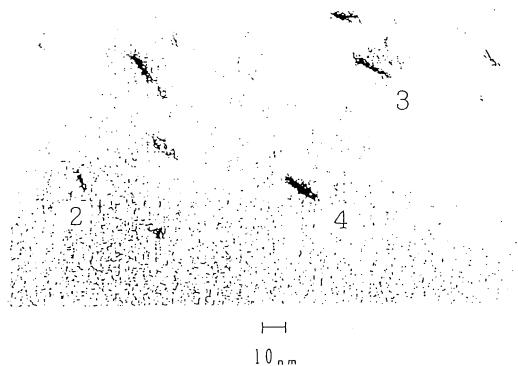


Figure 11.1. A photograph of PbI_2 ultrathin nanocrystals taken by a transmission electron microscope. The c -axis locates in the photograph. A number indicates a layer number of the ultrathin nanocrystal.

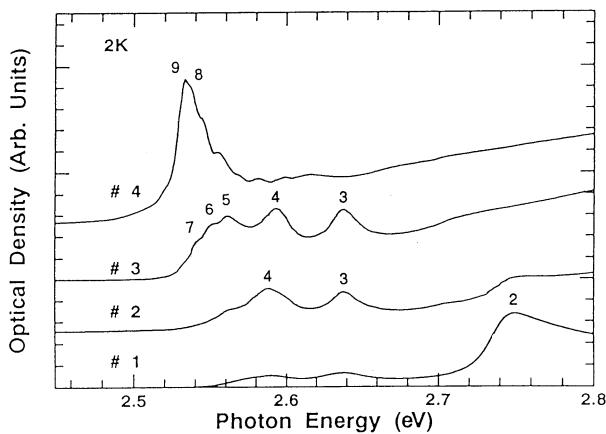


Figure 11.2. Absorption spectra of four polymer films containing PbI_2 ultrathin nanocrystals with different layer numbers at 2 K. A number in this figure shows a layer number of the nanocrystal corresponding to the peak.

where the conduction electron orbit with the lowest energy has a p-like symmetry, A_4^- and the valence band orbit with the highest energy has a s-like symmetry, A_4^+ , according to Schlüter and Schlüter's paper [10]. The lowest energy exciton is called a cationic exciton, and consists of the A_4^- electron and the A_4^+ hole. A dipole transition of the exciton is allowed. The energy of the exciton absorption band is around 2.5 eV depending on the polytype.

Figure 11.2 shows exciton absorption spectra of four samples at 2 K. The average thickness of PbI_2 nanocrystals embedded in the polymer film becomes

larger as the sample number increases. If the thickness is uniform, a single band should be expected to appear near the energy region of the A exciton of the bulk which has the lowest energy. In fact, as the polymer contains ultrathin nanocrystals with different numbers of monolayers, we find some exciton bands in the spectra. It is because the exciton energy depends on the thickness as well as the lateral size of the nanocrystal owing to the size quantization of the exciton energy. The lateral diameter, however, is larger than 6 nm, so that the energy difference of the exciton due to the size distribution is not so large. Hence, the distributed lateral sizes are contributed to the band width.

It is essential to know the thickness of the nanocrystal corresponding to each exciton band for analysis of a quantum size effect on the exciton energy. To determine the thickness, the resonant Raman scattering was measured where the exciting light energy coincides with the exciton band energy. We observed a rigid layer mode of the platelet-like nanocrystal with finite thickness. The energy depends on the thickness, but not on the lateral size. Comparing the measured Raman shift with the theoretical value, we determined the layer number of the nanocrystal, i.e. the thickness. A number in this figure shows the layer number of the nanocrystal corresponding to the exciton band. For example, two monolayer nanocrystals are seen to be mainly contained in the polymer of sample 1, and three to seven monolayer nanocrystals are seen to be contained in sample 3, from the absorption spectra of [figure 11.2](#). Consequently, an exact correspondence between the exciton formation energy and the thickness of the ultrathin nanocrystal is obtained.

In [figure 11.3](#), the exciton energy is shown by an open circle as a function of $1/L^2$, where L is a thickness of the ultrathin nanocrystal. The upper abscissa represents the layer number. As seen in figures 11.2 and 11.3, the exciton band energy becomes larger with decreasing the thickness. This effect is considered to be due to spatial confinement of the exciton in the c -direction of the nanocrystal. In the simple model where the exciton translational motion is confined in the one-dimensional quantum well with an infinite potential, the exciton energy E in an effective mass approximation is expressed by

$$E = E_B + \frac{\hbar^2}{2M} \frac{\pi^2 p^2}{L^2}, \quad p = 1, 2, 3, \dots, \quad (11.1)$$

where E_B is an exciton formation energy in the bulk, M is exciton total mass, L a width of the potential well and p a quantum number of the confinement. Equation (11.1) with the exciton translational mass of the bulk, $M = 1.0m_0$ [11] and with an adjustable parameter, $E_B = 2.524$ eV is represented by a full straight curve in [figure 11.3](#) for the lowest energy exciton, $p = 1$, on the assumption that L equals the thickness of the nanocrystal. The value of E_B is close to that of the lowest energy exciton in the bulk, which depends a little on the polytype.

This relation corresponds to the weak confinement limit as stated before. In the strong confinement regime, the electron and the hole are individually confined, and hence, the mass M in equation (11.1) should be the reduced mass of the

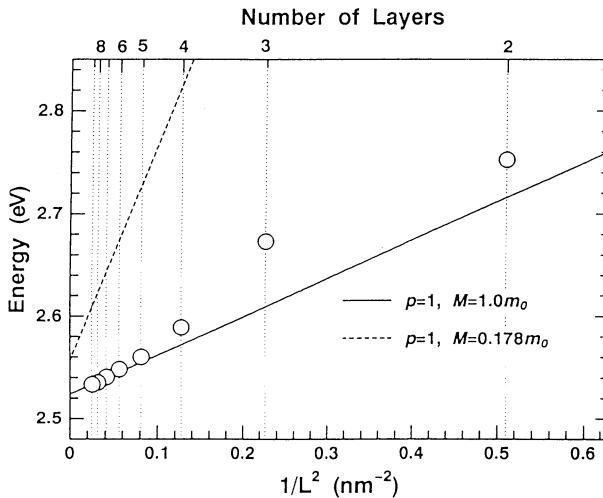


Figure 11.3. Exciton band energies as a function of $1/L^2$ are shown by open circles. L is a thickness of the nanocrystal. The upper abscissa represents the layer number. A full straight curve shows a relation expressed by equation (11.1), where $M = 1.0m_0$, $p = 1$ and $E_B = 2.524$ eV. A broken curve shows a relation expressed by equation (11.1), where $M = 0.18m_0$, $p = 1$ and $E_B = 2.56$ eV.

electron and the hole. The reduced mass is obtained as $0.18m_0$ by Skölnick and Bimberg [12], where m_0 is a free electron mass. In figure 11.3 a broken line represents equation (11.1) for the lowest energy exciton, $p = 1$, with this mass value and the bandgap energy 2.56 eV as E_B . The exciton band energies of ultrathin nanocrystals with more than five monolayers coincide with the full straight curve. Hence, it is concluded that the translational motion of the exciton is quantized in these thicker nanocrystals.

As the exciton Bohr radius, 2 nm [13], corresponds to the thickness of three monolayer nanocrystals the internal motion of the exciton is considered to be kept unchanged in the nanocrystal with more than five monolayers. The observed exciton energies of two, three and four monolayer nanocrystals deviate from the full straight curve to the higher energies. This deviation may be caused by the change from exciton confinement to electron–hole confinement. The exciton energy of the two monolayer crystal approaches the straight curve, again. This phenomenon is against a size confinement theory based on an effective mass approximation. The contradiction means that not only a breakdown of the effective mass approximation but also a large change of the electronic band structure occurs in the two monolayer nanocrystals because of the too thin layer crystal.

11.3 Size confinement of an exciton internal motion in ultrathin PbI₂ nanocrystals [14]

To study an exciton internal motion in the ultrathin nanocrystal we measured a diamagnetic shift of the exciton absorption band energy in ultrahigh magnetic fields. The sample was set at the centre of one turn coil and cooled by helium gas. A pulsed magnetic field up to 150 T was applied to the sample. A pulsed white light from a xenon lamp was transmitted through the sample, analysed by a monochromator and detected by a streak camera. Figures 11.4 and 11.5 show the exciton absorption spectra of the polymer film containing three and four monolayer nanocrystals for σ^+ and σ^- circular polarizations, respectively, in the Faraday configuration at 15 K and 11 K, respectively. The magnetic field for each curve is shown on the left-hand side and a number on the lowest curve without the magnetic field shows a layer number of the nanocrystal corresponding to the exciton band. Structures below 2.56 eV appear owing to insufficient cancellation, because the spectrum of the xenon flash lamp used as a light source is too complicated.

Here we discuss the diamagnetic shift of the PbI₂ nanocrystal. The diamagnetic shift $\Delta E_{d\parallel}$ for the magnetic field B_z parallel to the c -axis is expressed by

$$\Delta E_{d\parallel} = \sigma B_z^2. \quad (11.2)$$

In the configuration $B \perp c$, however, the exciton state is not an eigenstate for circularly polarized light. In fact, the apparent diamagnetic shift of the bulk is less than half [13]. Taking account of the apparent shift of the bulk, the diamagnetic shift for B perpendicular to the c -axis $\Delta E_{d\perp}$ may be written as

$$\Delta E_{d\perp} = p\sigma(B_x^2 + B_y^2), \quad (11.3)$$

where $p = 0$ and 0.25 for σ^+ and σ^- polarizations, respectively.

The Zeeman splitting ΔE_z is given by

$$\Delta E_z = \pm \frac{1}{2}g\mu_B B, \quad (11.4)$$

where signs + and - correspond to σ^+ and σ^- polarizations, respectively. From an analogy with the bulk it is assumed that ΔE_z is isotropic. On the other hand, the oscillator strength of the lowest energy exciton for $E \parallel c$ polarization is known to be one-fourth of that for $E \perp c$ [10]. It is determined by the transmission electron microscopy that the average lateral size of the microcrystallites is about ten times as large as the thickness. Under the assumption that the absorption band shape is Lorentzian and the c -axis of the microcrystal are randomly oriented, the absorption spectra in the magnetic fields of zero and 150 T for σ^+ and σ^- polarizations are calculated and shown by dotted curves in figure 11.6.

Adjusting the calculated dotted curves to the experimental full curves, we obtain the diamagnetic shift for the three and four monolayer nanocrystals as

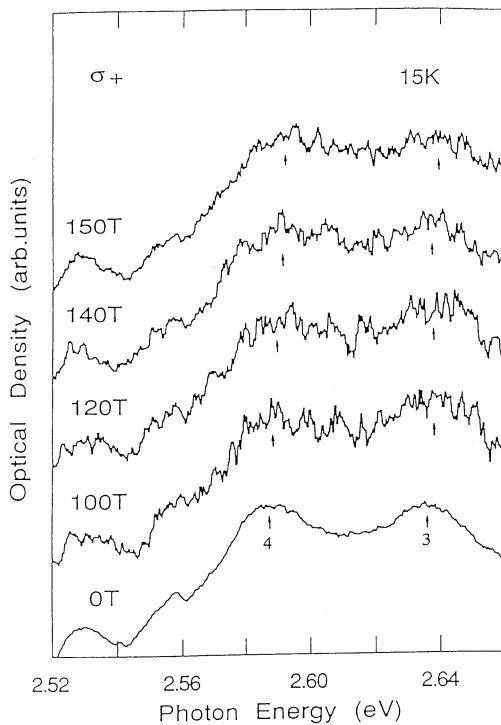


Figure 11.4. Exciton absorption spectra in the different magnetic fields for σ^+ polarization at 15 K. A number in the lowest spectrum without magnetic field indicates a layer number corresponding to the exciton band.

$\sigma = 3 \times 10^{-7} \text{ eV T}^{-2}$ which is smaller compared to diamagnetic coefficients 6.1 and $7.0 \times 10^{-7} \text{ eV T}^{-2}$ for three and four monolayer nanocrystals respectively, calculated by using formulae of Basterd *et al* [15] on the basis of an effective mass approximation. The smaller value may originate from a dielectric screening effect [16], because the PbI₂ microcrystal is embedded in polymer with the smaller dielectric constant. Namely, the exciton wavefunction shrinks as the Coulomb interaction between electron and hole becomes stronger outside the crystal. From the diamagnetic constant, the exciton binding energy is estimated to be 200 meV which is 3.3 times as large as the bulk value.

11.4 Nonlinear optical processes in CuCl nanocrystals [17]

Optical nonlinearity of quantum dots, i.e. nanocrystals, has interested us very much, because the nonlinearity has been found to be much larger than that of the bulk materials [18]. These studies of optical nonlinearity and dynamics are

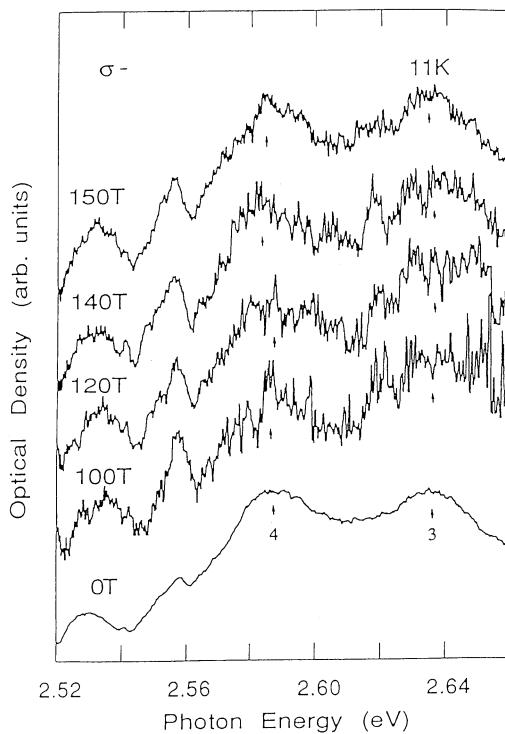


Figure 11.5. Exciton absorption spectra in the different magnetic fields for σ^- polarization at 11 K.

useful for the development of new materials and applications of ultrafast optical devices. The dynamics of carriers confined in CuCl nanocrystals has been studied by an ultrafast pulse laser. Here we introduce temporal dependences of the absorption change due to a strong pulsed laser light and of the luminescence in CuCl nanocrystals embedded in a NaCl crystal at 77 K. A CuCl nanocrystal is a typical material for zero-dimensional quantum confinement of the exciton translational motion, because the Bohr radius 0.7 nm is small compared to the radius of the prepared nanocrystals. The lowest energy exciton of a CuCl crystal with the zinc blende-structure is dipole allowed and consists of a Γ_6 conduction electron and a Γ_7 valence hole which is called a Z_3 exciton. As the Γ_6 and Γ_7 orbitals are orbitally singlet, the excitonic nature is very simple. The second lowest energy exciton, $Z_{1,2}$, consists of a Γ_6 conduction electron and a spin-orbit split-off Γ_8 valence hole, and is also dipole allowed. We used a CuCl nanocrystal embedded in a NaCl single crystal and the exciton kinetic energy ΔE is assumed to be quantized in a spherically shaped nanocrystal with an infinite potential well

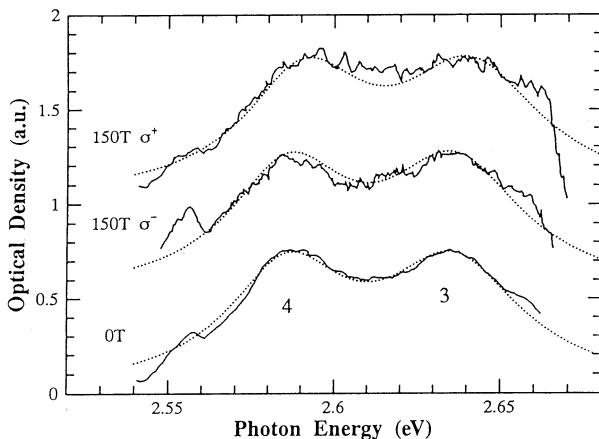


Figure 11.6. Upper two full curves show the absorption spectra in the magnetic field of 150 T for σ^+ and σ^- polarizations. The lowest curve is the absorption spectrum without magnetic field. A dotted curve shows the calculated absorption spectrum on the assumption that $\sigma = 3 \times 10^{-7} \text{ eV T}^{-2}$.

as

$$\Delta E = \frac{\hbar^2}{2M} \cdot \frac{\pi^2}{a^{*2}}, \quad a^* = a - 0.5a_B, \quad (11.5)$$

where a is the real nanocrystal radius, a_B the exciton Bohr radius and a^* the effective radius [19]. The exciton is believed to be unstable at the surface layer of the sphere which is called a dead layer, and hence, the effective radius a^* is smaller than the real radius a .

Figure 11.7 shows absorption spectra with (broken curve) and without (full curve) irradiation of subpicosecond pulsed laser light with the energy corresponding to the band-to-band transition region. The pump light has a wavelength of 307 nm, a repetition rate of 1 kHz, a pulse width of ~ 200 fs and an intensity of 1 mJ cm^{-2} at the surface of the sample. The luminescence spectrum in the excitation of the same light as the pump is shown by a dash-dotted curve. These spectra were measured at 77 K in a nanocrystal with the average radius of 4 nm. The crystal radius is distributed around 4 nm.

A luminescence band at 3.23 eV appears at the energy position of the Z_3 absorption band and is associated with radiative annihilation of the Z_3 exciton and named E_x . A luminescence band at 3.17 eV originates from annihilation of the bi-exciton leaving the Z_3 exciton.

The Z_3 and $Z_{1,2}$ exciton absorption bands are blue-shifted and broadened when the sample is excited by the strong pulsed light. At the M band region, a negative absorption band appears, which means existence of an optical gain. The appearance of the M band luminescence suggests that more than two excitons are

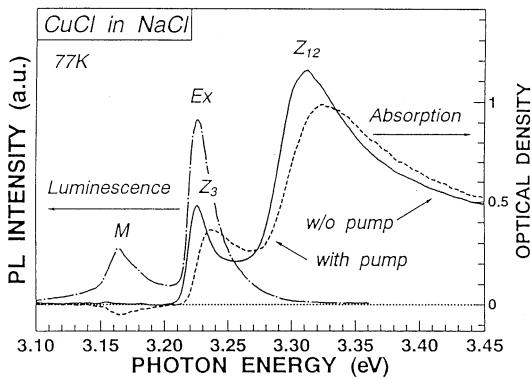


Figure 11.7. A luminescence spectrum and absorption spectra of CuCl nanocrystals embedded in a NaCl single crystal. Full and broken curves show absorption spectra without excitation by a subpicosecond laser light and 10 ps after the excitation, respectively, at 77 K. Absorption bands at 3.225 eV and 3.31 eV are associated with Z_3 and $Z_{1,2}$ excitons, respectively.

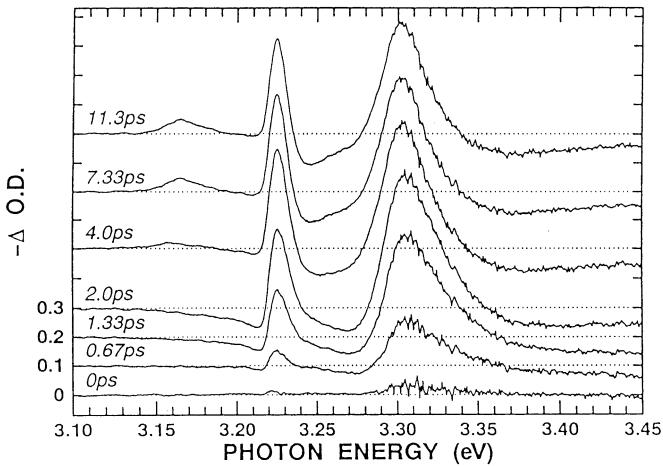


Figure 11.8. Changes of the absorption spectra at various delay times within 11.3 ps after the excitation.

photogenerated in a CuCl nanocrystal.

Figures 11.8 and 11.9 show changes of the absorption spectra at different delay times after the strong pulsed light excitation. The delay time of the probe light pulse from the pump light pulse is shown on the left-hand side. The ordinate represents the negative change in the optical density by the pump light irradiation. To analyse the change of the exciton absorption bands in the nanocrystals with

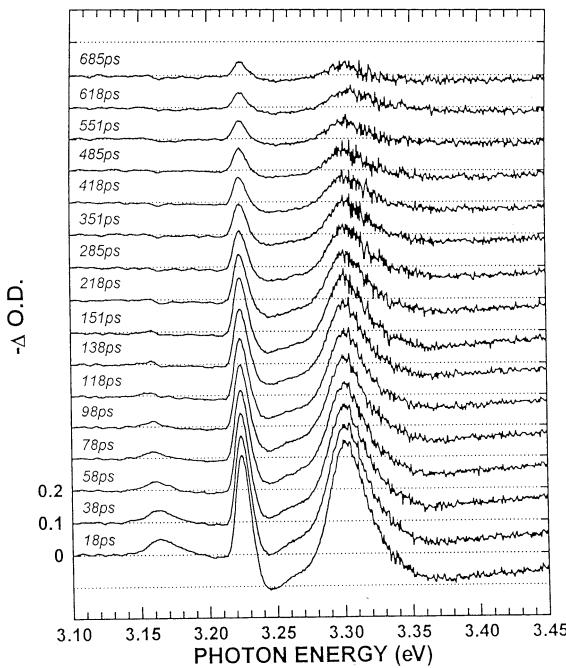


Figure 11.9. Changes of the absorption spectra at various delay times between 18 ps and 685 ps.

the different sizes, a moment analysis is useful [19]. The first order moment $M_1 = \int EI(E) dE/M_0$, where E is the photon energy and $M_0 = \int EI(E) dE$ and $I(E)$ is the intensity, is calculated and shown in figure 11.10 for the Z_3 (triangles) and $Z_{1,2}$ (squares) bands as a function of delay time.

Figure 11.11 shows the gain ($\Delta O.D. < 0$) and the induced absorption ($\Delta O.D. > 0$) at 3.165 eV in the M band region by open circles as a function of delay time. The negative change in the optical density, i.e. the optical gain, is observed in the delay time less than 100 ps, while the positive change, i.e. the induced absorption, is observed in the delay time more than 100 ps. In figure 11.12, temporal responses of the E_x and M luminescence intensities are shown by a full curve. The M luminescence intensity decays faster than the E_x , in contrast, the E_x luminescence intensity scarcely increases and then slowly decays.

Here we assume that existence of two excitons in a nanocrystal causes a peak shift twice as large as that of one exciton. By taking the Poisson's distribution into account, the average number of excitons per nanocrystal is estimated to be two. The gain at the M band region is caused by an induced transition between the exciton state and the bi-exciton state.

Dotted curves in figures 11.10, 11.12 and a full curve in figure 11.11,

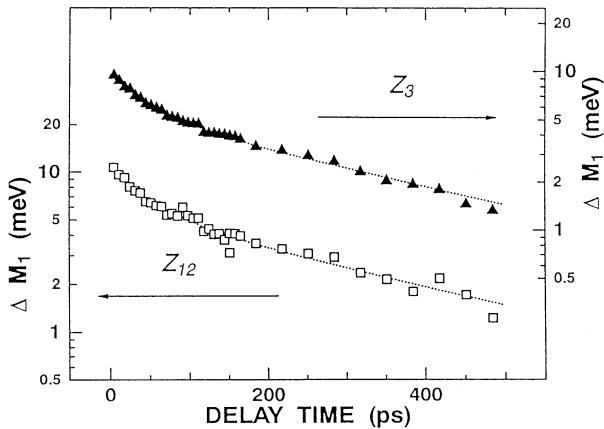


Figure 11.10. Delay time dependences of the first moment change ΔM_1 of the Z_3 (triangle) and $Z_{1,2}$ (squares) absorption bands in the time range greater than 10 ps. Dotted curves show calculated curves on the basis of a cascade relaxation model.

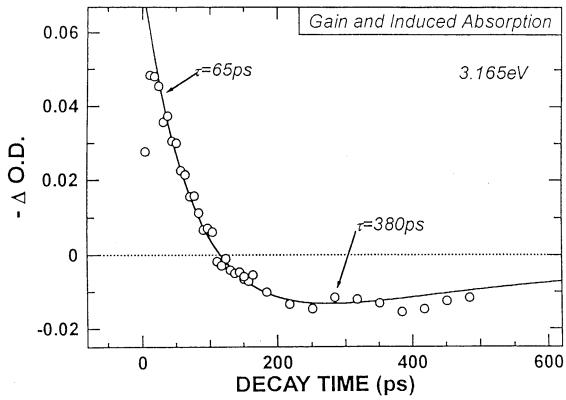


Figure 11.11. An absorption change at 3.165 eV in the M luminescence band region as a function of the delay time in the time range greater than 10 ps. A full curve shows a calculated curve on the basis of the cascade relaxation model.

represent calculated curves on the basis of an effective cascade relaxation model. From this analysis, the lifetimes of the exciton and the bi-exciton are obtained as 380 ps and 65 ps, respectively.

Next, we analyse time responses of the first moment, of the gain or induced absorption and of the luminescence in the picosecond time range. The time response of the luminescence was measured by an up-conversion method. Delay time dependences of the first moment change ΔM_1 in the Z_3 and $Z_{1,2}$ exciton

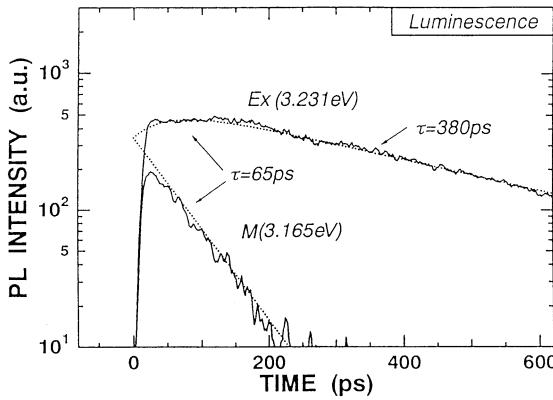


Figure 11.12. Time dependences of the E_x and M luminescence intensities. Dotted curves show calculated lines.

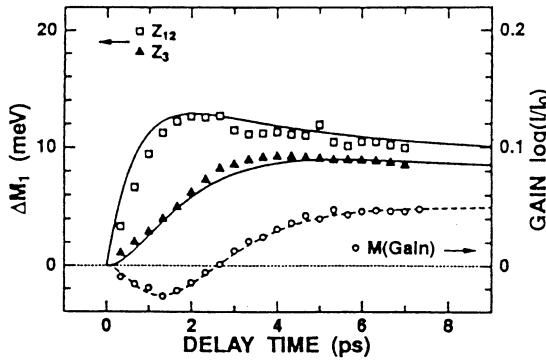


Figure 11.13. Temporal change of ΔM_1 in the Z_3 (triangles) and $Z_{1,2}$ (squares) exciton bands, and the gain (circles) at 3.165 eV.

bands are shown by triangles and squares, respectively, in figure 11.13. The gain at 3.165 eV in the M band luminescence region is shown by circles. The negative gain seen below 3 ps means the induced absorption. Figure 11.14 shows the temporal behaviour of the luminescence intensity at 77 K. Open and closed circles represent the exciton (E_x) and bi-exciton (M) luminescences, respectively. The E_x luminescence intensity is multiplied by five. The E_x luminescence intensity rises faster compared to the M luminescence. Full and broken curves in figure 11.13 and full curves in figure 11.14 represent calculated curves from rate equations on the assumption that a photogenerated electron–hole pair in a nanocrystal relaxes to an exciton with a rate k_1 and that two photogenerated electron–hole pairs relaxes to an exciton and an electron–hole pair with a rate $2k_1$. We also assume that the two excitons are combined to the bi-exciton immediately.

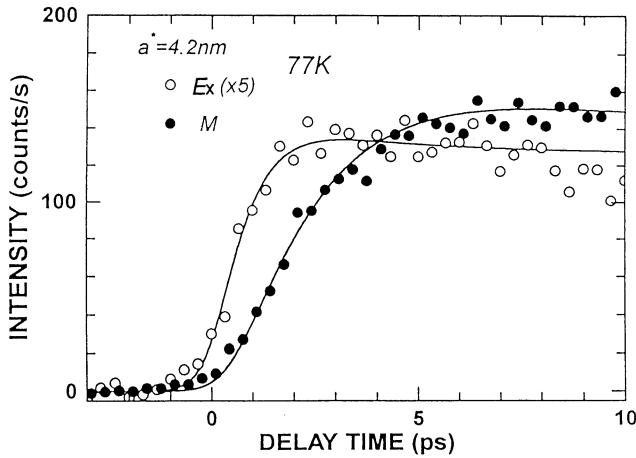


Figure 11.14. Intensities of the Ex (open circles) and M (closed circles) luminescences are shown as a function of delay time. Full curves represent the calculated lines.

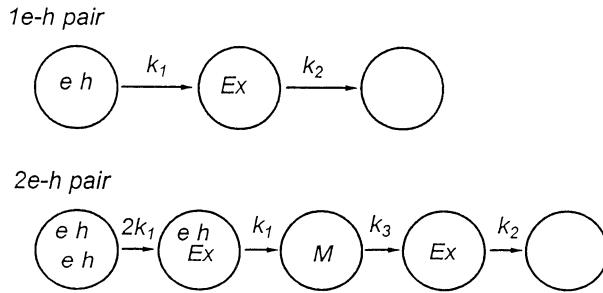


Figure 11.15. Schematic picture showing the cascade relaxation process of the photogenerated electron–hole pair. Upper and lower parts show the processes when one and two electron–hole pairs, respectively, are generated in a CuCl nanocrystal.

From the best fitting to the experimental data of the first moment, the gain and the luminescence intensity, the rate k_1 is estimated to be $1/1.6 \text{ ps}^{-1}$. Namely, the relaxation time of the electron–hole pair to the exciton is 1.6 ps. Dynamical processes of photogenerated carriers in a CuCl nanocrystal for two cases of one and two electron–hole pair formations are summarized in figure 11.15. A circle shows a nanocrystal and a symbol $e\ h$ means an electron–hole pair. If one electron–hole pair is generated in a nanocrystal by the pump light irradiation (upper part), the electron–hole pair recombines to an exciton with a rate k_1 . The exciton is radiatively annihilated with a slow rate k_2 . If two electron–hole pairs are generated in a nanocrystal (lower part), a bi-exciton is formed through one electron–hole pair and one exciton from two electron–hole pairs. The bi-

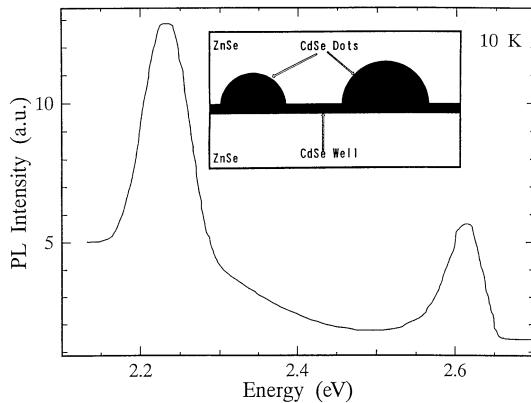


Figure 11.16. A luminescence spectrum of CdSe quantum dots in ZnSe under the excitation of a 325 nm light at 10 K. The insert is a schematic picture of the sample.

exciton annihilates radiatively leaving an exciton with rate k_3 . Finally, the exciton recombines with a rate k_2 . Relaxation rates k_1 , k_2 and k_3 are estimated to be 1.6 ps^{-1} , 380 ps^{-1} and 65 ps^{-1} , respectively.

11.5 Photoinduced phenomena in luminescence spectra of a single CdSe nanocrystal [20]

Bawendi's group [21] has found a spectral diffusion and an on-off behaviour of the luminescence from a single CdSe nanocrystal synthesized as colloids for the first time and interpreted these phenomena as due to a Stark effect originating from a trapped electron outside the nanocrystal generated by a photoionization.

Our group has also found the similar phenomena in a CdSe quantum dot on a CdSe quantum well prepared by self-organized growth. Figure 11.16 shows a luminescence spectrum in the excitation by 325 nm light from a He-Cd laser at 10 K. The inset is a schematic picture of the sample. The sample was grown by molecular beam epitaxy. A 120 nm thick buffer layer of ZnSe was grown at the (100) surface of GaAs at 300 °C. The CdSe quantum dots were formed after deposition of about three monolayers of CdSe in average under Cd rich conditions at 300 °C, and capped by a 50 nm thick ZnSe layer. To obtain a higher uniformity of dots, two quantum dot sheets were stacked with a ZnSe spacer of 12 nm and a cap layer of 50 nm. The total thickness of the structures was kept below 200 nm to avoid relaxation of the barrier material.

The luminescence band at about 2.6 eV originates from recombination of the size quantized electron and hole in the two monolayers of CdSe, and that at about 2.23 eV from the CdSe dots.

In order to directly generate an electron-hole pair in the two-monolayer

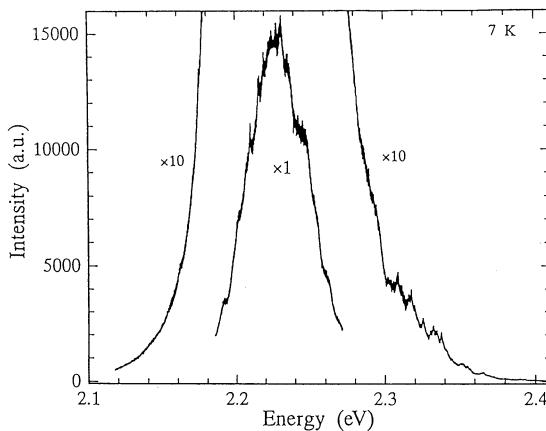


Figure 11.17. A microscopic luminescence spectrum with one second excitation time at 7 K, when the surface area of $1 \mu\text{m}^2$ on the sample is excited by a 476.5 nm light.

CdSe well, the sample was excited by a 476.5 nm (2.602 eV) light from a cw Ar⁺ laser. The excitation energy in the CdSe quantum well transfers to the quantum dot with the lower electron–hole pair energy and then annihilates radiatively. The laser beam was focused at the sample surface by a microscope objective lens. The laser light spot diameter was about 1 μm and the power density was about 1 kW cm⁻² at the sample surface. The sample was cooled up to 7 K by conduction method using an Oxford microscope cryostat. The luminescence was collected through the same microscope objective lens and analysed by a JY Ramanor 64000 spectrometer with the spectral resolution of 0.1 meV.

Figure 11.17 shows a luminescence spectrum with one second integration time at 7 K. The luminescence spectrum consists of many sharp and small peaks. The spectrum on the high-energy tail of the overall luminescence band in figure 11.17 is shown in [figure 11.18](#) on an expanded energy scale. The luminescence spectrum was measured nine times, and the time interval between two consecutive spectra was 15 s. A number in the figure indicates an order of the measurement. In the experiment, we found that the energy of the sharp peaks changed with a jump when the position of the sample was slowly moved. That is to say, the peak A disappeared suddenly, the peak B or C became stronger or disappeared or another peak appeared suddenly. Thus, the sharp peaks are position sensitive. When the size of the laser light spot was large, the sharp peaks disappeared. These facts strongly suggest that the observed sharp peaks originate from single CdSe quantum dots. As a spot area of the laser light at the sample surface is about $1 \mu\text{m}^2$, many dots with different sizes are excited, and hence, there appear many sharp lines near the peak of the overall band. For the quantum size effect on the electronic transition energy, the transition energy in the quantum dot much smaller than the average one is expected to be largely separated to the

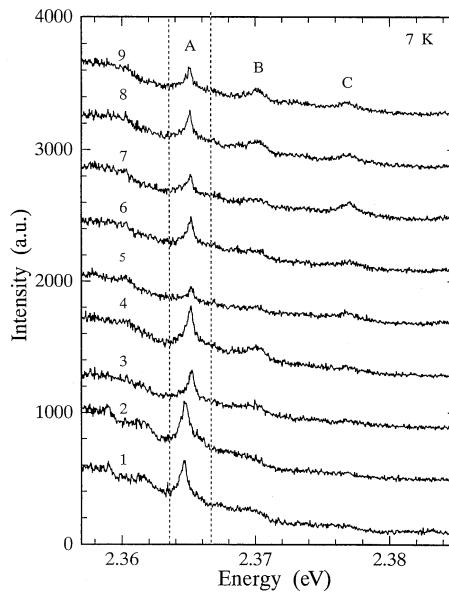


Figure 11.18. The luminescence spectrum in the high-energy tail of the band in figure 11.17 on the expanded energy scale. The excitation time of each spectrum is 1 s and the time interval of successive spectra is 15 s.

higher energy. The A line which is associated with such a small dot has a full width at half maximum less than 1 meV. From this figure, we see the energy shift and the intensity change of the A line with time. These phenomena are similar to those in CdSe dots synthesized as colloids [21]. They interpreted them as follows. The previously photogenerated electron or hole is Auger recombined resulting in ejection of an electron or a hole from a quantum dot and the ejected electron or hole is trapped at an impurity centre outside the dot. The successive photogenerated electron and hole in the dot are influenced by the electric field created by the trapped electron or hole in the vicinity of the dot. As a result, the recombination energy of the second electron and hole shifts and the transition probability becomes small. In the present experiment, the spectral change is one order of magnitude quicker, and the spectral diffusion range is only about 1 meV, which is much smaller than that of CdSe dots synthesized as colloids, more than 10 meV.

The quicker spectral change may originate from the shorter lifetime of the trapped electron in the CdSe well. The smaller amount of the spectral diffusion suggests that the Auger ionized electron (hole) is trapped far from the photoexcited nanocrystal. This is because the electron (hole) can move more easily in the CdSe quantum well which is the base of the dots. Hence, the Stark effect due to the trapped electron (hole) is expected to be smaller.

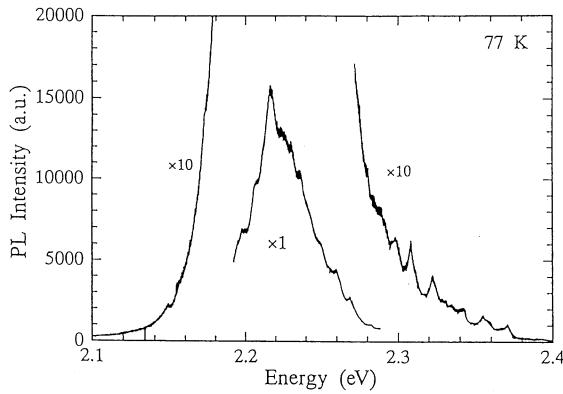


Figure 11.19. A microscopic luminescence spectrum of the sample excited by a 476.5 nm light at 77 K. The excitation time is 4 s.

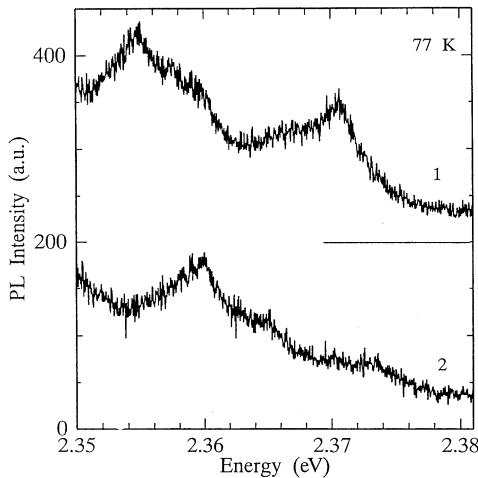


Figure 11.20. The expended luminescence spectra in the high-energy tail of the main broad band in figure 11.19. The time interval of spectra 1 and 2 is 4 s.

Figure 11.19 shows the microscopic photoluminescence of the sample excited by the 476.5 nm light at 77 K with a 4 s integration time. The luminescence spectrum consists of many small peaks. These peaks are considered to be due to the luminescences of single CdSe quantum dots, as in the CdSe at 7 K. These small peaks are broader than those in [figure 11.17](#) at 7 K. Figure 11.20 shows two consecutive 4 s integration spectra expanded in the smaller energy range at the higher energy side of the main structure of figure 11.19 at a given position of the sample at 77 K. The time interval between the two spectra is 4 s.

We notice that the spectral shape changes significantly as time goes on. Namely, the peak shift and the intensity change of the sharp lines becomes quicker as the temperature rises from 7 K to 77 K. As a result, the linewidth becomes larger at 77 K. These facts are interpreted by the proposed mechanism where a trapped electron (hole) in the CdSe quantum well is thermally released at 77 K and hence, the lifetime of the trapped electron becomes shorter.

In summary, we introduce a size quantization of excitons in an ultrathin nanocrystal of a layer type semiconductor PbI₂ and enhancement of the exciton binding energy in the ultrathin nanocrystals with three and four monolayers.

Second, temporal behaviours of absorption changes and luminescences in CuCl nanocrystals embedded in NaCl are analysed by a cascade relaxation model. The relaxation time of the electron–hole pair to the exciton, that of the bi-exciton to the exciton and the exciton lifetime are estimated to be 1.6 ps, 65 ps and 380 ps, respectively.

Third, we have observed a spectral diffusion of a single nanocrystal of CdSe on the CdSe quantum well base embedded in ZnSe. The spectral range and the diffusion time are much smaller than those of CdSe nanocrystals synthesized as colloids. These results suggest that the lifetime of an electron (hole) outside the nanocrystal is very short and the trapped electron (hole) locates apart from the nanocrystal.

References

- [1] Lavallard P 1998 *J. Cryst. Growth* **184/185** 352
- [2] Miller R C, Gossard A C, Tsang W T and Munteanu O 1982 *Phys. Rev. B* **25** 25
- [3] Naoe K, Zimin L G and Masumoto Y 1994 *Phys. Rev. B* **50** 18 200
- [4] Gaponenko S V 1996 *Semiconductors* **30** 315
- Zhang L X, Edamatsu K, Itoh T, Ono Y and Sato K 1997 *Nonlinear Opt.* **18** 341
- [5] Zhang L X, Edamatsu K, Itoh T, Ono Y and Sato K 1997 *Nonlinear Opt.* **18** 341
- Masumoto Y, Kawabata K and Kawazoe T 1995 *J. Lumin.* **52** 7834
- [6] Empedocles S A and Bawendi M G 1992 *Science* **278** 2114
- [7] Chepic D I, Efros Al L, Ekimov A I, Ivanov M G, Kharchednikov V A, Kudriavtsev I A and Yazeva T V 1990 *J. Lumin.* **47** 113
- [8] Masumoto Y, Okamoto S, Yamamoto T and Kawazoe T 1995 *Phys. Status Solidi b* **188** 209
- [9] Saito S and Goto T 1995 *Phys. Rev. B* **52** 5929
- [10] Schlüter I and Schlüter M 1974 *Phys. Rev.* **89** 1652
- [11] Hayashi T 1986 *J. Phys. Soc. Japan* **55** 2043
- [12] Skölnick M S and Bimberg D 1978 *Phys. Rev. B* **18** 7080
- [13] Nagamune Y, Takeyama S and Miura N 1989 *Phys. Rev. B* **40** 8099
- [14] Goto T and Saito S 1996 *J. Lumin.* **70** 435
- [15] Basterd G, Mendez E E, Chang L L and Esaki L *Phys. Rev. B* **26** 1974
- [16] Hanamura E, Nagaosa N, Kumagai M and Takagahara T 1988 *Mater. Sci. Eng. B* **1** 215

- [17] Itoh T, Iwai S, Edamatsu K, Yano S and Goto T 1995 *Japan. J. Appl. Phys.* **34** (Suppl. 34-1) 1
Edamatsu K, Iwai S, Itoh T, Yano S and Goto T 1995 *Phys. Rev. B* **51** 11 205
Yano S, Yamamoto A and Goto T 1998 *Phys. Rev. B* **57** 7203
- [18] Masumoto T 1988 *Appl. Phys. Lett.* **53** 1527
Tokizaki T, Akiyama H, Tanaka M and Nakamura A 1992 *J. Cryst. Growth* **117** 603
Yano S, Itoh T and Goto T 1996 *J. Appl. Phys.* **79** 8216
- [19] Itoh T, Iwabuchi Y and Kirihara T 1988 *Phys. Status Solidi b* **146** 531
Henserdon and Imbuch G F 1989 *Optical Spectroscopy of Inorganic Solids* (Oxford: Clarendon) p 285
- [20] Shen M Y, Goto T, Kurtz E, Zhu Z and Yao T 1998 *J. Phys.: Condens. Matter* **10** L171
- [21] Empedocles S A, Norris D J and Bawendi M G 1996 *Phys. Rev. Lett.* **77** 3873

Chapter 12

Electronic properties of InAs/GaAs quantum dots

**R Heitz¹, I Mukhametzhanov², O Stier¹, A Hoffmann¹,
A Madhukar² and D Bimberg¹**

¹ *Institut für Festkörperphysik, Technische Universität Berlin,
Berlin, Germany*

² *Photonic Materials and Devices Laboratory, University of
Southern California, Los Angeles, CA 90089-0241, USA*

We report a detailed study of the electronic properties of self-organized InAs/GaAs quantum dots (QDs) by photoluminescence (PL), time-resolved PL, and PL excitation (PLE) experiments. High-quality InAs/GaAs QDs of tunable size were obtained using the variable deposition amount approach in MBE-growth, yielding ultimately room-temperature emission at $1.3\text{ }\mu\text{m}$ for island area densities of $\sim 400\text{ }\mu\text{m}^{-2}$. The experiments emphasize the role of a slowed down carrier relaxation in the QDs, being important, for example, for energy transfer processes between QDs and the temperature dependence of the carrier capture processes. The quantum size effect of the excited-state spectrum is revealed in PLE experiments and shown to be in good agreement to numerical results for pyramidal QDs based on eight band $\mathbf{k}\cdot\mathbf{p}$ theory. Finally, phonon-assisted recombination processes are identified demonstrating an enhanced exciton–LO-phonon coupling. Excellent agreement with estimations in the adiabatic approximation suggests that this enhancement is the consequence of the particular quantum confinement and the piezoelectricity in the strained low-symmetry QDs.

12.1 Introduction

The formation of nanoscale coherent islands in highly strained semiconductor epitaxy has been extensively studied as a means to generate optically active

quantum dots (QDs) [1]. In spite of the large size inhomogeneity [2–5] of $\sim 10\%$ such Stranski–Krastanow QDs have been successfully employed in devices [6–10], demonstrating partly the predicted advantages of QD-based devices [11, 12]. Though self-organized QDs are easily incorporated in conventional device structures, the interdependent nature of the QD density and size makes the adaptation of the QDs to the device needs difficult. Additionally, the inhomogeneous broadening of the discrete density of states for self-organized QD ensembles hampers detailed investigations of the excited-state spectrum and of energy relaxation (and recombination) processes, which are both of basic physical interest and critical for design and performance of devices. In fact, it is often necessary to distinguish between extrinsic ensemble effects and intrinsic properties of single QDs, for example in describing the carrier dynamics [13, 14]. In recent years, extensive work has been devoted to the study of the excited states, as well as the temperature dependence and dynamical behaviour of the optical properties of self-organized QDs [4, 6, 15–27]. Furthermore, the prediction [28] of slowed-down carrier relaxation due to restricted inelastic phonon scattering in QDs was evidenced by the observation of multi-phonon resonances in photoluminescence excitation (PLE) spectra [24, 29–31]. However, the electronic properties of such self-organized QDs are still controversially discussed owing to the inhomogeneous broadening and the wide spread of reported structural properties of the individual QDs. Further progress in the understanding of the electronic properties of the self-organized QDs might be stimulated by improved samples.

In this paper, we review optical investigations of a new class of self-organized InAs/GaAs QD samples giving improved insight into the electronic properties of such QDs. The samples were grown in the variable deposition amount (VDA) approach [32] exploiting the concept of vertical stacking [33] to control the island formation process. A suitable seed layer and the InAs deposition amount in the second layer (Θ_2) are used to control, respectively, the island density and size for given growth conditions. Additionally, the island size distribution is improved in this growth approach enabling the realization of full widths at half maximum (FWHMs) down to 18 meV. The VDA approach allows for high-quality high-density InAs QDs emitting at $1.3 \mu\text{m}$ at room temperature [32, 34] and new studies, for example, of non-resonant energy transfer processes between neighbouring QDs [35, 36]. In the following, we will describe photoluminescence (PL), PLE, and time-resolved PL (TRPL) results for samples with widely varying carrier localization as a function of temperature. The results give insight into the carrier dynamics as well as the excited-state spectrum of self-organized InAs/GaAs QDs. The experimental results are discussed based on eight band $\mathbf{k} \cdot \mathbf{p}$ calculations for pyramidal InAs/GaAs QDs [37]. We find that the seed QDs in the VDA samples allow us to tune the QD size and improve the size and shape uniformity of the optically active QDs, whereas the effect on their electronic properties is negligible.

12.2 Experimental

The investigated InAs/GaAs QD samples were grown by molecular beam epitaxy on semi-insulating GaAs(001) substrates. InAs was deposited at 500 °C, an As₄ partial pressure of 6×10^{-6} Torr, and a growth rate of 0.22 MLs⁻¹. Special care was taken in capping the InAs QDs in order to maintain the island shape and minimize In/Ga interdiffusion. Therefore, after the InAs deposition the sample temperature was lowered to 400 °C for the deposition of the GaAs cap layer using migration enhanced epitaxy (MEE) as described in [4, 32]. Here we report results for single-layer samples and a variety of bilayer VDA samples with a 1.74 ML seed layer and 36 to 54 ML MEE-grown GaAs spacer layer all grown under the same growth conditions. The VDA growth approach provides enhanced control of the island formation process [32]. [Figure 12.1\(e\)](#) shows a schematic of the VDA growth approach, which exploits the well-established tendency of islands to form vertical stacks in multilayered structures due to the modulated surface strain (σ_{surface}) induced by the buried islands [33]. However, in contrast to previous attempts to generate a high density of identical islands in multilayered samples [25, 38–40], here the islands in the first layer (dubbed *seed* layer) are used to control the island density in the second layer. Then, the generally different InAs deposition Θ_2 in the second layer (and thus the InAs amount available for each island) allows controlled tuning of the average islands size in the second layer for given growth conditions. Additionally, the VDA samples benefit from an improved uniformity of the island ensemble [32].

Panels 1(a)–(d) of figure 12.1 show cross-sectional transmission electron microscopy (TEM) images of the 1.74 ML seed layer ($\Theta_2 = 0$) and three VDA samples with $\Theta_2 = 1.74$, 2.00 and 3.00 ML, respectively, and 36 ML spacers. The 36 ML spacer ensures almost perfect vertical alignment in QD pairs even for largely mismatched InAs depositions in the seed and the second layer in good agreement with previous observations for symmetric depositions [33]. The strain contrast of the islands in the second layer increases with increasing Θ_2 indicating an increasing island size. Finally, even for the 3.00 ML VDA sample, having the largest islands, we find only very few defected islands. More quantitative information on the islands was obtained from AFM investigations for identically grown but uncapped samples [32]. (Note that the growth of capped samples deviates only after the initial cooling down to 400 °C.) The AFM investigations yielded a QD density of 3.8×10^{10} cm⁻² { 4.8×10^{10} cm⁻²}, an average height of (38 ± 9) Å { (97 ± 7) Å}, and an average width of (169 ± 18) Å { (256 ± 12) Å} for the 1.74 ML single-layer {3.00 ML VDA} sample [32]. The AFM results indicate improved island uniformity in the VDA sample. Note, that the AFM measurements overestimate the lateral extension of the islands. Cross-sectional TEM images (not shown) for the uncapped 3.00 ML VDA sample indicate an average island baselength of ~ 19 nm.

The PL and PLE measurements were performed in a continuous-flow He-cryostat, providing temperatures between 3.5 and 360 K. PL was excited either

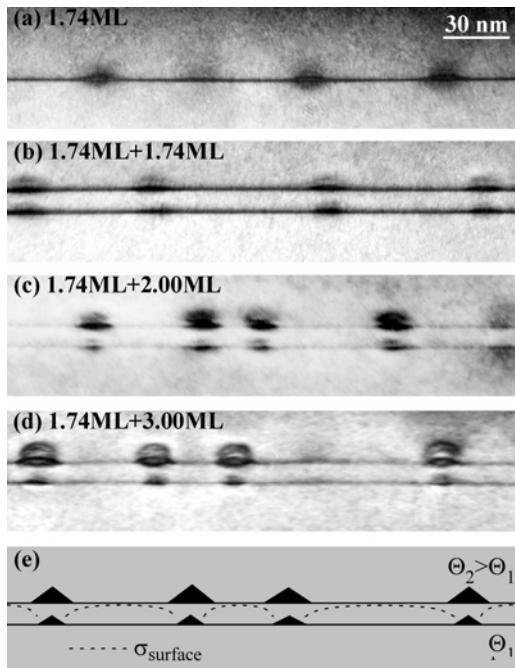


Figure 12.1. Cross sectional $g = (200)$ dark field TEM images of (a) a 1.74 ML single-layer sample, and double-layer samples with a 1.74 ML seed layer and (b) 1.74 ML, (c) 2.00 ML, and (d) 3.00 ML InAs in the second layer. Panel (e) shows a schematic of an asymmetrically stacked double-layer (VDA) sample. The surface stress σ_{surface} induced by the buried islands (indicated by the broken curve) causes vertical self-organization. The InAs depositions Θ_1 and Θ_2 control, respectively, the island density and size in the second layer.

by an Ar^+ laser or a tungsten lamp dispersed by a 0.27 m double-grating monochromator as tunable, low excitation density ($<0.02 \text{ W cm}^{-2}$) light source. The PL was dispersed in a double-grating monochromator and detected by a cooled Ge-diode (North Coast 817L). Time-resolved PL was excited by 6 ps pulses of a synchronously pumped dye-laser at 670 nm in the GaAs barrier, spectrally dispersed by a 0.35 m subtractive double-grating monochromator, and detected with a multichannel plate multiplier with a S1-cathode in photon-counting mode. The system response with a FWHM of ~ 50 ps was taken into account in the analysis of the transients.

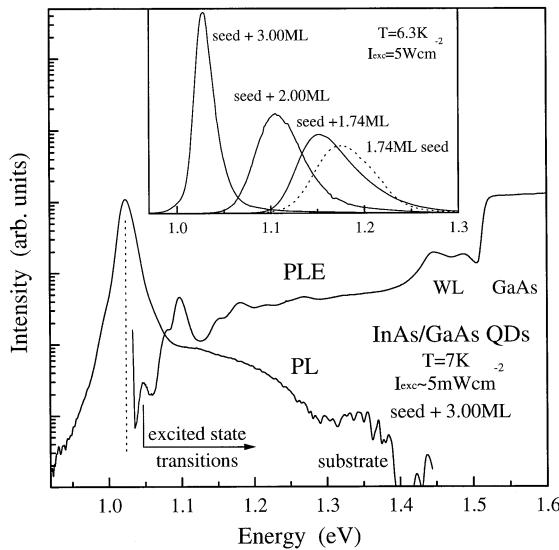


Figure 12.2. PL and PLE spectra for a 3.00 ML VDA sample. The FWHM of the ground-state PL is 18 meV. The inset shows the effect of a varying InAs deposition Θ_2 in the second layer.

12.3 Experimental results and discussion

Figure 12.2 depicts PL spectra of typical samples grown in the VDA approach. The inset compares PL spectra of the series of samples for which TEM images are shown in figure 12.1 revealing the influence of the InAs deposition amount Θ_2 in the second layer. At low excitation densities all samples show only a single PL peak attributed to the QD ground-state transition. For the 1.74 ML seed layer sample ($\Theta_2 = 0$) the PL peak is centred at 1.175 eV and has a FWHM of 75 meV. Increasing Θ_2 to 3.00 ML, the transition energy is decreased to 1.028 eV with an improved FWHM of 25 meV. This behaviour closely reflects the evolution of the structural properties of the islands in the second layer as observed in AFM and TEM investigations. Although the VDA samples contain asymmetric (with respect to size) QD pairs (AQDPs), see figure 12.1, obviously only the larger QD in the second layer is optically active. As will be shown later, the 36 ML spacer leads to energy transfer processes in the AQDPs, which efficiently quench the luminescence of the smaller seed QDs [35].

The main part of figure 12.2 depicts PL and PLE spectra of a second 3.00 ML VDA sample on a semilogarithmic scale. The practically matching ground-state transition energy (1.024 eV) demonstrates the excellent reproducibility of this growth approach and the FWHM of only 18meV of the almost Gaussian PL peak is the lowest one we have yet obtained. Note that semilogarithmic

PL spectra reveal for most samples a weak but extended high-energy shoulder. TRPL investigations indicate this luminescence to originate from the ground-state transition of smaller QDs, i.e. the seed QDs and/or (as indicated by some AFM data [32]) also a small number of unpaired islands in the second layer. More information on excited states, the wetting layer (WL), and relaxation processes are provided by PLE investigations. [Figure 12.2](#) shows a typical PLE spectrum recorded at the PL maximum, which reveals localized excitation via a series of intradot excited-state transitions for $E < 1.4$ eV as well as efficient delocalized excitation via the WL and the GaAs barrier [25]. The WL energy of 1.445 eV is typical for the employed growth conditions and corresponds to an ~ 1 ML thick InAs quantum well [41]. For the growth conditions used, the stacking process only marginally affects the WL transition energy and, therefore, the morphology of the WL in the second layer. The varying ground-state transition energy of the QDs (inset of figure 12.2) reflects the variation of the exciton localization with respect to the WL ranging from 270 meV for the 1.74 ML seed layer to 430 meV for the 3.00 ML VDA sample.

In the following, the various aspects of the PL and PLE spectra shown in figure 12.2 are discussed in detail for a variety of single layer and VDA samples providing new insight into the electronic properties of self-organized InAs/GaAs QDs. The experimental data are compared to results of numerical calculations of the electronic states of pyramidal InAs/GaAs QDs on the basis of the eight band $k \cdot p$ model described in detail in [37].

12.3.1 Energy transfer processes

As shown above the ([figure 12.1](#)) VDA growth approach generates well-defined AQDPs, for which the ground-state transition energies of both constituents are clearly distinguishable. The VDA samples show, however, only a single PL peak, while the expected PL of the smaller seed-layer QDs (SQDs) at ~ 1.2 eV is missing or at least two orders of magnitude weaker than the lower energy PL peak, figure 12.2. Obviously, only the larger QDs (LQDs) in the second layer are optically active and efficient energy transfer processes within the AQDPs quench the luminescence of SQDs [35, 36]. Similar energy transfer processes have been reported for mixtures of differently sized colloidal QDs and attributed to dipolar long-range resonance transfer [42], whereby the interdot separation was varied via the surface passivation layer. For the AQDPs the interdot separation is controlled on a ML-scale by the grown GaAs spacer thickness.

[Figure 12.3](#) compares PL spectra for a series of 2.00 ML VDA samples with GaAs spacers of 36, 45 and 54 ML thickness. The PL spectrum of the 36 ML sample shows only one, almost Gaussian, peak at 1.102 eV with a FWHM of 58 meV. With increasing spacer thickness an additional PL peak becomes evident on the high-energy side at ~ 1.21 eV, resulting in a clear double peak structure for the 54 ML spacer sample. Both, excitation in the GaAs barrier as well as in the WL yield similar results but show a small preference of SQDs in the former case.

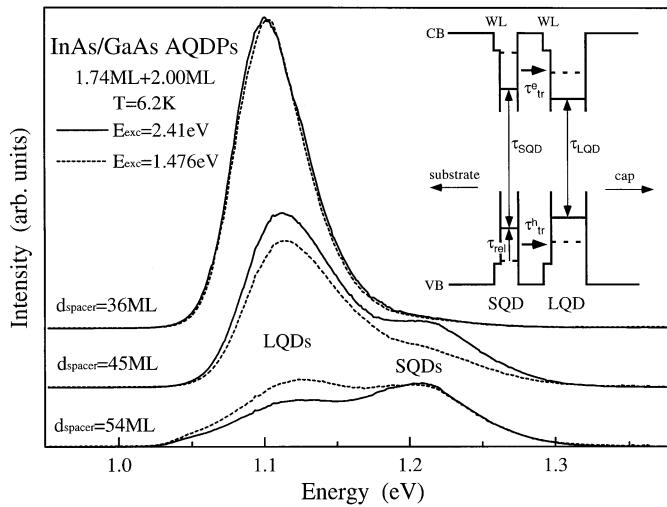


Figure 12.3. Low-temperature PL spectra for 2.00 ML VDA samples with varying spacer layer thicknesses. The inset shows a schematic energy diagram of the AQDPs.

For GaAs excitation the majority of carriers is generated on the substrate side of the QD layers and favourably captured by the SQDs ‘masking’ the LQDs. PLE measurements (not shown here) demonstrate that the energy transfer processes take indeed place between localized QD states, i.e. the carriers are first captured into the SQDs before they are efficiently transferred to the LQDs. The increase of the PL yield of the SQDs with increasing spacer thickness is attributed to the decreasing excitation transfer probability with increasing tunnel barrier thickness. Note that the actual tunnel barrier width is much smaller than the grown spacer thickness, since the 3D seed islands significantly protrude into the spacer. We identify the actual tunnel barrier with the separation between the seed island tip and the second InAs layer, which can be estimated based on AFM data for the seed islands to range from ~ 5.8 to ~ 11 nm for the investigated spacer thicknesses. The ratio of the integrated intensities of PL from SQDs and LQDs allows us to estimate the transfer efficiency η , ranging from nearly 100% for the 36 ML spacer down to $\sim 15\%$ for the 54 ML spacer [35].

As a result of the statistical fluctuations in island size and shape, we have to assume that in general the eigenstates of the SQD and LQD in the AQDPs are non-resonant. Furthermore, slowed-down intradot carrier relaxation [25, 43] on the timescale of some 10 ps might prevent carriers to reach the ground state in the SQDs before energy transfer takes place. TRPL investigations of the series of 2.00 ML VDA samples with various spacer thicknesses give further insight into the energy transfer mechanisms in the AQDPs. The transients have been recorded at low excitation densities for which all state filling effects could be neglected.

The transients recorded for various detection energies in the regions of PL from SQDs ($\gtrsim 1.16$ eV) and LQDs ($\lesssim 1.16$ eV) are well fitted by an exponential rise and decay convoluted with the system response function giving a resolution of ~ 10 ps for the PL rise times. Figure 12.4 depicts the PL rise (*a*) and decay (*b*) times as a function of the detection energy. Surprisingly, the decay time is practically independent of the detection energy and the spacer thickness. For symmetrically stacked samples grown under the same conditions also almost size-independent recombination times have been observed and attributed to the fact that the investigated QDs are close to the strong confinement regime [25]. However, the energy transfer processes, which efficiently quench the SQD PL yield, should be manifest in a pronounced but spacer thickness dependent decrease of the SQD decay time, which is not observed. Obviously, the excitation channel and not the SQD ground state itself is affected by the energy transfer processes in good agreement with the predicted strong vertical localization of the ground-state wavefunction [37]. In contrast, the PL rise time, describing carrier diffusion and relaxation processes, shows a characteristic spacer-dependent evolution. For the 36 ML and 45 ML spacer samples the rise time is much shorter for SQDs than LQDs, whereas no such difference is observed for the 54 ML spacer. The finite rise time (~ 30 ps) of the SQD PL observed for thin spacers is tentatively attributed to carrier diffusion processes. Obviously, the energy transfer processes compete with intradot carrier relaxation and, thus, originate from excited states of the QDs. The observed efficient excitation transfer in the AQDPs is therefore closely connected to the slowed-down relaxation processes in the QDs.

Further information on the nature of the energy transfer processes is obtained from the long-term dynamics of the QD PL. For samples with intermediate transfer yield, an additional slow decay component on a 10 ns timescale is observed. Typical transients are shown in figure 12.5 for a 2.00 ML VDA sample with 45 ML spacer, whereby 16 ML GaAs spacer have been replaced by AlGaAs to increase the tunnel barrier. A possible explanation for the second slow decay component of the LQD PL might be separate transfer of the electron and hole on different timescales. The PL yield of the SQDs is determined by the transfer of the faster carrier, forming an intermediate spatially indirect exciton state with the electron and hole in different QDs. The limited wavefunction overlap in such a spatially indirect exciton leads to a negligible recombination probability. The slower transfer of the second carrier determines the long living PL of the LQDs. Calculated eigenstate spectra and the corresponding wavefunctions suggest that most likely electron transfer is much faster than hole transfer. On one hand, the electron wavefunctions are less localized in the growth direction and, on the other hand, the hole states have smaller substate splittings allowing for faster relaxation. From the experimental data on the vertical energy transfer processes we estimate that for the investigated samples lateral energy transfer processes should become important only for QD sheet densities well above 2×10^{11} cm $^{-2}$ [35]. However, it is not clear yet how the transfer probability scales with the localization energy and therefore the QD size.

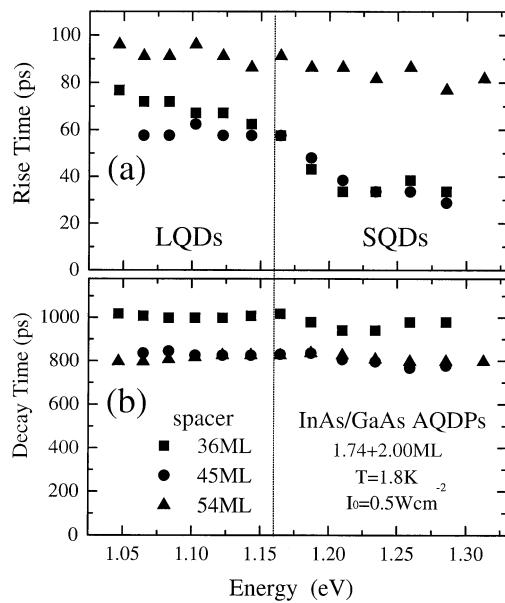


Figure 12.4. PL rise (a) and decay (b) times derived from low-excitation density transients for 2.00 ML VDA samples with various spacer thicknesses.

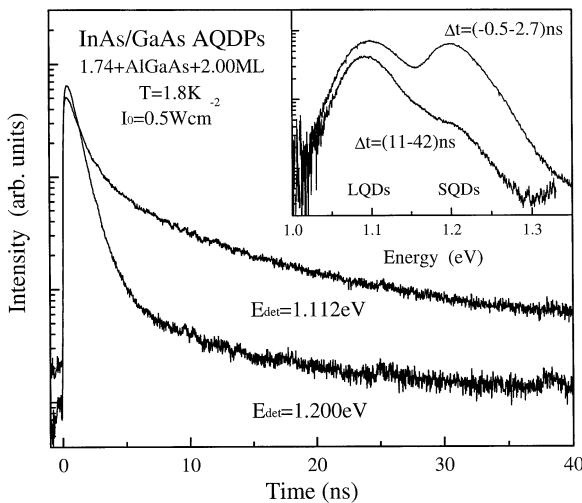


Figure 12.5. PL transients for the maxima of SQD and LQD PL in a 2.00 ML VDA sample with an additional 16 ML AlGaAs tunnel barrier between SQD and LQD. The inset depicts corresponding time-delayed PL spectra.

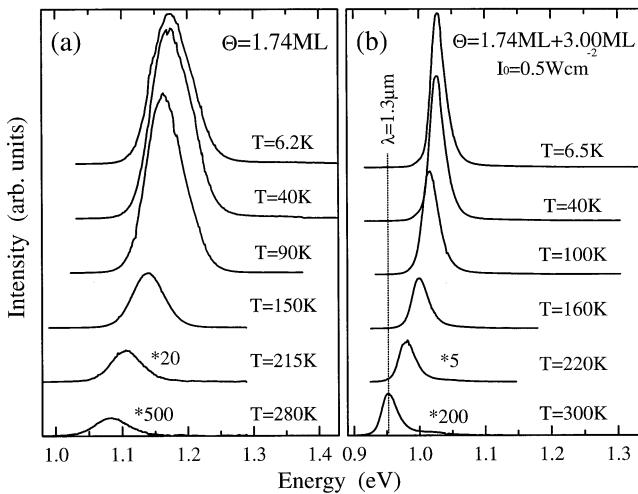


Figure 12.6. PL spectra for the 1.74 ML (a) and the 3.00 ML VDA (b) for various sample temperatures and an excitation density of 0.5 W cm^{-2} .

12.3.2 Temperature dependence

One main driving force for device applications of QDs is the temperature-stability of their properties resulting from the combination of the discrete density of states and a potentially large substate splitting. As shown in the inset of figure 12.2 the VDA growth approach allows us to grow for one and the same growth conditions and comparable island densities QD samples with different carrier localization energies. In the following we will compare properties of the 1.74 ML seed layer and the 3.00 ML VDA samples, which are governed by ensemble and intrinsic properties, respectively. Both samples are distinguished by the relative broadening of the exciton localization energy, which is much larger for the 1.74 ML sample (27%) than for the 3.00 ML (5.8%) sample.

Figure 12.6 shows PL spectra of the 1.74 ML single layer (panel (a)) and the 3.00 ML VDA (panel (b)) samples for various temperatures. The QDs in the 3.00 ML VDA sample show narrower and more temperature stable PL than the seed QDs as a result of their larger size and better uniformity. These PL spectra provide information on a variety of ensemble and intrinsic properties of the QDs, which will be discussed in more detail below. The main features are the temperature dependences of the PL energy and FWHM, the integrated PL intensity and the appearance of a high-energy sideband with increasing temperature. Note that the large QDs with an area density of $\sim 4 \times 10^{10} \text{ cm}^{-2}$ obtained in the VDA approach emit at $1.3 \mu\text{m}$ (FWHM = 29 meV) at room temperature; a wavelength of fundamental importance to fibre optic communications.

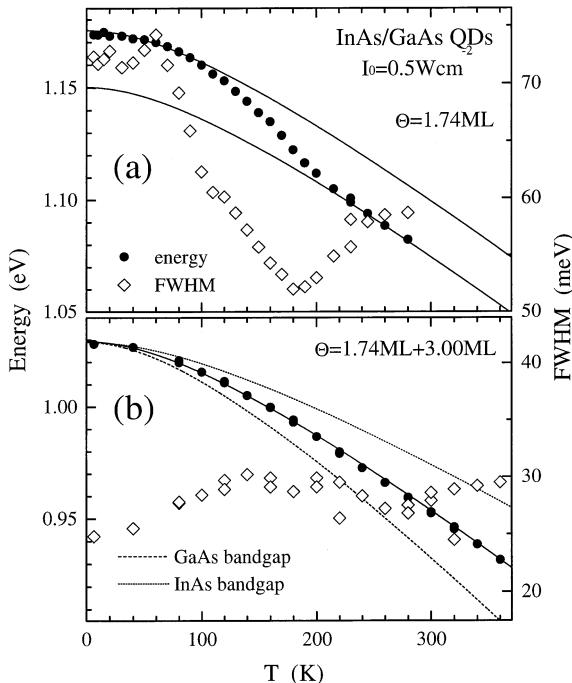


Figure 12.7. FWHM (diamonds) and energy (dots) of the ground-state PL for the 1.74 ML (a) and the 3.00 ML VDA (b). The temperature change of the bulk InAs and GaAs bandgaps are depicted by the dotted and broken curves in panel (b), respectively. Full curves show the variation of the QD ground-state transition energy in the 3.00 ML sample fitted by Varshni's formula [47].

Figure 12.7 depicts the energy (full dots) and FWHM (open diamonds) of the QD ground-state transition for the 1.74 ML single layer (panel (a)) and the 3.00 ML VDA (panel (b)) samples as a function of temperature, demonstrating a qualitatively different behaviour. For the 1.74 ML sample the FWHM decreases by $\sim 30\%$ for temperatures above ~ 80 K and the ground-state transition energy shows a simultaneous low-energy shift. We interpret this behaviour as an ensemble effect, which indicates preferential quenching of luminescence from the smaller QDs in the ensemble, providing less carrier localization [21, 22]. Such preferential quenching is obviously negligible for the more homogeneous 3.00 ML VDA sample, for which the FWHM increases slightly up to ~ 120 K and the ground-state transition shifts smoothly to lower energies. The uniformity of the QD ensemble in this sample enables the intrinsic properties of the single QDs to dominate the temperature dependences even if the experiment probes the inhomogeneous ensemble. The slight increase of the FWHM observed for temperatures up to 120K is attributed to the thermal population of excited hole

states expected a few meV above the ground state [37]. We note that the low-energy tail of the QD PL peak maintains its shape over at least two orders of magnitude (compare figure 12.2 for a semilogarithmic plot). Obviously, the inhomogeneous broadening dominates for temperatures up to 360 K and fits assuming a Voigt profile centred at the peak maximum place an upper limit of ~ 4 meV on the FWHM of the Lorentz contribution of single QDs at room temperature. At temperatures below 50 K FWHMs of less than $150 \mu\text{eV}$ have been reported for the homogeneous PL lines of single QDs [44, 45]. Phonon scattering processes are, however, expected to cause much larger homogeneous FWHMs at higher temperatures [46].

The temperature-dependent energy shift of bulk bandgaps is phenomenologically described by Varshni's formula [47]. The full line in panel (b) of figure 12.7 represents a corresponding fit to the transition energy in the 3.00 ML VDA sample. The temperature dependent shift falls in between of those of the bulk InAs (dotted curve) and GaAs (broken curve) bandgaps, which are expected to be the limits for large and small QDs, respectively. For the 1.74 ML sample the dominant ensemble effects make a straightforward fit of the energy data meaningless. However, both the low- and high-temperature regions are well described by the temperature-dependent shift observed for the 3.00 ML VDA sample (full curves in panel (a)), whereas in the intermediate region ensemble effects prevail and enhance the temperature-dependent red-shift. A detailed understanding of the temperature dependence of the ground-state transition energy of the self-organized InAs/GaAs QDs would require numerical calculations with temperature-dependent material parameters, which are not available yet.

The temperature evolution of the integrated QD PL intensity is depicted by full dots in figure 12.8 for the 1.74 ML (panel (a)) and 3.00 ML (panel (b)) samples. The intensity begins to decrease above ~ 100 K in good agreement with the observed onset of preferential quenching of the smallest QDs in the 1.74 ML sample. Above ~ 200 K the intensity-decrease becomes steeper. The integrated intensity is well fitted (full curves) assuming two thermally activated non-radiative recombination processes. The obtained activation energies are close to the excited-state splitting observed in high-density PL spectra and the localization energy of the ground-state exciton with respect to the WL, respectively. However, we emphasize caution in interpreting these activation energies. PLE spectra of the 3.00 ML VDA sample (inset of figure 12.8(b)) show a much stronger quenching of the efficiency for non-resonant (GaAs and WL) excitation compared to near-resonant excitation via excited-state transitions. Obviously, the decrease of the capture efficiency with increasing temperature dominates the quenching of the PL intensity upon non-resonant excitation. A similar behaviour has been observed for smaller QDs, too [25].

This effect becomes quite obvious comparing the integrated PL intensity with the lifetime of the ground-state exciton measured upon non-resonant excitation, open diamonds in figure 12.8. The exciton decay monitors the actual ground-state population and does not depend on the initial population yield.

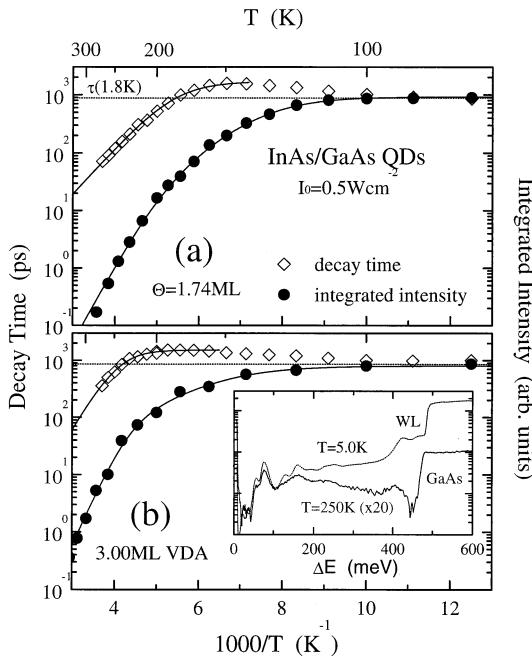


Figure 12.8. PL decay time (diamonds) and integrated PL intensity (dots) for the 1.74 ML (a) and the 3.00 ML VDA (b) samples. The horizontal dotted curves indicate the ground-state decay time at 1.8 K, respectively. Full curves show fits assuming one and two thermally activated non-radiative recombination processes, respectively. The inset of panel (b) compares PLE spectra recorded in the maximum of the QD PL ($\Delta E = E_{\text{exc}} - E_{\text{det}}$).

Below $\sim 100 \text{ K}$ the PL decay time is practically independent of temperature and QD size, being $(790 \pm 50) \text{ ps}$ and $(840 \pm 50) \text{ ps}$ (indicated by dotted curves), respectively, for the 1.74 ML single layer (panel (a)) and the 3.00 ML VDA (panel (b)) samples. Above $\sim 100 \text{ K}$ the lifetime increases up to $\sim 1600 \text{ ps}$ before it finally starts to decrease. The *radiative* recombination probability of a given exciton transition is temperature independent, accounting for the constant lifetime observed at low temperatures when the substate splitting is sufficiently large to prevent significant excited-state population. With increasing temperature, however, thermal population of excited states becomes important, resulting in an effective lifetime, which depends on the oscillator strength of the excited-state transitions. The hole states are much narrower spaced than the electron states [37] and, thus, populated first, which together with the low probability of transitions between excited hole states and the electron ground state explains the increase of the effective lifetime. Finally, thermally induced escape will dominate reducing the lifetime again. Qualitatively similar results for smaller QDs with $\sim 120 \text{ meV}$ exciton localization have been successfully simulated considering

one localized exciton state and delocalized WL states [48]. In the present case, however, a complete simulation is hampered by the need to include a spectrum of excited-state transitions. Fits of the exciton lifetime assuming a constant radiative lifetime ($\tau_{\text{rad}} = 1600$ ps) and complete non-radiative recombination for carriers activated to the barrier yields activation energies E_τ of (170 ± 30) meV and (240 ± 30) meV, respectively, for the 1.74 ML and the 3.00 ML sample. These activation energies indicate electron evaporation to the WL to be the driving mechanism behind the decreasing exciton lifetime. Indeed, owing to the fact that the exciton binding energy of ~ 20 meV in the QDs [37] is much smaller than the carrier localization energies, single carrier escape might be expected to dominate. Room temperature (300 K) lifetimes of ~ 35 and ~ 150 ps are estimated, respectively, which practically correspond to the electron escape times.

Mukai *et al* [19] proposed thermally activated non-radiative recombination in the barrier to explain the decreasing excitation efficiency of the QD PL in their samples. The activation energies were thereby characteristic for the non-radiative recombination centres. The present experimental results, however, suggest the carrier capture probability itself to be temperature dependent, i.e. to decrease with increasing temperature. A possible reason might be the slowed-down carrier relaxation in QDs, allowing thermally induced escape to dominate intradot relaxation for carriers captured into excited states, thus reducing the capture efficiency. Finally, the PLE spectra (inset of [figure 12.8\(b\)](#)) for the 3.00 ML sample demonstrate size-selective excitation at temperatures as high as 250 K, well above the regime where lateral exciton transfer should become important. Efficient carrier loss due to surface recombination and diffusion into the substrate prevent lateral coupling of the QDs, which would be a precondition for the formation of a common Fermi-level for the QD ensemble. Placing the active QD region between two AlGaAs diffusion barriers should suppress such loss mechanisms and increase the temperature stability of the QD PL intensity.

12.3.3 Local-equilibrium emission

The self-organized InAs/GaAs QDs provide strong carrier localization and a high-temperature stability of the PL intensity, enabling the exploitation of the thermal population of excited QD states in the optical investigation of their electronic properties. [Figure 12.9](#) depicts normalized PL spectra for the 3.00 ML VDA sample with respect to the energy of the peak maximum for temperatures up to 360 K in order to compare the lineshapes. In addition to the slight broadening of the ground-state peak I_g the excited-state transition I^* becomes visible shifted ~ 72 meV to higher energies for temperatures above ~ 180 K. At this stage the thermal population of the excited electron and hole states becomes significant enough to resolve corresponding transitions in the PL spectra. The inset of [figure 12.9](#) depicts the intensity ratio I^*/I_g over the inverse temperature. The full curve represents a fit to the intensity ratio assuming a three-level model with the degeneracies g_i and oscillator strengths f_i , yielding an activation energy of

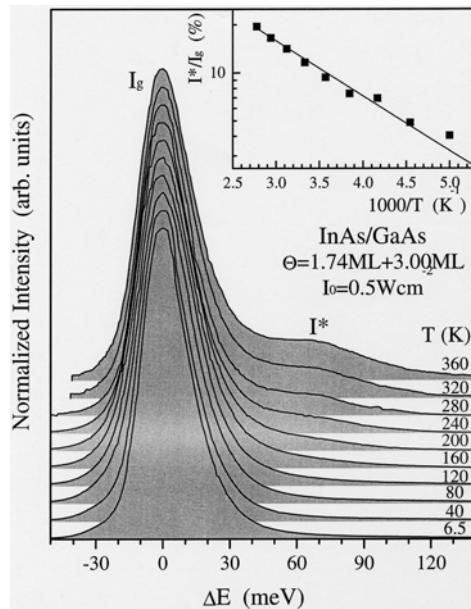


Figure 12.9. Normalized PL spectra for the 3.00 ML VDA sample for various temperatures. The insert shows the intensity ratio I^*/I_g on a semilogarithmic scale together with an exponential fit.

(70 ± 5) meV and a weighting factor $g^* f^* / g_g f_g$ of 1.7. The activation energy corresponds to the energy shift between the ground- and excited-state transitions and the weighting factor contributes a number of superimposed excited-state transitions as well as a possibly higher oscillator strength of some of the excited-state transitions [37]. Although the electronic structure of the QDs is more complex than the assumed three-level model (e.g., excited electron and hole levels are involved), figure 12.9 demonstrates a local equilibrium distribution of carriers between ground and excited states for each single QD.

The resolution of the non-resonantly excited PL spectra (figure 12.2) is limited by inhomogeneous broadening, i.e. the non-uniformity of the QD ensemble. Size-selective PLE spectra (e.g., figure 12.2) show, however, rich fine structure in the near resonant excitation regime, suggesting that, for example, in high-density PL spectra [49–51] the electronic fine structure of the QDs is masked. The interpretation of the PLE spectra, which involve excited-state *absorption* and carrier *relaxation* processes, is, however, ambiguous in the case of QDs. On one hand, since the self-organized InAs/GaAs QDs vary in size and shape the excited-state spectrum of QDs with the same ground-state transition energy (E_{det}) is generally inhomogeneously broadened, limiting the resolution in PLE experiments [25]. On the other hand, the slowed down carrier relaxation

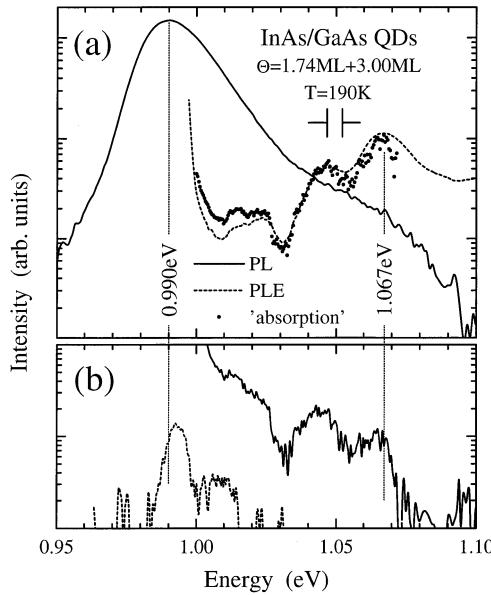


Figure 12.10. PL (full curves) and PLE (broken curves) spectra for the 3.00 ML VDA sample at 190 K. Panel (a) shows non-resonantly excited PL and PLE detected in the maximum of the PL spectrum. Panel (b) depicts ‘anti-Stokes’ PL excited in the ground-state transition (0.990 eV) and PLE spectrum for an excited-state transition (1.067 eV). The dots in panel (a) (dubbed ‘absorption’) show the resonantly excited PL spectrum divided by the thermal carrier distribution $\exp(-\Delta E/kT)$ in QDs occupied with one exciton.

makes the QDs prone to possible non-radiative recombination processes, the so-called ‘phonon-bottleneck’ effect [28]. Both effects together might lead to the occurrence of multi-LO-phonon resonances in PLE spectra, which indeed have been observed recently [29–31], demonstrating inelastic phonon scattering as the dominant relaxation process [24]. TRPL measurements show the extrinsic nature of the competing recombination processes, most likely non-radiative recombination at defects in the vicinity of the QDs, i.e. in the low-temperature GaAs barrier. Avoiding such non-radiative recombination by optimization of the growth of the GaAs cap layer has been proposed to allow the unhampered resolution of excited-state transitions in the PLE spectra [25, 52].

Here, we propose to take advantage of the *thermal population* of excited QD states upon *resonant excitation of the ground state* to investigate the ground-state transition as well as the excited-state spectrum of the QDs in more detail. Figure 12.10 shows optical spectra for the 3.00 ML VDA sample at an elevated sample temperature of $T = 190\text{K}$. Panel (a) shows a non-resonantly excited PL spectrum (full curve) and a PLE spectrum detected in a 3.4 meV window

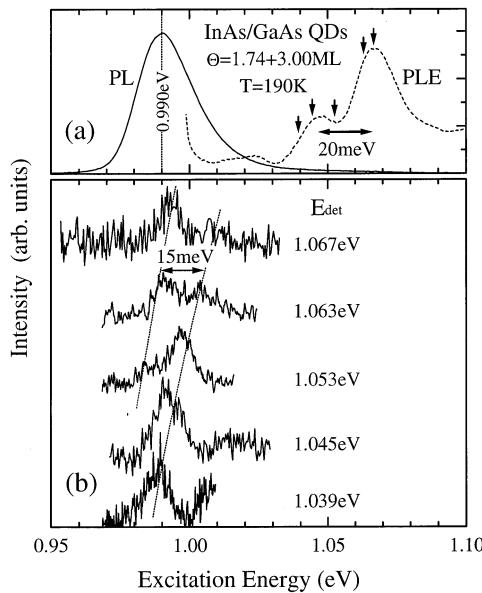


Figure 12.11. (a) PL and PLE spectra for the 3.00 ML VDA sample at 190 K. Panel (b) depicts ‘anti-Stokes’ PLE spectra for various detection energies (marked by arrows in panel (a)).

around the PL maximum at 0.990 eV on a semilogarithmic scale. The PL spectrum clearly shows the excited-state transition at ~ 1.065 eV, whereas the size-selective PLE spectrum reveals a richer finestructure with the most intense excitation resonance at 1.067 eV close to the excited-state PL peak position. For such conventional spectra the excitation energy is *larger* than the detection energy.

Panel (b) in figure 12.10 compares PL (full curve) and PLE (broken curve) spectra where the excitation energy is *smaller* than the detection energy. Thus, the excitons are excited in the QD ground state and subsequent thermal excitation into excited states leads to the detected excited-state luminescence. Exciting resonantly the ground-state transition ($E_{\text{exc}} = 0.990$ eV), a series of weak luminescence structures is observed on the *high-energy* side of the ground-state transition, which in non-resonantly excited PL spectra (panel (a)) are lumped together in one excited-state transition peak. These lines are attributed to recombination involving various excited states of the selectively excited QD subensemble. The PLE spectrum of the excited-state transition at 1.067 eV reveals the ground-state absorption at 0.993 meV. The 3 meV high-energy shift with respect to the PL maximum is attributed to the exponential modulation caused by the thermal carrier distribution favouring the PL of QDs with a smaller substate splitting, i.e. higher ground-state transition energy. This effect is more obvious in figure 12.11(b), which shows a series of PLE spectra recorded at

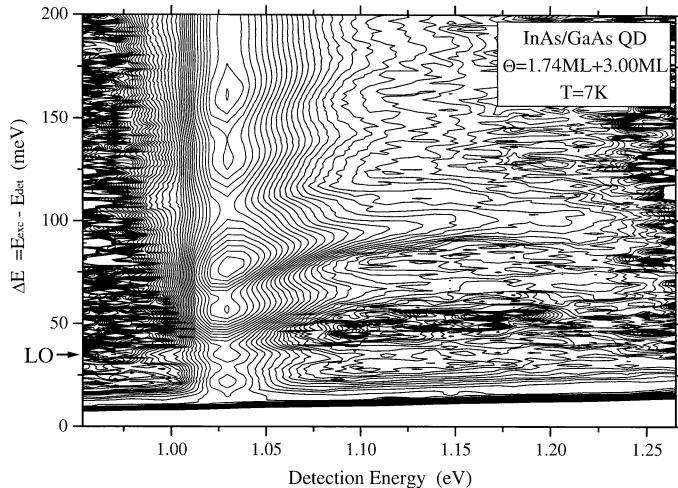


Figure 12.12. Contour plot of the PL intensity in the 3.00 ML VDA sample as a function of the detection energy and the excess excitation energy $\Delta E = E_{\text{exc}} - E_{\text{det}}$. The intensity is given on a logarithmic scale.

$T = 190$ K for different detection energies as indicated by arrows in panel (a). The excitation resonances show a red-shift with decreasing detection energy indicating size-selective excitation of the ground-state transition. Most PLE spectra reveal a doublet structure separated by ~ 15 meV, which is slightly smaller than the separation of the two dominating excited-state transitions (~ 20 meV, PLE spectrum in panel (a)). The doublet structure results from ground-state excitation of different sized QDs with one of these two excited-state transitions in resonance with the detection energy. Thus, the different energy splitting observed is the result of the QD size dependence of the excited-state splitting.

For the PLE spectrum of the ground-state transition (broken curve in figure 12.10(a)) and the resonantly excited PL spectrum (full curve in figure 12.10(b)) the roles of excitation and detection are somewhat reversed. PLE probes the excited-state absorption spectrum, but might be modified by restricted carrier relaxation processes and carrier localization at defects, whereas the PL spectrum is given by the absorption spectrum modified by the thermal carrier distribution among the localized QD states. The ‘absorption’ spectrum, extracted by dividing the resonantly excited PL by the thermal population distribution $\exp(-(E_{\text{exc}} - E_{\text{det}})/kT)$ for QDs occupied by one exciton, is given by the dotted spectrum in figure 12.10(a). The ‘absorption’ spectrum is found to reproduce the PLE spectrum within experimental accuracy. For the investigated samples the PLE spectrum indeed follows the absorption spectrum of a subensemble defined by the ground-state transition energy unhampered by the slowed down carrier relaxation dynamics. Additionally, the ground-state absorption and emission

energies are obviously perfectly matched for the self-organized InAs/GaAs QDs, demonstrating the intrinsic nature of the exciton recombination. Finally, the near resonant fine structure observed in the PLE spectra is practically unchanged for temperatures between 7 and 250 K (inset of figure 12.8(b)), suggesting carrier scattering between the different localized QD states to be faster than competing recombination processes even at low temperatures. Figure 12.10 shows, thus, that PLE spectroscopy gives reliable access to the excited-state spectrum of the QDs, providing high-quality samples with negligible non-radiative recombination.

12.3.4 Excited-state transitions

As outlined in the previous section, PLE spectroscopy presents for the investigated samples an ideal tool to study size-selectively the excited-state transition spectrum of the QDs. Varying the detection energy over the PL peak QDs of different average size are probed, enabling the measurement of the size-dependence of the excited-state spectrum for one sample exploiting the inhomogeneous broadening. Figure 12.12 shows for the 3.00 ML VDA sample a contour plot of the PL intensity as a function of the detection energy and the excess excitation energy $\Delta E = E_{\text{exc}} - E_{\text{det}}$ at 7 K. The PL intensity is given on a logarithmic scale. For the various excitation resonances the energy shift ΔE , corresponding to the combined excited-state energy of the involved electron and hole states (with respect to the respective ground-state energy), depends clearly on the detection energy and, therefore, the QD size. For example, with decreasing detection energy ΔE decreases from 101 to 70 meV for the dominating excitation resonance. The decrease of ΔE with decreasing detection energy demonstrates the quantum-size effect of the excited-state splitting for the self-organized InAs/GaAs QDs, supporting the notion of the PLE spectra revealing the absorption spectrum. The single exemption is the resonance at ~ 35 meV, for which ΔE is independent of the detection energy indicating phonon-assisted absorption as discussed in the next section. The various excitation resonances have FWHMs ≥ 10 meV, which are not limited by the experimental resolution (~ 4 meV). The combination of size *and* shape fluctuations of the self-organized QDs cause the excited-state spectrum for a given ground-state transition energy to be still inhomogeneously broadened [25]. The FWHM of the excitation resonances might be taken as a measure for the shape uniformity of the QDs.

Figure 12.13 depicts two PLE spectra of the 3.00 ML VDA sample taken near the PL maximum as a function of the excess excitation energy ΔE . The most intense excitation resonances appear at energies of ~ 75 , ~ 160 and ~ 240 meV corresponding roughly to the excited-state transitions observed in inhomogeneously broadened high-density PL spectra (not shown). However, additional weaker resonances are resolved at ~ 22 , ~ 60 , ~ 130 , ~ 190 , ..., meV and lineshape fits to the PLE spectra suggest the presence of even more resonances. The excited-state transition spectrum of the QDs is, therefore, denser

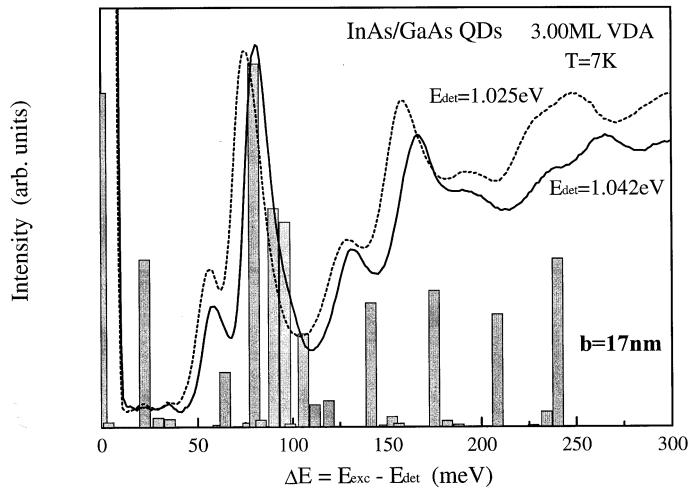


Figure 12.13. PLE spectra for the 3.00 ML VDA sample for two different detection energies. Bars show the absorption strength of various electron and hole transitions calculated for a 17 nm baselength pyramid [37]. The exciton ground-state energy of the pyramid is 1.027 eV.

than expected from inhomogeneously broadened spectra. These results give experimental evidence for the failure of a parabolic description of the electronic states of the QDs leading to a $\Delta n = 0$ selection rule for dipole transitions as well as regularly spaced excited-state transitions. Therefore, we compare the experimental results to numerical calculations based on the eight band $\mathbf{k} \cdot \mathbf{p}$ approximation for pyramidal InAs/GaAs QDs with a base length between 10 and 20 nm described in detail in [37]. The calculations indicate a variety of bound electron and hole states, whereby the substate splittings for the hole states are in the 10 meV region and those of the electron states are in the 50 meV region. Absorption spectra calculated based on the obtained wavefunctions show transitions between a variety electron and hole levels to have sufficient oscillator strength to be observed in experiments. The low symmetry of such strained self-organized QDs to cause a breakdown of the $\Delta n = 0$ selection rule, although the most intense transitions have still matching ‘quantum numbers’ of the electron and hole wavefunctions [37]. Note, that in the strained pyramidal QDs n is no longer a good quantum number, and has been replaced for the classification of the wavefunction by the somewhat arbitrary number of nodes [37]. The calculated absorption strengths of the lowest energy transitions in a 17 nm baselength pyramid is represented by bars in figure 12.13. The height of the bars corresponds to the relative transition probability. Figure 12.13 shows a good qualitative agreement between calculation and experiment supporting the eight band $\mathbf{k} \cdot \mathbf{p}$ calculations and the assumed QD shape [53]. The excitation

resonance at ~ 22 meV corresponds to the transition between the electron ground state and the second excited hole state. At higher energies transitions involving excited electron states contribute to the spectrum. We remind, that in the 3.00 ML sample the optically active QDs are grown on top of the smaller seed islands, which might influence their properties. However, for the comparatively thick (36 ML) spacer electronic coupling is negligible [25] and, additionally, numerical calculations for such AQDPs indicate the strain-interaction to lead to *blue*-shift of the transition energies by a few meV, which is negligible for the presently available experimental results.

12.3.5 Phonon-assisted exciton recombination

Finally, we will discuss the exciton–LO-phonon coupling in the self-organized InAs/GaAs QDs, which is of fundamental interest for carrier relaxation processes as well as a sensitive test for the electron and hole wavefunctions, i.e. for the numerical calculations based on idealized QD shapes as discussed in the previous section. Recently, multi-LO-phonon progressions have been observed in PLE spectra of such QDs and unambiguously attributed to the effects of the carrier dynamics following excited-state absorption [24, 30]. These results demonstrated inelastic phonon scattering to be the dominant carrier relaxation process at low carrier densities, but give no access to the actual exciton–LO-phonon coupling strength. However, for uniform InAs/GaAs QD samples with a ground-state PL FWHM smaller than typical LO-phonon energies characteristic low-energy shoulders are observed for the ground-state PL, see, for example, figures 12.2 and 12.14. These shoulders can be identified as phonon-assisted recombination processes giving direct access to the exciton–LO-phonon coupling strength [54].

The phonon-assisted exciton transitions are clearly resolved in site-selective PL and PLE spectra, circumventing the inhomogeneous broadening of the QD ensemble, figure 12.14(a). The pronounced modulation of the PLE efficiency (dotted spectrum) allows for the selective excitation of QDs with a ground-state transition energy in resonance with the inhomogeneous PL maximum. Exciting the PLE resonance at 1.103 eV, the QD PL peak at 1.027 eV (dubbed zero phonon line (ZPL)) becomes narrower (FWHM = 11 meV)¹ and a peak (dubbed –LO) is clearly resolved in the region of the first low-energy shoulder, having $\sim 1.5\%$ of the intensity of the ZPL. The –LO replica becomes a well-resolved separate peak exciting directly the ZPL. Gaussian fits of such higher resolved spectra yield an energy of 34.2 meV for the phonon mode in good agreement with previous results for the QD LO phonon energy [24, 30]. Figure 12.13(b) shows the dependence of integrated intensity of the +LO-phonon resonance observed in PLE spectra on the detection energy, proving the phonon-assisted transition to be resonant with the QD ground-state transition. Two resonances centred at 1.027 and 0.993 eV are revealed, which are resonant to the ground-state transition and red-shifted

¹ The weak shoulder at 1.044 eV results from smaller QDs excited via a different excited-state transition.

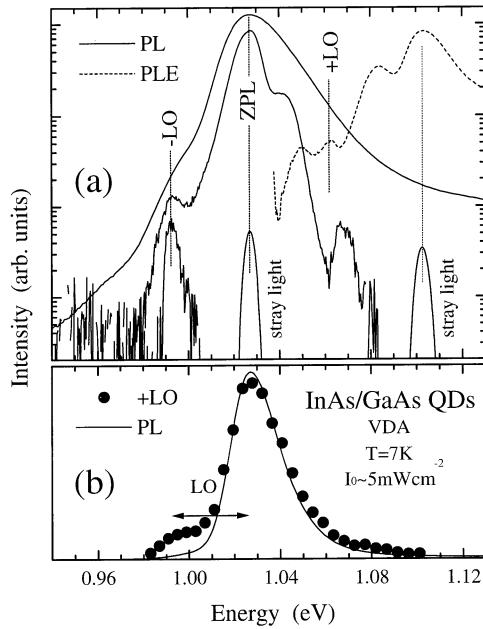
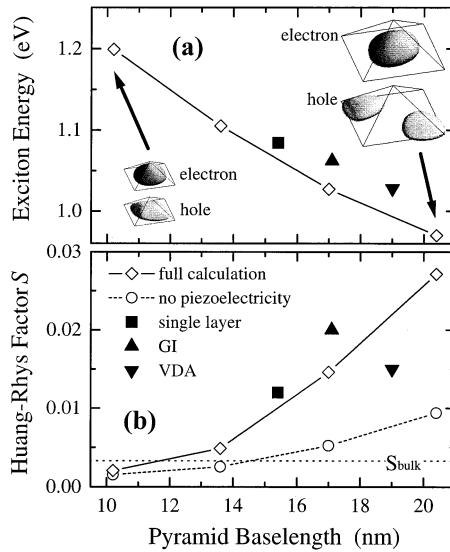


Figure 12.14. (a) Low-temperature PL (full curves) and PLE (broken curve) spectra for the 3.00 ML VDA sample on a semilogarithmic scale. The PLE spectrum was detected on the QD PL maximum at 1.028 eV (marked ZPL) and the PL spectra were excited in the GaAs barrier (uppermost spectrum), at 1.103 eV and via the ZPL. (b) Energy dependence of the intensity of the +LO resonance in PLE.

by one LO-phonon energy, respectively. The LO phonon is excited either in the generation or recombination of the ground-state exciton, respectively. The PLE spectra of the ZPL and the $-LO$ resonance (not shown) [54] show identical excitation resonances proving that in both cases the same QD subensemble is probed as expected for a phonon-assisted transition.

The observation of the phonon-assisted optical transitions allows the determination of the Huang–Rhys parameter S for the polar exciton–LO-phonon coupling in the self-organized InAs/GaAs QDs. At low temperatures, the Huang–Rhys parameter S is given by the ratio of the intensities of the first LO-replica ($-LO$) and the ZPL [55], yielding $S = 0.014$ for the 3.00 ML VDA sample. The experimental results for the ground-state transition energy and S are depicted in figure 12.15. Results from equivalent experiments (not shown) for smaller QDs in differently grown single-layer samples (dubbed GI and 3ML) are included, too [54].

The Huang–Rhys factors S determined for the ground-state exciton in the self-organized InAs/GaAs QDs are ~ 5 times larger than in bulk InAs ($S_{\text{bulk}} \sim$



baselengths of 10.2 and 20.4 nm, respectively. The electron wavefunction is predominantly s-like, whereas the hole wavefunction is partly p-like and, with increasing pyramid size, concentrates in the [1–10] corners of the pyramid as a result of the strain-induced piezoelectric potential.

The Huang–Rhys parameter S is estimated within the adiabatic approximation for the interaction of the ground-state exciton with bulk phonon modes, following the calculations of Nomura and Kobayashi (equation (8) in [56]) and using the above-described wavefunctions. The results are depicted as open diamonds in figure 12.15(b). S increases from 0.0020 to 0.027 by more than one order of magnitude doubling the pyramid baselength from 10.2 to 20.4 nm. The piezoelectricity modifying the hole wavefunction [37] decisive for the pronounced enhancement of the coupling strength. Neglecting the piezoelectricity results in an ~70% reduction of the Huang–Rhys parameter S (open dots in figure 12.15(b)). The good agreement between experimental (full symbols) and calculated S values indicates that the adiabatic approximation is able to explain the polar exciton–LO-phonon coupling in the coherent, defect-free InAs/GaAs QDs and supports the validity of the eight band $\mathbf{k} \cdot \mathbf{p}$ calculations. The apparent lack of a clear size-dependence in the experimental data (full symbols) results from the fact that the GI and 3 ML samples have single QDs whereas the VDA sample contains AQDPs. Numerical calculations for such AQDPs show the strain interaction between the small seed island and the optically active QD to cause mainly a reduction of the piezoelectric potential, leading to a significant decrease (~40%) of the Huang–Rhys parameter S compared to an equivalent unpaired QD.

12.4 Conclusions

We have reported on a thorough study of the optical properties of self-organized InAs/GaAs QDs using PL, PLE and TRPL techniques as a function of temperature. The investigated samples contain asymmetric QD pairs allowing to control the size and improve the uniformity of the optically active QDs in the second layer. For the 36 ML spacer used for most samples the electronic interaction of the QDs in such AQDPs is negligible, but only the larger one is optically active due to efficient intrapair energy transfer with a yield of ~1. Increasing the spacer thickness allows one to suppress the energy transfer processes and, thus, to tune the carrier dynamics from the transfer to the recombination limited regime. The VDA samples enable the growth of a high density (~400 μm^{-2}) of high-quality QDs providing room-temperature emission at 1.3 μm , exciton localization with respect to the WL of ~430 meV, and a PL FWHM of only 18 meV. Room-temperature exciton lifetimes of up to 150 ps are estimated, which are determined by the thermal escape of the electrons. The PL intensity for non-resonant excitation is, however, quenched due to a decreasing capture efficiency with increasing temperature. The results suggest that AlGaAs diffusion barriers might drastically enhance the room-temperature PL yield.

PLE spectra are shown to reveal the excited-state transition spectrum for the investigated VDA samples, showing a much richer fine structure than expected from inhomogeneously broadened high-density PL spectra. A comparison to eight band $\mathbf{k} \cdot \mathbf{p}$ calculations shows a breakdown of the $\Delta n = 0$ selection rule and leads to a qualitative understanding of the excited-state spectrum. The presence of the seed QD is shown to have only a negligible effect on the electronic properties of the optically active QDs in the second layer. Finally, phonon-assisted exciton recombination processes are identified indicating an enhanced polar exciton–LO-phonon interaction in the self-organized InAs/GaAs QDs. The experimental observations are well explained assuming adiabatic coupling to bulk-like phonon modes taking into account realistic wavefunctions. The particular quantum confinement and the piezoelectricity in the strained low-symmetry QDs result in the enhanced coupling strength. The presented results show that far-field spectroscopy on high-quality samples is able to provide a deep understanding of the electronic properties of self-organized InAs/GaAs QDs.

The variable deposition amount growth approach gives unprecedented control on the formation of self-organized QDs allowing to position, tune and shape these coherent defect-free QDs. In the present study we have used a self-organized seed layer. However, using artificially prepatterned stressors with a regular surface arrangement as seed layer might further improve this growth concept.

Acknowledgments

Parts of this work were supported by the US Air Force Office of Scientific Research, the US Office of Naval Research and the Deutsche Forschungsgemeinschaft in the framework of Sonderforschungsbereich 296.

References

- [1] For a recent review see: Bimberg D, Grundmann M and Ledentsov N N 1999 *Quantum Dot Heterostructures* (London: Wiley)
- [2] Leonard D, Pond K and Petroff P M 1994 *Phys. Rev. B* **50** 11 687
- [3] Moison J M, Houzay F, Barthe F, Leprince L, André E and Vatel O 1994 *Appl. Phys. Lett.* **64** 196
- [4] Xie Q, Chen P, Kalburge A, Ramachandran T R, Nayfonov A, Konkar A and Madhukar A 1995 *J. Cryst. Growth* **150** 357
- [5] Kobayashi N P, Ramachandran T R, Chen P and Madhukar A 1996 *Appl. Phys. Lett.* **68** 3299
- [6] Kirstaedter N *et al* 1994 *Electron. Lett.* **30** 1416
Bimberg D, Kirstaedter N, Ledentsov N N, Alferov Zh I, Kop'ev P S and Ustinov V M 1997 *IEEE Sel. Topics Quantum Electron.* **3** 196
- [7] Xie Q, Kalburge A, Chen P and Madhukar A 1996 *IEEE Photon. Technol. Lett.* **8** 965
- [8] Campbell J C, Huffaker D L, Deng H and Deppe D G 1997 *Electron. Lett.* **33** 1337

- [9] Yusa G and Sakaki H 1997 *J. Appl. Phys.* **70** 345
- [10] Finley J J, Skalitz M, Arzberger M, Zrenner A, Böhm G and Abstreiter G 1998 *Appl. Phys. Lett.* **73** 2618
- [11] Asada M, Miyamoto Y and Suematsu Y 1986 *IEEE J. Quantum Electron.* **22** 1915
- [12] Arakawa Y and Sakaki S 1982 *Appl. Phys. Lett.* **40** 939
- [13] Schmidt O *et al* 1996 *Electron. Lett.* **32** 1302
- [14] Grundmann M and Bimberg D 1997 *Phys. Rev. B* **55** 9740
Grundmann M and Bimberg D 1997 *Japan. J. Appl. Phys.* **36** 4181
- [15] Wang G, Fafard S, Leonard D, Bowers J E, Merz J L and Petroff P M 1994 *Appl. Phys. Lett.* **64** 2815
- [16] Marzin J-Y, Gérard J M, Izrael A, Barrier D and Bastard G 1994 *Phys. Rev. Lett.* **73** 716
- [17] Ledentsov N N *et al* 1996 *Solid-State Electron.* **40** 785
- [18] Yang W, Lowe-Webb R R, Lee H and Sercel P C 1997 *Phys. Rev. B* **56** 13 314
- [19] Mukai K, Ohtsuka N and Sugawara M 1997 *Appl. Phys. Lett.* **70** 2416
- [20] Adler F, Geiger M, Bauknecht A, Haase D, Ernst P, Dörnen A, Scholz F and Schweizer H 1998 *J. Appl. Phys.* **83** 1631
- [21] Xu Z Y, Lu Z D, Yang X P, Yuan Z L, Zheng B Z, Xu J Z, Ge W K, Wang Y, Wang J and Chang L L 1996 *Phys. Rev. B* **54** 11 528
- [22] Lubyshev D I, González-Borrero P P, Marega E, Petitprez E, La Scala N and Basmaji P 1996 *Appl. Phys. Lett.* **68** 205
- [23] Grundmann M, Heitz R, Bimberg D, Sandmann J H H and Feldmann J 1997 *Phys. Status Solidi b* **203** 121
- [24] Heitz R, Veit M, Ledentsov N N, Hoffmann A, Bimberg D, Ustinov V M, Kop'ev P S and Alferov Zh I 1997 *Phys. Rev. B* **56** 10 435
- [25] Heitz R, Kalburge A, Xie Q, Grundmann M, Chen P, Hoffmann A, Madhukar A and Bimberg D 1998 *Phys. Rev. B* **57** 9050
- [26] Vollmer M, Mayer E J, Rühle W W, Kurtenbach A and Eberl K 1996 *Phys. Rev. B* **54** R17 292
- [27] Schmidt K H, Medeiros-Ribeiro G, Oestereich M, Petroff P M and Döhler G H 1996 *Phys. Rev. B* **54** 11 346
- [28] Benisty H, Sotomayor Torres C M and Weisbuch C 1991 *Phys. Rev. B* **44** 10 945
- [29] Ledentsov N N *et al* 1995 *Proc. 22nd Int. Conf. on the Physics of Semiconductors (Vancouver, Canada, 1994)* vol 3, ed D J Lockwood (Singapore: World Scientific) p 1855
- [30] Heitz R, Grundmann M, Ledentsov N N, Eckey L, Veit M, Bimberg D, Ustinov V M, Egorov A Yu, Zhukov A E, Kop'ev P S and Alferov Zh I 1996 *Appl. Phys. Lett.* **68** 361
Heitz R, Grundmann M, Ledentsov N N, Eckey L, Veit M, Bimberg D, Ustinov V M, Egorov A Yu, Zhukov A E, Kop'ev P S and Alferov Zh I 1996 *Surf. Sci.* **361/362** 770
- [31] Steer M J, Mowbray D J, Tribe W R, Skolnick M S, Sturge M D, Hopkinson M, Cullis A G, Whitehouse C R and Murray R 1997 *Phys. Rev. B* **54** 17 738
- [32] Mukhametzhanov I, Heitz R, Zeng J, Chen P and Madhukar A 1998 *Appl. Phys. Lett.* **73** 1341
- [33] Xie Q, Madhukar A, Chen P and Kobayashi N P 1995 *Phys. Rev. Lett.* **75** 2542
- [34] Heitz R, Mukhametzhanov I, Madhukar A, Hoffmann A and Bimberg D *J. Electron. Mater.* to be published

- [35] Heitz R, Mukhametzhanov I, Chen P and Madhukar A 1998 *Phys. Rev. B* **58** R10 151
- [36] Heitz R, Mukhametzhanov I, Zeng J, Chen P, Madhukar A and Bimberg D 1999 *Superlatt. Microstruct.* **25** 97
- [37] Stier O, Grundmann M and Bimberg D 1999 *Phys. Rev. B* **59** 5688
- [38] Ledentsov N N *et al* 1996 *Phys. Rev. B* **54** 8743
- [39] Solomon G S, Trezza J A, Marshall A F and Harris J S 1996 *Phys. Rev. Lett.* **76** 952
- [40] Grundmann M *et al* 1996 *Proc. 8th Int. Conf. on InP and Related Materials (Schwäbisch Gemünd)* p 738
- [41] Heitz R, Ramachandran T R, Kalburge A, Xie Q, Mukhametzhanov I, Chen P and Madhukar A 1997 *Phys. Rev. Lett.* **78** 4071
- [42] Kagan C R, Murray C B, Nirmal M and Bawendi M G 1996 *Phys. Rev. Lett.* **76** 1517
- [43] Ohnesorge B, Albrecht M, Oshinowo J, Forchel A and Arakawa Y 1996 *Phys. Rev. B* **54** 11 532
- [44] Grundmann M *et al* 1995 *Phys. Rev. Lett.* **74** 4043
- [45] Ota K, Usami N and Shiraki Y 1998 *Physica E* **2** 573
- [46] Di Bartolo B 1968 *Optical Interactions in Solids* (New York: Wiley)
- [47] Varshni Y P 1967 *Physica* **34** 149
- [48] Yang W, Lowe-Webb R R, Lee H, Sercel P C 1997 *Phys. Rev. B* **56** 13 314
- [49] Grundmann M, Ledentsov N N, Stier O, Böhner J, Bimberg D, Ustinov V M, Kop'ev P S and Alferov Zh I *Phys. Rev. B* **53** R10 509
- [50] Mukai K, Ohtsuka N, Shoji H and Sugawara M 1996 *Appl. Phys. Lett.* **68** 3013
- [51] Raymond S, Fafard S, Poole P J, Wojs A, Hawrylak P, Charbonneau S, Leonard D, Leon R, Petroff P M and Merz J 1996 *Phys. Rev. B* **54** 11 548
- [52] Heitz R *et al* 1997 *Phys. Low-Dim. Struct.* **11/12** 163
- [53] Heitz R, Stier O, Mukhametzanov I, Schliwa A, Madhukar A and Bimberg D to be published
- [54] Heitz R, Mukhametzanov I, Stier O, Madhukar A and Bimberg D to be published
- [55] Schmitt-Rink S, Miller D A B and Chemla D S 1987 *Phys. Rev. B* **35** 8113
- [56] Nomura S and Kobayashi T 1992 *Phys. Rev. B* **45** 1305
- [57] Schmitt-Rink S, Miller D A B and Chemla D S 1987 *Phys. Rev. B* **35** 8113

Chapter 13

Graded and coupled quantum wells for emission of radiation by intersubband emission

K D Maranowski, A C Gossard and K L Campman

Electrical and Computer Engineering Department, University of California, Santa Barbara, CA 93106, USA

Recent advances have been made in both electrical and optical pumping of intersubband transitions in graded and coupled quantum well structures to realize semiconductor sources of radiation in the terahertz frequency range. In this frequency regime, difficulties such as low radiative efficiency and large optical losses have slowed progress. Various approaches to terahertz generation including electrical excitation of intersubband transitions are reviewed. Electrical and optical pumping of parabolically graded quantum wells have yielded 10–100 nW of power at terahertz frequencies. Advances in the understanding of quantum interference effects in intersubband transitions have enabled the design of coupled quantum well structures which have reduced absorption at energies between broadened intersubband transition energies. These interference effects may be useful in designing a new family of intersubband emission devices.

13.1 Introduction

The terahertz frequency regime is rich in physical phenomena and important for its potential applications in atmospheric and gas monitoring, remote sensing and communications. However, to realize this potential, more efficient and compact solid state sources and detectors must be developed. The range 1–10 THz is a crossover regime between electronic and photonic semiconductor devices. Transit times and parasitic time constants limit the high-frequency performance of electronic devices, while low radiative efficiencies at long wavelengths limit

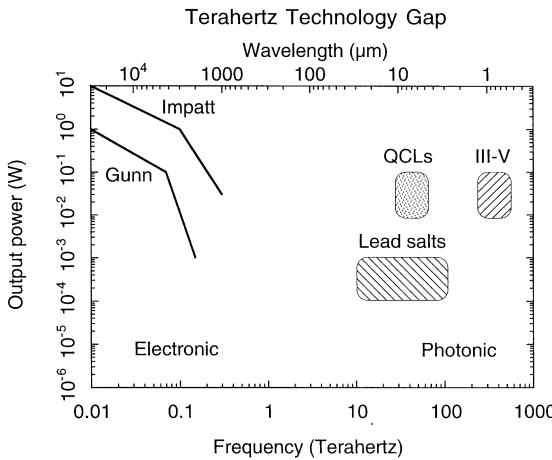


Figure 13.1. Output powers of various classes of semiconductor devices over the frequency spectrum. Approaching the terahertz regime from below, electronic transfer devices roll off due to transit and parasitic time constants. Approaching from above, optoelectronic devices suffer from rapidly decreasing quantum efficiencies at lower frequencies.

the performance of optoelectronic devices. These factors create a gap in viable semiconductor technologies in the range of 1–10 THz as shown in figure 13.1.

Much progress has been made toward filling this gap in the past decade. One successful approach to generating terahertz emission has been to use transitions between electronic subbands in quantum wells. The large electric dipole matrix elements that exist between subbands in quasi-two-dimensional electron systems make them attractive for resonant excitation and emission. While the oscillator strength of an intersubband transition can be large, the efficiency of the emission process is controlled by the relative rates for photon emission and phonon emission. The quantum efficiency, η , for light emission is given by the ratio of the non-radiative lifetime to the total lifetime

$$\eta = \frac{\tau_{\text{nr}}}{\tau_{\text{nr}} + \tau_{\text{rad}}} \quad (13.1)$$

where τ_{rad} is the radiative lifetime and τ_{nr} is the non-radiative lifetime. For a two-level system represented as a classical electron oscillator, the radiative lifetime is given by

$$\tau_{\text{rad}} = \frac{3\epsilon_0 m^* c \lambda_o^2}{2n\pi e^2} \quad (13.2)$$

where ϵ_0 is the permittivity of free space, m^* is the effective electron mass, c is the speed of light, λ_o is the wavelength, n is the refractive index and e is the electronic charge. The λ^2 dependence of the radiative lifetime poses a fundamental difficulty

in engineering long-wavelength emission sources. This lifetime is of the order of ten microseconds for far-infrared wavelengths, while the competing non-radiative processes are much faster. For energies greater than the optical phonon energy (36 meV in GaAs), non-radiative relaxation via optical phonon emission is very fast (≤ 1 ps). For transition energies below 36 meV, optical phonon emission is forbidden and acoustic phonon emission becomes the dominant relaxation mechanism. This non-radiative lifetime varies between $\approx 0.1\text{--}1$ ns depending on the energy and temperature of the electron system [1].

Another design consideration for intersubband transition devices involves coupling light out normal to the sample surface. The selection rules for an intersubband transition in a two-dimensional quantum well require that the transition can interact only with light which has an electric field polarized perpendicular to the two-dimensional confining potential, that is, perpendicular to the epitaxial layers. (Physically speaking, the electric field must point perpendicular to the walls of the quantum well to cause charge to slosh back and forth in the well.) Light with this polarization will of course travel parallel to the sample layers. By placing a metallic grating on the surface of the sample, this light can be coupled out normal to the surface since the near field of the grating allows modes with the electric field polarized perpendicular to the plane of the grating. The optimal coupling to the grating is achieved when the period of the grating is less than, but approximately equal to, the wavelength of the radiation inside the sample [2, 3].

13.2 Review of intersubband emission

For an intersubband transition to radiate terahertz light, the electron distribution must, of course, be heated relative to the surrounding lattice temperature. In other words, a non-equilibrium carrier distribution must be created. One obvious method for heating the electron population is by applying an electric field in the plane of the two-dimensional electron gas (2DEG). The first to use this approach in 1976, Gornik and Tsui [4] observed intersubband emission from the 2DEG formed at the surface (the inversion layer) in a Si-MOS (metal-oxide-semiconductor) structure. The electron density and the subband separation were both tunable with the gate voltage, while voltage pulses applied to source and drain contacts provided for heating the electrons. The detected emission signal versus gate voltage for an in-plane electric field of 20 V cm^{-1} is shown in figure 13.2. The dips in the spectra at $100\text{ }\mu\text{m}$ and $76\text{ }\mu\text{m}$ are characteristic of the Ge:As filter and the crystal quartz filter, respectively, and demonstrate the tunability of the emission with gate voltage.

Other early techniques for generating far-infrared light involved transitions between Landau levels [5]. This approach is most notable for its culmination in the p-Ge far-infrared laser. The first experimental observations of radiative transitions between Landau levels were made in bulk n-InSb samples [6, 7]. In

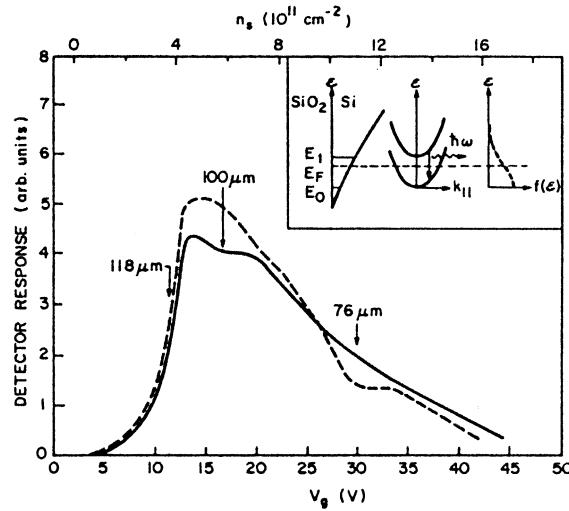


Figure 13.2. Detector response versus gate voltage with an electric field of 20 V cm^{-1} applied along the inversion layer of the Si-MOSFET at 4.2 K. The full curve was taken with a 2 mm thick Ge:As filter, and the dashed curve with a 4 mm thick crystal quartz filter. The inset illustrates the energy structure of the inversion layer and the E_1-E_0 radiative transition. $f(\epsilon)$ is the electron distribution function. From [4].

these experiments, electric fields of several V cm^{-1} were applied to sample platelets, causing electron heating to populate higher Landau levels, and power levels of the order of 10^{-8} W were observed. These initial investigations stimulated many other emission studies for investigating hot carrier distribution functions and band structure in magnetic fields, ultimately leading to the development of the p-Ge laser [8, 9]. The p-Ge laser relies on crossed electric and magnetic fields as well as intersubband scattering between light and heavy hole subbands to create an inverted hot hole distribution. The emission wavelength is tunable with the applied magnetic field as seen in figure 13.3.

The first electrically pumped intersubband emission source in the GaAs/AlGaAs system was reported by Helm *et al* [2, 10]. These sources consisted of GaAs/AlGaAs superlattices with wide wells ($350\text{--}400 \text{ \AA}$) designed to have several energy levels below the optical phonon energy. Initially, excitation was achieved with an electric field in the plane of the superlattice layers by applying voltage pulses to ohmic contacts at the edges of a $4 \times 4 \text{ mm}^2$ sample [2]. The surface area between the ohmic contacts was covered with a metallic grating to achieve surface emission. The sample temperature was kept between 5 and 10 K, and the emission was detected with a Si bolometer at 1.5 K after passing through an InSb filter which was tunable with a magnetic field. The InSb filter acts by absorbing light at its cyclotron resonance frequency, and so, the detected

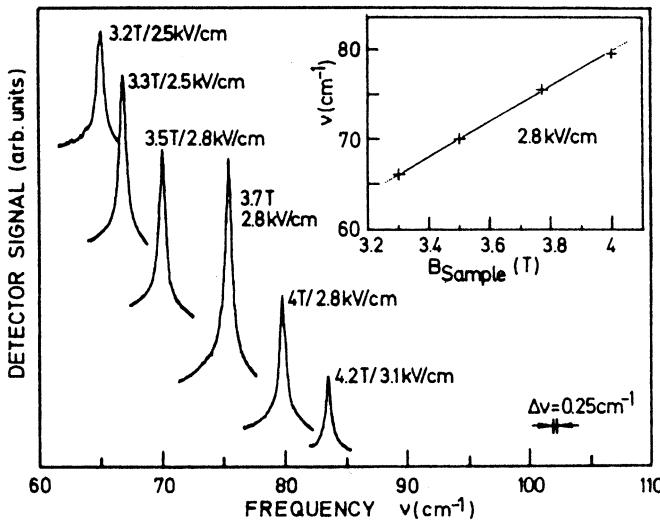


Figure 13.3. Spectra of stimulated emission from a p-Ge laser sample recorded with a GaAs detector. Each spectrum consists of a single laser line which is linearly tunable with magnetic field (inset). From [9].

signal at the bolometer has a minimum when the InSb cyclotron resonance frequency equals the peak frequency of the emitted radiation. The results of this experiment are shown in figure 13.4. The 2p–1s shallow donor transition (from the background concentration of $1 \times 10^{15} \text{ cm}^{-3}$) is seen for samples with and without gratings. For the sample with the $30 \mu\text{m}$ grating (curve (b)), the lowest intersubband transition is also clearly observed.

Shortly after, Helm *et al* achieved stronger emission from multiple intersubband transitions by using a vertical resonant injection scheme [10]. In these experiments, the superlattice was sandwiched between n+ top and bottom contact layers, and voltage was applied normal to the epitaxial layers allowing resonant tunnelling from the lowest energy level of one quantum well into higher energy levels of an adjacent well. Strong negative differential resistances (NDR) were observed in the current–voltage characteristics, confirming the resonant tunnelling nature of the vertical transport. Emission from as many as three separate intersubband transitions was observed at the highest biases as shown in figure 13.5. Since the fourth energy level is above the optical phonon energy, the 4–3 transition has a much lower quantum efficiency and the emission from this transition is very weak. The power ratio of the 3–2 and 2–1 transitions gives an estimate of the population ratio. This ratio, n_3/n_2 , is calculated to be 0.23 ± 0.05 , corresponding to an intersubband temperature of $140 \pm 20 \text{ K}$ [10]. Although a population inversion was not achieved, this experiment was the first to use resonant tunnelling to create a hot carrier distribution capable of emitting

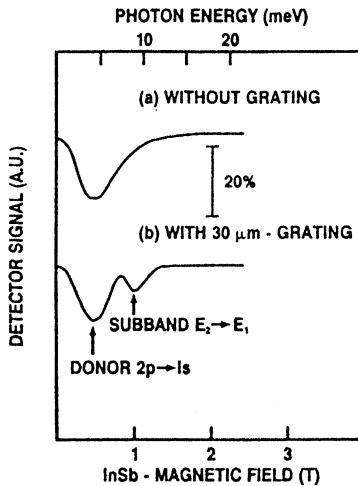


Figure 13.4. Detector signal versus magnetic field of InSb-filter and corresponding photon energy. The 2p–1s donor transition and E_2 – E_1 subband transition are indicated. From [2].

radiation, and therefore represented a fundamental step towards an intersubband laser.

Five years later, using the InGaAs/AlInAs system lattice matched to InP, Faist *et al* reported the first intersubband laser, dubbed the quantum cascade laser (QCL) [11]. While the first QCL operated at a wavelength of 4.2 μm , this range has since been extended beyond 11 μm [12], demonstrating the powerful flexibility of the design. The basic QCL lasing scheme is a three level system with the laser transition occurring between the upper two levels as shown in figure 13.6. The necessary inversion of the upper levels is achieved by designing levels 1 and 2 to be resonant with the optical phonon energy. The lifetime of state 2 is then very short (≈ 0.6 ps) and state 2 is efficiently depopulated. Fast tunnelling out of state 1 and into state 3 are of course also necessary, and these processes are aided by the superlattice between successive active regions which acts as an electron Bragg reflector [12].

The QCL represented another major breakthrough in intersubband emission sources, but it is unclear whether the QCL design can be successfully applied in the terahertz regime. It becomes increasingly difficult to invert a system as the lasing transition energy decreases, particularly if tunnel injection is to be used. Large optical losses compound the problem. An encouraging result was the observation of spontaneous emission at a frequency of ≈ 6 THz from a quantum cascade type structure by Xu *et al* [3]. Optical pumping of similar structures may also offer a solution [13]. Very recently, electroluminescence has been observed at 88 μm wavelength in another quantum cascade type structure [14].

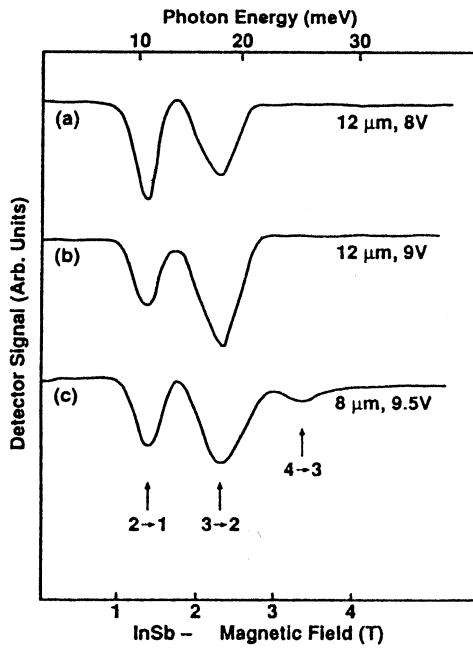


Figure 13.5. Detector signal versus magnetic field of InSb filter and corresponding photon energy for samples with different grating periods at bias voltages as indicated. The observed intersubband transition are marked. The peak absorption is about 20%; the magnitude of the signal in curve (b) has been reduced by a factor of 2.5. From [10].

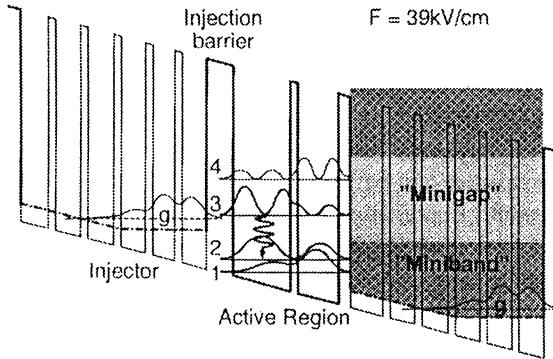


Figure 13.6. Schematic conduction band diagram of a portion of a quantum cascade laser (QCL) structure (one active region bordered by two injector regions), under positive bias corresponding to laser threshold. From [12].

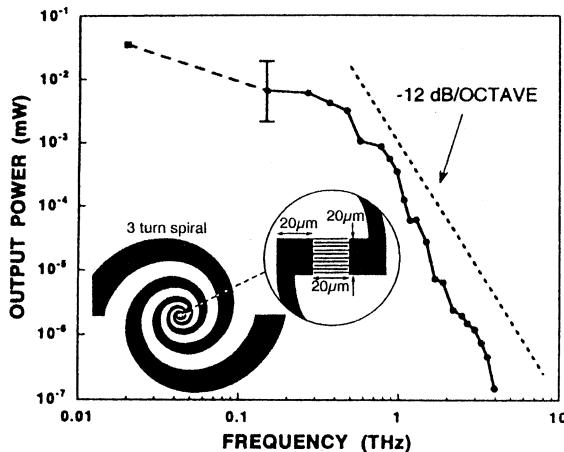


Figure 13.7. Experimental output power from a $1.8 \mu\text{m}$ gap interdigitated electrode photomixer. The inset shows the three-turn spiral antenna and an expanded view of the interdigitated fingers at the driving point of the antenna. From [16].

Another recent advance in terahertz generation is photomixing in low-temperature grown GaAs (LTG-GaAs) [15, 16]. This technique employs two visible or near-infrared lasers which are focused onto a $20 \mu\text{m} \times 20 \mu\text{m}$ area containing interdigitated fingers $0.2 \mu\text{m}$ wide with $1.8 \mu\text{m}$ gaps between. The difference frequency of the two lasers (which is tunable) causes a beating effect in the absorption in the LTG-GaAs epilayer. A large voltage ($\sim 30 \text{ V}$) is applied between the fingers, and the photogenerated carriers are swept to the oppositely biased fingers. This is essentially an electronic transport device modulated by absorption where the modulated voltage is radiated by a spiral antenna attached to the interdigitated fingers. LTG-GaAs is advantageous for this application since it has a high electrical breakdown field ($E_B > 5 \times 10^5 \text{ V cm}^{-1}$), and the extremely short electron-hole photocarrier lifetime in LTG-GaAs ($\tau_{\text{eh}} = 0.27 \text{ ps}$ [15]) allows this mixing process to continue into the terahertz regime (see figure 13.7). With these devices, high output powers are achieved but roll off relatively quickly above 1 THz.

13.3 Emission from parabolic quantum wells

Parabolically graded quantum well (PQW) structures form high-quality electron resonators with resonant frequencies in the terahertz range. One of the most interesting features of parabolically confined potentials is that they will absorb long-wavelength light only at the bare harmonic oscillator frequency, ω_0 , independent of the number of electrons in the well (i.e. independent of the electron-electron interaction [17]). This can be considered as a generalization

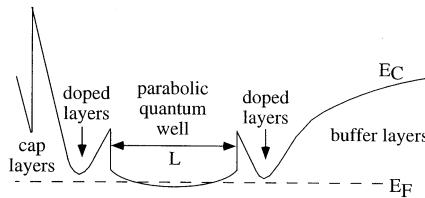


Figure 13.8. Schematic band diagram of the general structure for both PQW emission samples. The growth direction is to the left.

of Kohn's theorem which states that cyclotron resonance absorption is unaffected by electron-electron interactions [18]. The bare harmonic oscillator frequency for PQWs is given by

$$\omega_0 = \sqrt{\frac{8\Delta}{L^2 m^*}} \quad (13.3)$$

where Δ is the energetic depth of the well, L is the well width, and m^* is the effective mass. The generalized Kohn's theorem has been confirmed for far-infrared (FIR) absorption in PQWs [19]. Here we will show that, in agreement with the generalized Kohn's theorem, PQWs will emit FIR radiation at the designed harmonic oscillator frequency when the electron distribution is heated.

In this study, we have used two approaches to heating the electron distribution. The first approach is simply to apply a voltage in the plane of the electron gas via ohmic contacts to the gas. This in-plane excitation relies on intersubband scattering to heat the electron distribution. A potentially more efficient approach is injection of electrons directly into the higher energy levels of the well by using a vertical (i.e. normal to the epitaxial layers) injection scheme. This approach can possibly produce higher powers and efficiencies since the injected hot electrons can decay radiatively through several levels. However, the resonant tunnelling process itself may be hindered by the 'bottom-heavy' parabolic well shape. In other words, the barriers between wells are effectively thicker for tunnelling out of the lowest energy states.

For the in-plane excitation experiments, two samples were grown by molecular beam epitaxy on semi-insulating substrates. A schematic band diagram of the general structure used for each sample is shown in figure 13.8. The layer structure for the 2000 Å PQW sample is as follows: 2000 Å $Al_{0.3}Ga_{0.7}As$ buffer / 100 period smoothing superlattice (14 Å GaAs / 6 Å AlAs) / 4 delta doping sheets (total $8 \times 10^{11} \text{ cm}^{-2}$), each separated by 10 Å of $Al_{0.3}Ga_{0.7}As$ / 100 Å $Al_{0.3}Ga_{0.7}As$ spacer / 2000 Å PQW with $x = 0.2$ at the edges / 100 Å $Al_{0.3}Ga_{0.7}As$ spacer / 4 delta doping sheets ($5 \times 10^{11} \text{ cm}^{-2}$ total) each separated by 10 Å $Al_{0.3}Ga_{0.7}As$ / 17 delta doping sheets ($1 \times 10^{12} \text{ cm}^{-2}$ total) each separated by 10 Å $Al_{0.3}Ga_{0.7}As$ / 800 Å $Al_{0.3}Ga_{0.7}As$ / 100 Å GaAs cap. The layer structure for the 760 Å PQW sample is as follows: 2000 Å GaAs buffer / 300 period smoothing superlattice (14 Å GaAs / 6 Å AlAs) / one delta doping sheet of

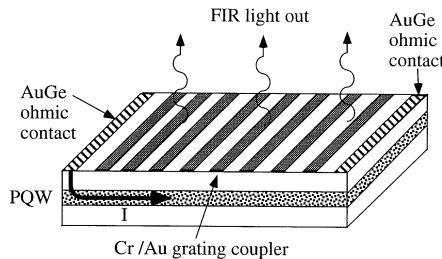


Figure 13.9. Device geometry for parabolic quantum well (PQW) emitter structures. The radiating area is approximately $5 \times 5 \text{ mm}^2$.

$5 \times 10^{11} \text{ cm}^{-2}$ / 200 Å Al_{0.3}Ga_{0.7}As spacer / 760 Å PQW ($x = 0.1$ at the edges) / 200 Å Al_{0.3}Ga_{0.7}As spacer / 2 delta doping sheets ($5 \times 10^{11} \text{ cm}^{-2}$ total) separated by 20 Å Al_{0.3}Ga_{0.7}As / 26 delta doping sheets ($5.2 \times 10^{12} \text{ cm}^{-2}$ total) separated by 20 Å Al_{0.3}Ga_{0.7}As / 100 Å GaAs cap. Each parabolic well was grown using the digital alloy technique [20] with 20 Å periods.

Standard Hall measurements were performed on both wells in the van der Pauw configuration. The low-temperature (10 K) carrier concentration and mobility for the 2000 Å well are $2.7 \times 10^{11} \text{ cm}^{-2}$ and $120\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively, and for the 760 Å well they are $3.9 \times 10^{11} \text{ cm}^{-2}$ and $170\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

The final device geometry is shown in figure 13.9. Standard AuGe ohmic contacts were made as large stripes 5 mm apart. Then a metallic grating (50 Å Cr / 2000 Å Au) was deposited parallel to the ohmic contact stripes to act as the grating coupler. (The grating coupler is necessary to couple light out normal to the surface, as discussed previously. Also, edge emission is much weaker due to self absorption.) Following Helm *et al* [2], who found that the optimal coupling was achieved when the grating period was slightly smaller ($\approx 10\%$) than the wavelength of emitted radiation in air, we used 50 μm and 20 μm grating periods for the 2000 Å and 760 Å wells, respectively. The duty cycle was 50% for each grating.

To measure the FIR radiation from the PQWs, the emitter was placed at one end of a metallic light pipe with a lightly doped ($6 \times 10^{13} \text{ cm}^{-3}$) n-InSb photoconductive detector at the other end. The entire light pipe is immersed in liquid helium to obtain 4 K background conditions. The emitter and detector are separated by 20 cm, and the detector sits in a superconducting magnet. The n-InSb photoconductive detector is biased with a constant current and operated in the cyclotron resonance mode. The spectral width of the cyclotron resonance photoconductive response [21] is about 3 cm^{-1} . By tuning the magnetic field, the spectral sensitivity can be changed according to the cyclotron resonance frequency $\omega_c = eB/m^*$ (where B is the magnetic field and e is the electronic charge) resulting in a tuning constant of about $65 \text{ cm}^{-1} \text{ T}^{-1}$. The emission spectra

data are actually a convolution of the true emission spectra and the detector response. However, the linewidth of the detector response (about 3 cm^{-1}) is narrower than the linewidth of the emitted radiation from the PQWs (see later), so the linewidth of observed spectra is limited by the true emission linewidth.

[Figure 13.10\(a\)](#) shows the emission spectra obtained for the 2000 \AA PQW. The peak frequency of the FIR emission is approximately 52 cm^{-1} , in good agreement with the designed harmonic oscillator frequency of 47 cm^{-1} . Also, the peak emission frequency does not change as the input power is varied. The linewidth also remains essentially constant at 18 cm^{-1} for various input powers. The peak intensity increases approximately linearly as the electric field is increased up to 8 V cm^{-1} , and above 10 V cm^{-1} the emission intensity saturates.

To estimate the output power from the 2000 \AA PQW, we compare its emission with the cyclotron resonance emission at 0.8 T (i.e. $\approx 52\text{ cm}^{-1}$) from a lightly doped ($6 \times 10^{13}\text{ cm}^{-3}$) n-InSb emitter for which the emission power is known [5]. This comparison yields an order of magnitude estimate of $7 \times 10^{-9}\text{ W}$ emitted from the 2000 \AA PQW sample at an electric field of 8 V cm^{-1} . This result can be compared with a rough theoretical estimate based on the intersubband lifetimes. The electrical input power P_{in} is simply multiplied by the coupling efficiency η_c (which includes the grating coupling efficiency and the collection efficiency) and the ratio of the non-radiative lifetime, τ_{nr} , to the radiative lifetime, τ_{rad} (for $\tau_{\text{nr}} \ll \tau_{\text{rad}}$).

$$P_{\text{out}} = P_{\text{in}}\eta_c \frac{\tau_{\text{nr}}}{\tau_{\text{rad}}} \quad (13.4)$$

The non-radiative intersubband lifetime for energies below the optical phonon energy was measured to be $\approx 1\text{ ns}$ in a coupled quantum well structure by Heyman *et al* [1]. The radiative lifetime for a classical electron oscillator is given by

$$\tau_{\text{rad}} = \frac{3\epsilon_0 m^* c \lambda_o^2}{2n\pi e^2} \quad (13.5)$$

where ϵ_0 is the permittivity of free space, c is the speed of light, n is the refractive index, and λ_o is the wavelength. At $\lambda_o = 200\text{ }\mu\text{m}$, $\tau_{\text{rad}} = 3 \times 10^{-5}\text{ s}$. The average electrical input power is 40 mW and assuming a coupling efficiency of 5% , $P_{\text{out}} = 6 \times 10^{-8}\text{ W}$ which is within an order of magnitude of the experimental result. We estimate the temperature of the electron distribution to be $\approx 30\text{ K}$ from the data of Hirakawa *et al* [22] showing the electron temperature as a function of power input per electron for a two-dimensional system.

The emission spectra for the 760 \AA PQW are shown in figure 13.10(b). Again, the peak emission frequency at 104 cm^{-1} agrees well with the designed frequency of 97 cm^{-1} . (The increased noise in figure 13.10(b) is due solely to the larger noise of the photoconductive detector at higher magnetic fields.) As in the case of the 2000 \AA PQW, the linewidth for each spectrum is $\approx 18\text{ cm}^{-1}$.

To improve device performance, we have begun to explore the possibility of using a vertical injection scheme. As an initial test to understand the feasibility of resonant tunnelling injection into wide PQWs, we grew simple resonant

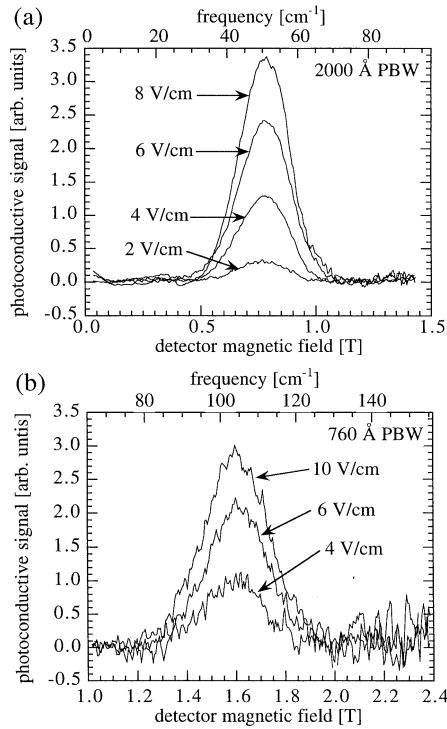


Figure 13.10. Emission spectra for different in-plane electric fields for the (a) 2000 Å PQW and the (b) 760 Å PQW. The frequency axis (top abscissa) is simply the detector magnetic field multiplied by the cyclotron resonance tuning of the photoconductive detector ($65 \text{ cm}^{-1} \text{ T}^{-1}$).

tunnelling diode (RTD) structures with a parabolically graded quantum well. This type of structure was first studied by Sen *et al* for a narrow 435 Å parabolic quantum well [23]. Three samples were grown with 435 Å, 1000 Å and 2000 Å PQWs. Each sample had 35 Å AlAs barriers and n+ GaAs contact layers above and below the wells.

The current–voltage characteristics at low temperatures (10 K) from the 435 Å RTD were essentially identical to those measured by Sen *et al* [23] showing at least ten negative differential resistances (NDR). These NDRs are evenly spaced as would be expected from the evenly spaced energy level ladder for a parabolic potential. The current–voltage curves for the 1000 Å and 2000 Å well RTDs are shown in figures 13.11(a) and (b), respectively. For the 1000 Å well, there are no NDRs, but evenly spaced resonances are still clearly observed in the derivative of the current with respect to voltage. This indicates that there continues to be coherent tunnelling through the quantized energy levels. In the

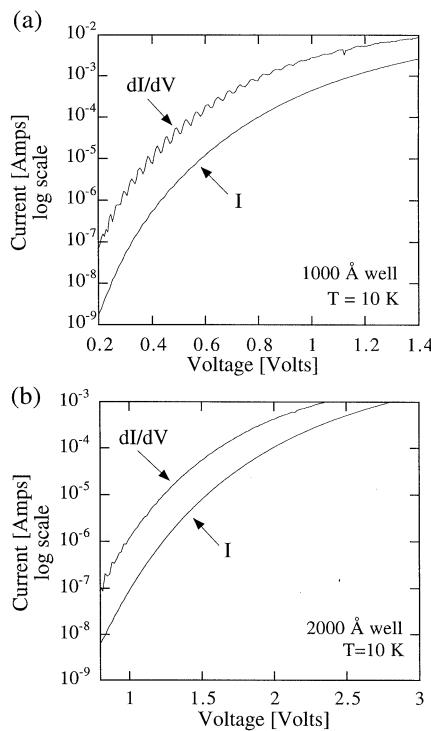


Figure 13.11. Current–voltage characteristics and current derivatives for (a) 1000 Å well resonant tunnelling diode (RTD) and (b) 2000 Å well RTD.

case of the 2000 Å well RTD, no clear resonances are observed in either the current or its derivative since the coherent tunnelling is lost because of scattering across the large width of the well.

With these design considerations, a vertical injection structure was grown consisting of ten 1000 Å PQWs with 70 Å Al_{0.3}Ga_{0.7}As barriers between the wells. To ensure adequate conductivity of the sample, a small amount of doping ($5 \times 10^{10} \text{ cm}^{-2}$) was placed in the middle of each barrier. Heavily doped n+ layers were placed at each end of the structure for the top and bottom contacts. The emission device used a 20 μm period AuGeNi grating which also functioned as the top contact. The n+ contact layer was etched out between the fingers of the grating to reduce free carrier absorption.

The emission spectrum for this vertical injection structure at a bias of 8 V is shown in figure 13.12. The designed frequency of the 1000 Å PQWs was 100 cm^{-1} , and the peak in the spectrum agrees fairly well with this value. This spectrum is about twice as broad as the spectra for in-plane excitation. This additional broadening is most likely caused by thickness or alloy variations

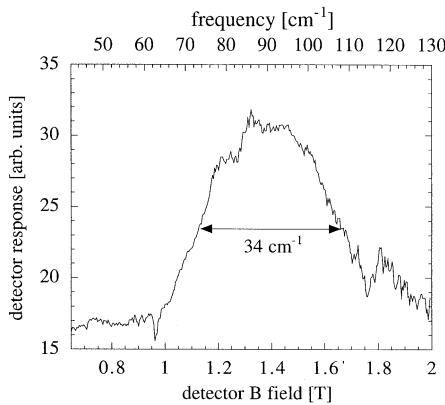


Figure 13.12. Emission spectrum from the 10 period PQW vertical injection structure at an applied bias of 8 V.

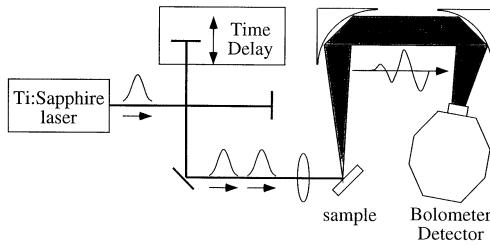


Figure 13.13. The experimental setup for measuring the picosecond response of samples with femtosecond laser pulse excitation. An autocorrelation function showing several cycles of terahertz emission is generated from the time delay between the two laser pulses. From [26].

across the ten PQWs in the growth direction. The total emitted power from this device is estimated to be 30 nW, based on the sensitivity of the InSb detector and comparison with the in-plane excitation data. This emission power of 30 nW is about a factor of four improvement over the in-plane excitation data. Further experiments and optimization are necessary to determine the efficiency and potential for this type of vertical injection device.

13.4 Optical excitation of parabolic quantum wells

Another approach to generating terahertz emission from PQWs is to use an ultrafast laser pulse as an optical excitation to create a decaying charge oscillation in the PQW which emits a few cycles of terahertz radiation. The general experimental setup for this type of measurement is illustrated in figure 13.13.

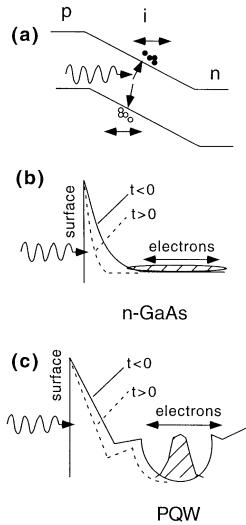


Figure 13.14. Schematic representation of ultrafast laser pulse excitation. In (a) the photogenerated carriers in the intrinsic region of a p-i-n structure undergo plasma oscillations in the built-in electric field. In (b) and (c), photogenerated carriers near the surface screen the surface depletion field (broken curves) and cause oscillation of electrons in the (b) bulk or (c) PQW.

In experiments on p-i-n structures by Sha *et al* [24], terahertz emission results from plasma oscillations of the photogenerated carriers in the built-in field present in the intrinsic region. The frequency of the emitted terahertz radiation is the plasma frequency corresponding to the density of photogenerated carriers (which is controlled by the strength of laser excitation). In bulk n-type GaAs samples, photogenerated carriers can be excited in the surface depletion field. These carriers screen the surface field and subsequently excite the extrinsic electron plasma in the bulk [25, 26]. These different mechanisms are illustrated schematically in figure 13.14. In contrast to the p-i-n case, the frequency of the emitted terahertz radiation from n-GaAs samples matches the plasma frequency corresponding to the extrinsic doping level of the samples and does not depend on the laser excitation strength [26]. This observation confirms that the emission mainly originates from plasma oscillations of cold electrons in the bulk. Time resolved spectra (and their Fourier transforms) for three n-GaAs samples with different doping levels are shown in [figure 13.15](#) (from Kersting *et al* [26]).

When a modulation doped PQW is placed adjacent to a surface absorption layer (shown in figure 13.14(c)), the cold electrons in the PQW are excited by the change in the surface depletion field (which is screened by the photogenerated carriers). [Figure 13.16](#) shows the Terahertz emission from a PQW sample at room temperature for different laser excitation powers [27]. The peak frequency

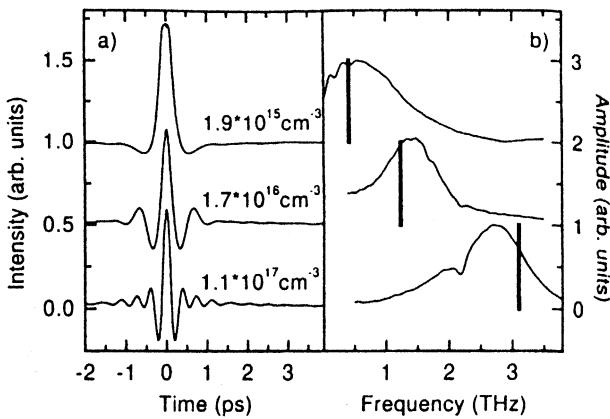


Figure 13.15. (a) Time-resolved correlation data for three n-GaAs samples n-doped at 1.9×10^{15} , 1.7×10^{16} and $1.1 \times 10^{17} \text{ cm}^{-3}$. (b) Fourier spectra. The bars indicate the plasma frequencies calculated from the extrinsic doping levels. The absorption dip at 2.2 THz results from the polyethylene window of the bolometer. From [26].

agrees well with the harmonic oscillator frequency of the parabolic potential. The slight increase in the peak frequency at higher pulse power may be attributed to additional PQW filling by photogenerated carriers.

13.5 Quantum interference effects

We return again to the basic questions about the potential of intersubband transitions for generating radiation at long wavelengths. As discussed previously, intersubband lasers (QCLs) have been achieved utilizing precise growth technology to control energy level spacings and relaxation times accurately and thereby achieve a population inversion. However, an entirely new class of optical devices has recently been proposed based on interference between intersubband transitions in coupled quantum wells. These infrared devices include an intersubband laser that operates without a population inversion [28], and efficient nonlinear optical elements, which rely on a destructive interference between intersubband transitions to sharply reduce the rate of absorption [29]. The most direct application of intersubband interference phenomena would be for semiconductor lasers which would not require a population inversion. In a standard laser system where emission and absorption probabilities for the lasing transition are equal, a population inversion (i.e. more electrons in the excited state than the ground state) is necessary to achieve gain. If it is possible to force an asymmetry in the emission and absorption probabilities by decreasing the likelihood of absorption and increasing the likelihood of emission, a population inversion is no longer necessary for lasing. This idea was first proposed for a four-

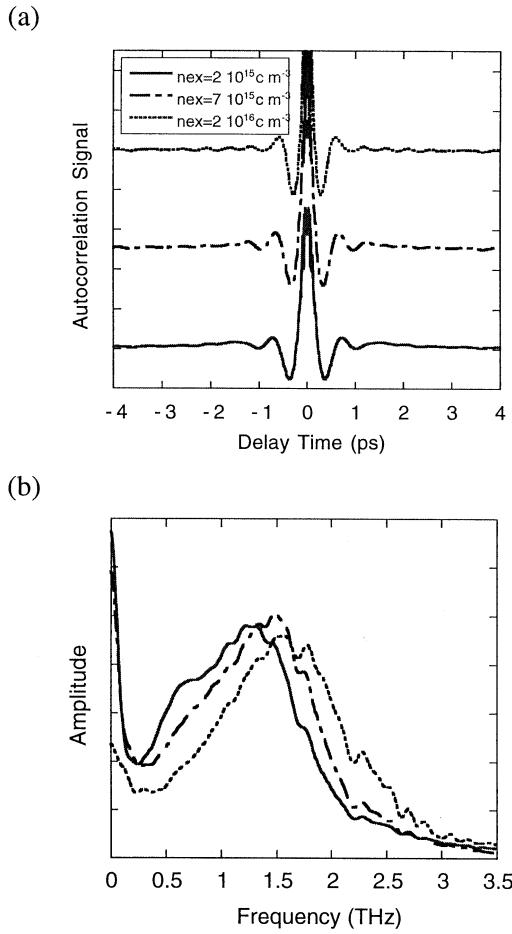


Figure 13.16. (a) Auto-correlation signal and (b) Fourier transformed spectra of the THz emission from the PQW sample at three different laser excitation powers (courtesy of K Unterrainer).

level atomic system [30], and soon after generalized for more readily available and experimentally achievable atomic systems [31]. Initially, the reduction of absorption as a result of destructive interference was observed in atomic systems experimentally [32]. More recently, lasing without inversion has also been observed in these systems [33, 34].

The combination of the research areas of intersubband transitions in semiconductor quantum wells and lasing without inversion in atomic systems was initially proposed by Imamoglu and Ram [28]. Previous attempts to observe amplification without inversion relied on atomic systems with parameters such

as energy levels and decay rates fixed by nature. The proposed semiconductor structure utilizes bandgap engineering to artificially fabricate a three-level lasing system in which the energy levels and decay rates can now be controlled by simply adjusting the quantum well growth parameters. Another advantage of a semiconductor system is that the coupling of the upper levels occurs through coherent resonant tunnelling across a barrier, instead of via an electromagnetic field, removing the need for a coupling laser. However, there is an important disadvantage to the semiconductor scheme. Atomic systems, although inflexible, present a more ideal system with regard to phase randomizing collisions which can destroy interference effects. Dephasing rates can be negligibly small compared to other relevant rates such as the decay rate to the continuum and the frequency of the coupling laser. In contrast, electrons in a semiconductor are embedded in a crystal lattice, and are therefore subject to more scattering events. Scattering times from crystal imperfections and from the emission of optical and acoustic phonons can be quite short, with possible detrimental effects on the interference phenomena. In this section, we seek to understand these scattering and dephasing times, which then enable the design of a resonant tunnelling induced transparency structure.

13.5.1 Absorption in coupled quantum wells

Samples were grown by MBE using the digital alloy technique [20] and measured with a Bruker IFS 66V Fourier Transform Infrared (FTIR) spectrometer. The samples were polished with 45° facets for a multipass geometry, and reference spectra were obtained by either depleting the electrons with a Schottky gate on the sample surface or by comparing TM versus TE polarized spectra (see figure 13.17). In order to understand the coupling strength of the upper two levels, Ω_{23} , a first set of experiments used samples in which the upper levels are bound states, rather than states that are quasi-bound to a continuum on one side via a tunnelling barrier. In the bound state case, there are no quantum interference effects to be included in the analysis, since there is no tunnelling coupling to a continuum.

The necessary three-level system can be created with two asymmetric coupled quantum wells. In a study of intersubband absorption in single quantum wells of different well widths and alloy compositions [35], it was found that the absorption linewidth was consistent with broadening dominated by interface roughness. It was also found that the linewidth was relatively unaffected by alloy composition in the quantum well, for a constant well width. Therefore, rather than using a narrow GaAs well with large interface roughness broadening, an alloyed $\text{Al}_x\text{Ga}_{1-x}\text{As}$ well can be used to create the same energy level configuration.

This design technique was employed in the structure shown in figure 13.18. This sample (sample A) was a multiquantum well stack with n+ contact layers above and below the quantum wells. For performing absorption experiments, the TM versus TE referencing technique described above was used since a Schottky

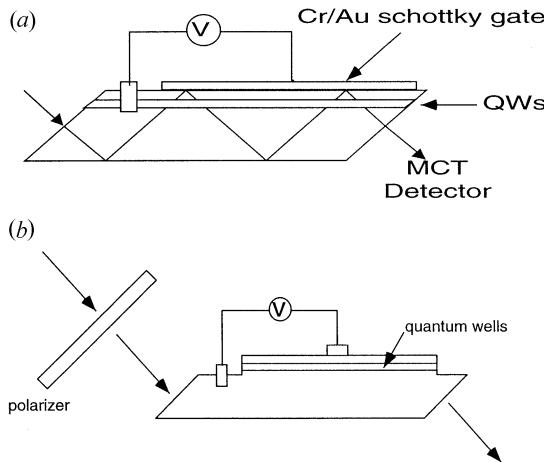


Figure 13.17. Schematic of two techniques (side view) for intersubband absorption experiments using the multipass geometry. In (a), a Schottky gate is deposited on top of the sample, and contact is made to the electron gas so that reference spectra can be taken by depleting carriers from the quantum well. In (b), a TM polarized spectrum is compared to a TE polarized spectrum to isolate the intersubband response.

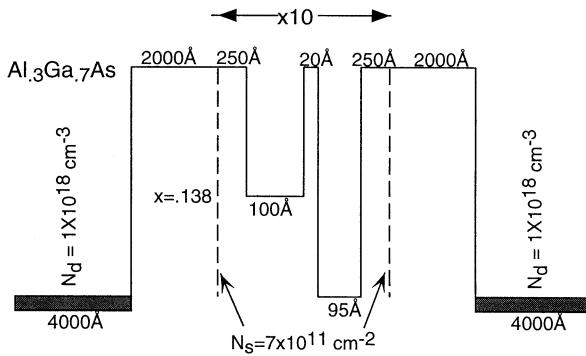


Figure 13.18. Multiquantum well sample structure (sample A) used for observation of coherent coupling of the upper two levels.

gate is incapable of depleting the 10 period quantum well stack. The sample bias was applied between the $n+$ regions, and the measured I-V characteristics were symmetric.

The low-temperature (4.2 K) absorption spectra for this sample are shown in figure 13.19 with the peak positions plotted in figure 13.20. Each scan was taken at a different applied bias (0.25 V steps between adjacent curves) with the bottom curve in the figure corresponding to -3.0 V, just prior to breakdown. In this case,

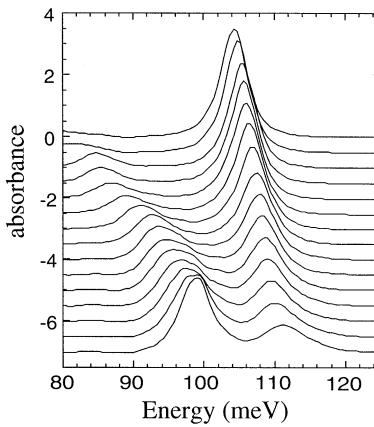


Figure 13.19. Intersubband absorption spectra for the coupled quantum well structure of figure 13.18. The bias ranges from 0.5 V for the top curve to -0.3 V for the bottom curve in 0.25 V steps.

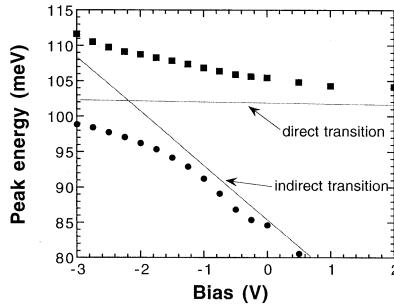


Figure 13.20. Measured absorption peak positions for the coupled quantum well structure of figure 13.18. The expected bias dependence of the spatially direct and spatially indirect transitions are also sketched.

we refer to positive bias as one that lowers the potential on the right, or substrate, side of the structure shown in figure 13.18.

At large positive biases (top of figure 13.19), the level in the left quantum well is much higher than the excited state in the GaAs quantum well. In this case, the coupling of levels is quite weak and only a single, spatially direct transition is observed. With increasing negative bias, the level in the left AlGaAs well is pulled down, closer to resonance, and a weak, spatially indirect, lower energy peak appears. At the anti-crossing when the levels are nicely aligned, the excited states are a superposition of the bare quantum well levels, extended across both the GaAs and AlGaAs quantum wells.

Figure 13.20 shows the measured peak positions as a function of the applied

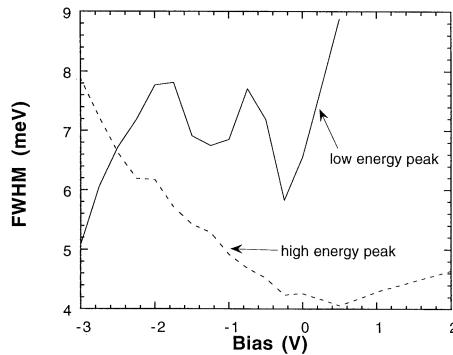


Figure 13.21. Intersubband transition linewidths extracted from fits to the spectra of figure 13.20 using a sum of Lorentzians form. At higher biases than resonance (> -2.5 V), the high-energy peak is the dominant peak corresponding to the direct transition.

bias. Considering only the direct intersubband transition in the GaAs quantum well, we expect a transition energy that remains relatively constant with a change in applied bias, because the carriers in the well can effectively screen the electric field. Since there are no carriers in the AlGaAs well, we expect that this energy level will move linearly with applied bias as sketched in figure 13.20. These sketched lines show the expected peak positions in the absence of coupling of the upper levels, which would be the case for strongly localized states. The interaction between the upper levels leads to a peak repulsion, or ‘anti-crossing’ of states, with a minimum peak splitting at resonance. This minimum splitting is the quantity Ω_{23} and is found to be 12 meV for this structure.

To obtain an estimate for the size of the interference effects that can be observed in double quantum wells coupled to a continuum, it is important to understand the additional scattering or inhomogeneities resulting from coupling the quantum wells. Figure 13.21 shows the measured intersubband transition linewidths for both the high- and low-energy peaks. This data was extracted by fitting the spectra to a sum of Lorentzians form. (Again, since the states are bound, there is no interference to be included.) At positive biases, only a single high-energy peak is observed corresponding to the direct transition. The linewidth for this peak is about 4 meV, about one meV larger than would be expected for a single 95 Å GaAs quantum well based on the measurements presented in [35]. Since there is substantial penetration of the wavefunction into the left well even out of resonance, it is reasonable to expect additional scattering from the added interfaces in the other well, alloy scattering in the barrier, or additional impurity incorporation. For the indirect transition far out of resonance (near zero bias), we find a much broader line, of the order of 6–8 meV. (The data is quite jumpy at this point since it is difficult to get a good fit for the small peaks, especially with the wavelength-dependent background of the polarizer.) It is this number

(the 6–8 meV indirect transition linewidth) that can be used as an estimate of the dephasing rate in the left well, γ_{31} , in a coupled quantum well structure because far out of resonance, the low-energy peak linewidth is mostly determined by scattering in the left quantum well. As the levels are brought closer to resonance, the low-energy peak takes on more character of the direct transition causing a narrowing of the lineshape, while the high-energy peak becomes more indirect with the opposite effect on linewidth. At resonance (the bias at which the splitting is minimum), the linewidths of the two transitions are equal which is expected if there are equal contributions to the total wavefunctions from the left and right quantum wells.

13.5.2 Interference of intersubband transitions

In the previous section, by measuring intersubband absorption in a bound double quantum well system, experimental estimates were obtained for two of the three parameters (Ω_{23} and γ_{31}) that enter into the equation for the absorption profile of a double well system coupled to a continuum. No interference effects were observed in the bound systems because the bound intersubband transitions are not lifetime broadened. To observe interference phenomena, it is necessary to couple these wells to a continuum through a thin tunnelling barrier. This coupling will then be responsible for determining the final quantity needed to complete the picture, Γ_2 .

To get an estimate of the tunnelling rates that can be achieved in our structures, we performed transmission matrix calculations [36] to look at the transmission versus energy curves for a single 90 Å GaAs quantum well coupled to a continuum on one side by an Al_{0.3}Ga_{0.7}As tunnelling barrier of varying widths. At the energy of the quasi-bound state, there is a pronounced peak in the transmission versus energy curve. We take the FWHM of this curve to be the tunnelling contribution to the FWHM of the intersubband transition that would be measured in this structure. Results of the calculation are plotted in figure 13.22, showing that with reasonable growth parameters for the barrier, the tunnelling induced energy broadening, Γ_2 , can be made large (FWHM \approx 10–20 meV) compared to the dephasing rates in single quantum wells (FWHM \approx 3–4 meV). For fast tunnelling rates, the transition no longer has a perfect Lorentzian lineshape since tunneling becomes faster at higher energies. Assuming a FWHM of 15 meV, there is about a 7% difference in tunnelling rates between the energies at half of the maximum intensity, corresponding to a small, high-energy tail.

The case of a double quantum well coupled to a continuum was also considered. The structure consists of a 94 Å GaAs quantum well coupled to a 96 Å Al_{0.1}Ga_{0.9}As quantum well by an 18 Å barrier. These wells are then coupled to the continuum by a 20 Å Al_{0.3}Ga_{0.7}As barrier. Transfer matrix calculations of the position and tunnelling width of the upper states were performed as a function of the electric field applied across the device. These are empty well calculations, and do not consider the additional electric fields caused by the electrons and the

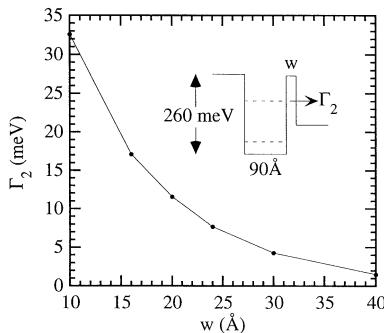


Figure 13.22. Calculation of Γ_2 (the FWHM of the transmission versus energy curve of the tunnelling calculation) for a quasi-bound state in a 90 Å GaAs single quantum well coupled to a continuum on one side by an $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ tunnelling barrier.

remote ionized donors left behind. Since we are dealing with transitions to levels which are high above the bottom of the wells, the effects of the additional fields should be small. However, the calculations were performed on a structure with a lower energy left well than used in the experiments since the absence of the electric field from the carriers across the left well raises the potential in this region substantially.

Results of the calculation for the energy separation of the two upper levels are shown in figure 13.23(a). The peak splitting has a parabolic shape with a minimum of 17.2 meV (at which point the levels are in resonance) near an applied bias of -2.5 kV cm^{-1} . Negative biases correspond to a higher conduction band energy on the left side of the structure. The calculated peak widths are given in figure 13.23(b). At large negative biases, the low-energy peak corresponds to the direct intersubband transition, with the high-energy peak being the indirect transition. Out of resonance, the direct transition has a large linewidth since electrons tunnel quickly out of this level, but the indirect level has a small linewidth because its wavefunction is mostly bound in the left well. As the levels are brought into resonance, the wavefunctions of both levels become distributed across the two wells tending to equalize the linewidths. At the minimum splitting, however, the transitions have linewidths that differ by 1.2 meV (5.2 meV and 6.4 meV). This difference occurs because there is a 17.2 meV difference in energy between the levels, and the tunnelling rate is a strong function of the energy. Note that these resonance linewidths are roughly half of the Γ_2 determined from figure 13.22. We can expect larger differences in linewidth between the two levels at resonance if we increase the splitting, or if we increase the coupling to the continuum. At large positive biases, the direct transition (high-energy peak) is more narrow than the direct peak at negative biases (low-energy peak) because in the latter case, the electric field tends to push the wavefunction towards the left quantum well, away from the continuum, reducing the coupling. The general

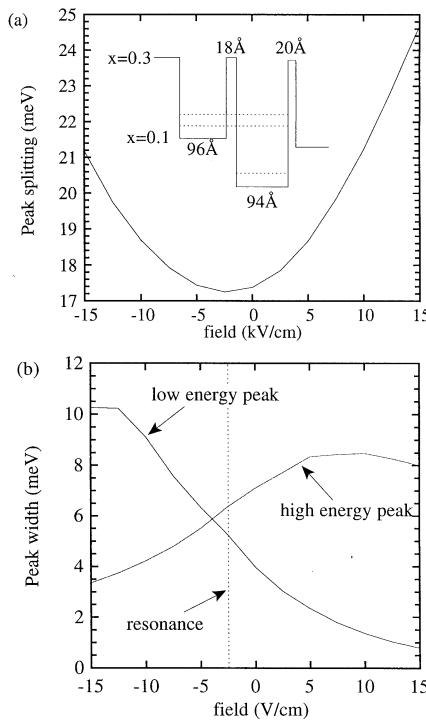


Figure 13.23. (a) Calculated energy splitting between excited states and (b) peak widths as a function of applied electric field across the device. The sample structure is given in the inset of (a). At resonance conditions (broken curve in (b)), the peak widths are not equal because of an energy-dependent tunnelling rate.

trend is to have a high-energy peak with a linewidth that remains relatively flat as a function of gate bias as compared to the low-energy peak.

The above calculations are a useful aid for designing the quantum wells and the tunnelling barrier to the continuum, but there are several additional design issues for creating a successful sample for the observation of interference effects. The double quantum well samples with bound states used in the previous section were simply modulation doped in the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barriers on both sides of the quantum wells. However, if a continuum is to be introduced on the right (substrate) side of the wells, then most of the doping must be in the left (surface side) barrier. This creates a large undesirable electric field across the quantum wells which further splits the two upper levels. This can be remedied by raising the energy level in the left well by either raising the Al mole fraction in the left well or by decreasing the well width, or some combination of the two. Another technique for reducing the built-in field across the well is to put a small amount of doping in the continuum. Here, caution must be used because the Si donor

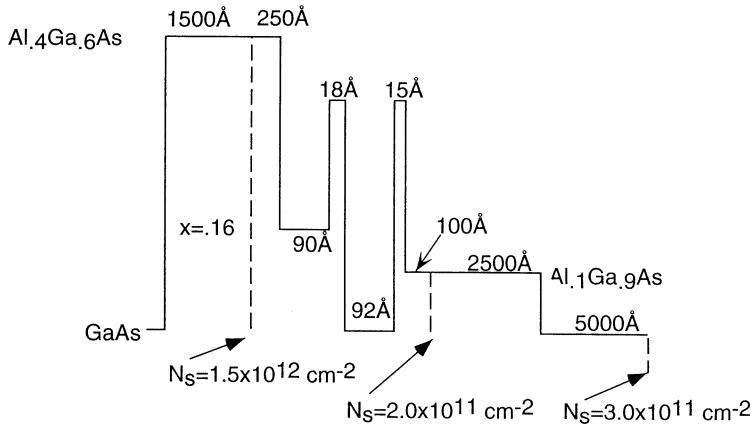


Figure 13.24. Sample structure for a double quantum well coupled to a continuum (sample B).

level is shallow for the continuum composition of $\text{Al}_{0.1}\text{Ga}_{0.9}\text{As}$, and all donors will be ionized even at low temperatures. Therefore, the amount of doping must be kept small ($\approx 2.0 \times 10^{11} \text{ cm}^{-2}$) to ensure that all electrons transfer to the quantum wells. Additional doping is also necessary below the continuum layers to reduce the electric field that would result from the fermi-level pinning in the semi-insulating substrate. A final design consideration is to increase the Al mole fraction in the top barrier from $x = 0.3$ to $x = 0.4$ to prevent any tunnelling of electrons to the triangular quantum well formed by the positively charged ionized delta-doping sheet.

All of these design considerations were used to design and grow sample B whose layer sequence is given in figure 13.24. In designing the sample, the layer parameters were typically chosen so that the upper levels are aligned when the sheet density of electrons in the well is approximately $5 \times 10^{11} \text{ cm}^{-2}$. Low-temperature (10 K) Hall measurements on this structure yield an electron density of $6.1 \times 10^{11} \text{ cm}^{-2}$ with mobility of $162\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

The Schottky gate referencing technique was used to obtain the absorption spectra shown in figure 13.25(a). A negative gate bias is used to deplete electrons from the quantum well, while at the same time raising the energy of the left side of the quantum well relative to the right. Figure 13.25(b) shows the peak positions as a function of the electron concentration showing a clear anti-crossing behaviour. (The carrier concentration values for the abscissa were obtained by comparing the integrated intensity of each spectrum at a given bias with the zero-bias spectrum, which corresponds to a density of $6.1 \times 10^{11} \text{ cm}^{-2}$ obtained from the Hall measurement.)

The measured FWHM of the high- and low-energy peaks are plotted in figure 13.26. Qualitatively, this plot looks similar to the simulation presented

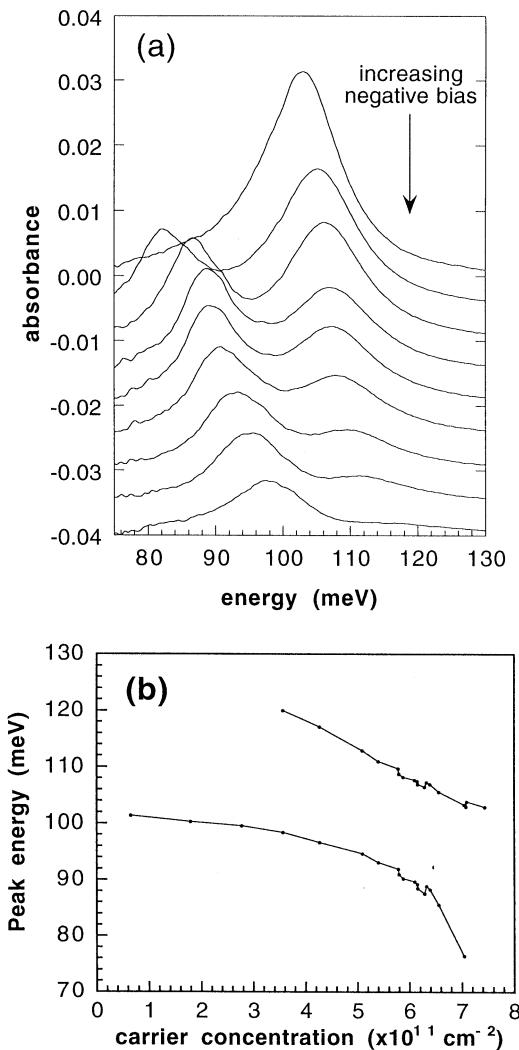


Figure 13.25. (a) Absorption spectra of sample B for different Schottky gate biases. Moving from top to bottom, the gate bias is becoming more negative. (b) Measured peak positions plotted as a function of the electron sheet density in the quantum well.

in figure 13.23(b). The high-energy peak is relatively flat as a function of gate bias (or carrier concentration), whereas the low-energy peak has a much stronger dependence. From this data, we can get estimates of γ_{31} , the dephasing in the left quantum well, and Γ_2 , the rate of tunnelling from the right well to the continuum. Based on the linewidth of the low-energy peak far from resonance, we estimate

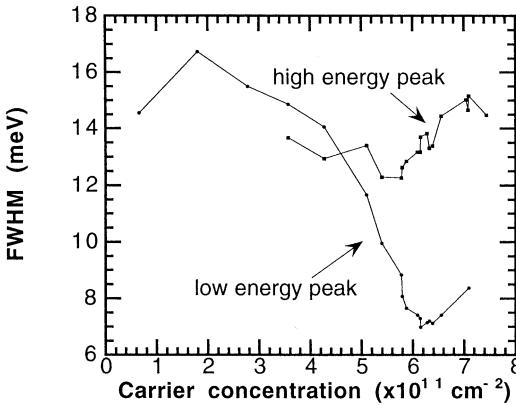


Figure 13.26. Measured FWHM of both the high- and low-energy peaks for sample B as a function of carrier concentration.

that $\gamma_{31} = 7$ meV, and from the high-energy peak far from resonance, we estimate $\Gamma_2 = 14$ meV.

All of the data in figures 13.25(b) and 13.26 on the peak positions and widths were extracted from the measured spectra by fitting the curves to a sum of two Voigt profiles (a convolution of Lorentzian and Gaussian profiles). This type of lineshape is to be expected if there is no interference whatsoever. In general, the Voigt profiles tended to be dominantly Lorentzian (of the order of 70–80% Lorentzian as determined by our curve fitting routine) as would be expected for lifetime broadened lineshapes. For bias conditions far out of resonance, the spectra were fit very well throughout the entire range of energies, but for conditions near resonance, the fits were not good in the energy range between the two transition energies as would be expected for interfering intersubband transitions. Figure 13.27(a) shows one of the absorption spectra near resonance conditions along with its curve fit.

Comparing the minimum absorbance between the transition energies for the curve fit and the measured data, there is a reduction of absorption of almost 16% relative to the sum of Voigt profiles. Figure 13.27(b) shows the reduction of absorption determined in the above manner as a function of the carrier concentration. In addition, the splitting between the intersubband transition energies is plotted to show the range of biases over which the upper states are in resonance. The destructive interference can only occur when the upper states are strongly coupled, and indeed the measured reduction in absorption falls to zero away from resonance. We can also estimate the coupling strength, Ω_{23} , to be nearly 18 meV for this structure based on the measured splitting shown in figure 13.27(b).

The key to increasing the size of the interference effect is reducing the dephasing parameter, γ_{31} , which reflects the dephasing in the left quantum well.

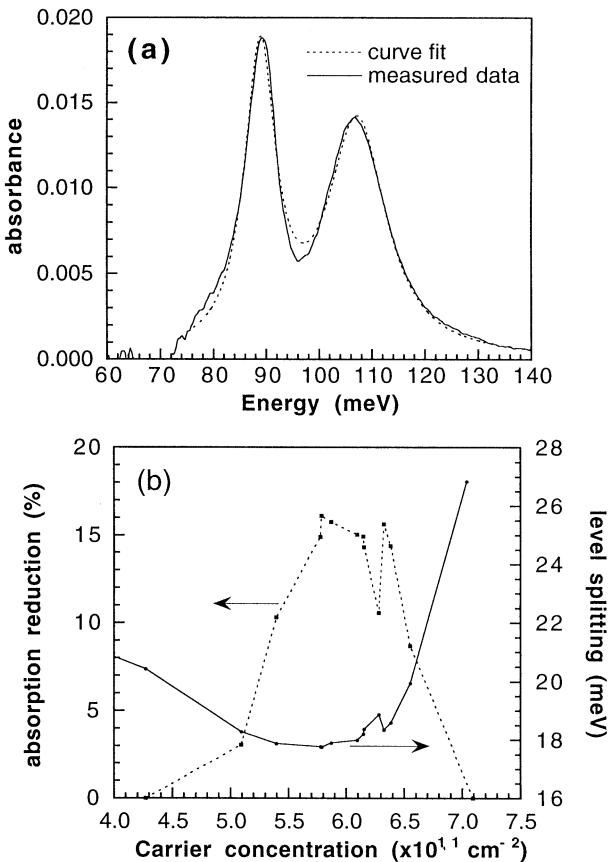


Figure 13.27. (a) Absorption spectrum for sample B near resonant biasing conditions along with a curve fit to the data. The absorption is reduced at energies between transitions relative to the sum of Voigt profiles fit. (b) Plot of the reduction of absorption for the measured data in figure 13.26(a) as compared to the sum of Voigt profiles fits. Also plotted is the measured splitting between the upper levels, which is at a minimum when the upper levels are closely aligned.

With interface roughness scattering being identified as the dominant dephasing mechanism in quantum wells, the next natural step towards reducing dephasing in the left quantum well would be to make the well wider than the 90 Å well used in sample B above. This can, however, present a problem since the strong electric fields across the left well create a triangular potential, instead of a square well potential. For wider wells, the ground state in the left well does not lie above the triangular potential, making the effective barrier for tunnelling much wider. Much wider wells can be used if the Al content in the left well is graded,

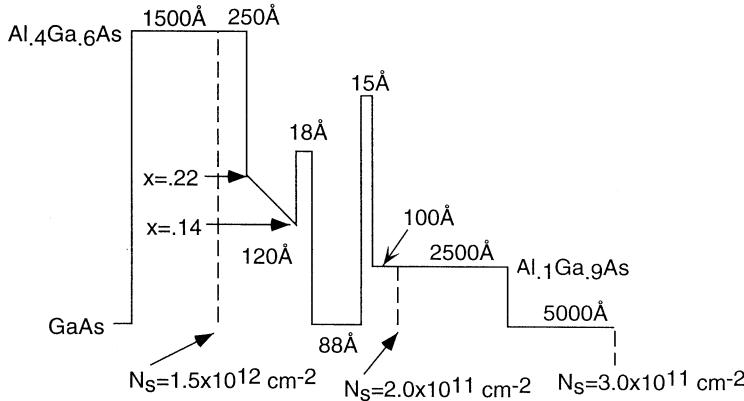


Figure 13.28. Sample design for sample C, including a graded, left quantum well design to offset the electric field between the carriers and the ionized dopants.

to offset the electric field from the dopants. In this way, the self-consistent band profile appears almost flat, even though there is a large electric field across the well ($\approx 50 \text{ kV cm}^{-1}$).

Figure 13.28 depicts the layer design for sample C, showing the grading of the left well. This grading has been done using the digital alloy technique with a 12 Å period. Hall measurements at 12 K on this sample give a sheet carrier density of $6.2 \times 10^{11} \text{ cm}^{-2}$ with a mobility of $157\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. A self-consistent calculation of the band profile using this electron density shows that the left well potential is indeed approximately flat at zero bias conditions.

Figure 13.29(a) shows the zero bias absorption spectrum for sample C, which was obtained using the Schottky gate referencing technique. As a guide to the eye, centre lines are drawn at the peak positions, and the asymmetry of each peak is clearly seen. This asymmetry is clearly much stronger than in the previous sample (B), indicating the presence of stronger interference between the transitions. Similar asymmetric absorption spectra have recently been observed by Faist *et al* [37]. Our measured data are not fit accurately by either a sum of Voigt profiles or a sum of Lorentzians, as is seen in figure 13.29(a). The reduction of the absorption at energies between the two transitions is approximately 50% compared to the Voigt profile fit, and over 67% relative to the sum of Lorentzians fit.

The measured data are, however, fit extremely well, as shown in figure 13.29(b), by a more general interference formula [33, 38], which is written in terms of the dressed state parameters, Γ_{2d} and Γ_{3d} (total decay rates for levels 2d and 3d), μ_{12d} and μ_{13d} (dipole matrix elements from level 1 to levels 2d and 3d), and $\Delta\omega_{2d}$ and $\Delta\omega_{3d}$, the energy detunings from levels 2d and 3d. The

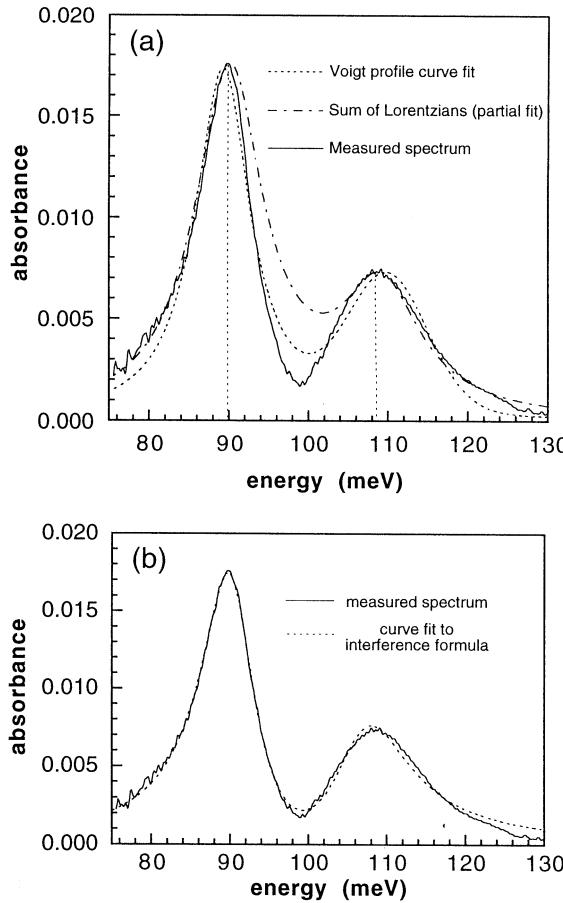


Figure 13.29. Zero bias absorption spectrum for sample C with fits to (a) the sum of Voigt profiles and the sum of Lorentzians, and (b) the interference formula given by equation (13.3).

expressions for $\chi^{(1)}$, and the absorption profile ($\text{Im } \chi^{(1)}$) are then [33, 39]

$$\chi^{(1)} \propto \frac{2(\Delta\omega_{2d}\mu_{13d}^2 + \Delta\omega_{3d}\mu_{12d}^2) - i(\Gamma_{2d}\mu_{13d}^2 + \Gamma_{3d}\mu_{12d}^2 + 2\kappa\mu_{12d}\mu_{13d})}{4\Delta\omega_{2d}\Delta\omega_{3d} - \Gamma_{2d}\Gamma_{3d} + \kappa^2 - 2i(\Gamma_{2d}\Delta\omega_{3d} + \Gamma_{3d}\Delta\omega_{2d})} \quad (13.6)$$

$$\begin{aligned} \text{Im}(\chi^{(1)}) \propto & \frac{4(\Delta\omega_{2d}\mu_{13d}^2 + \Delta\omega_{3d}\mu_{12d}^2)(\Gamma_{2d}\Delta\omega_{3d} + \Gamma_{3d}\Delta\omega_{2d})}{(4\Delta\omega_{2d}\Delta\omega_{3d} - \Gamma_{2d}\Gamma_{3d} + \kappa^2)^2 + 4(\Gamma_{2d}\Delta\omega_{3d} + \Gamma_{3d}\Delta\omega_{2d})^2} \\ & - \frac{(\Gamma_{2d}\mu_{13d}^2 + \Gamma_{3d}\mu_{12d}^2 + 2\kappa\mu_{12d}\mu_{13d})(4\Delta\omega_{2d}\Delta\omega_{3d} - \Gamma_{2d}\Gamma_{3d} + \kappa^2)}{(4\Delta\omega_{2d}\Delta\omega_{3d} - \Gamma_{2d}\Gamma_{3d} + \kappa^2)^2 + 4(\Gamma_{2d}\Delta\omega_{3d} + \Gamma_{3d}\Delta\omega_{2d})^2}. \end{aligned} \quad (13.7)$$

An additional parameter, κ , enters the formula, and gives an estimate of the size of the interference effect. Physically, this term corresponds to the coupling of the upper levels through the continuum and is given by

$$\kappa = \sqrt{\Gamma_{\text{lifetime}(2d)} \Gamma_{\text{lifetime}(3d)}}. \quad (13.8)$$

The total decay rates in the formula (Γ_{2d} and Γ_{3d}) can be split into lifetime and dephasing components

$$\Gamma = \Gamma_{\text{lifetime}} + \Gamma_{\text{deph}}. \quad (13.9)$$

Then the quantity

$$\frac{\kappa}{\sqrt{\Gamma_{2d} \Gamma_{3d}}} \quad (13.10)$$

is a number which varies from zero, in the absence of interference ($\Gamma_{\text{lifetime}} = 0$), to a value of one in the case of perfect interference ($\Gamma_{\text{deph}} = 0$). The measured data can then be fit to equation (13.7) to obtain information about the size of the interference effects. This fitting to the interference data is reported elsewhere [39], so we will briefly discuss these results here.

The peak positions and peak splitting are plotted in [figures 13.30\(a\) and \(b\)](#), respectively, as a function of carrier concentration. In contrast to the behaviour of the previous sample, B, the anti-crossing and the minimum peak splitting occur here at zero bias since the design of the left quantum well in sample C compensated the built-in field that is present. This indicates that the interference effect is optimized in this structure for zero bias, and indeed, the interference ratio of equation (13.10) gives a maximum value of 0.64 at zero bias.

We are now in a position to estimate the intersubband transition broadening contributions from both lifetime broadening and dephasing. From the interference formula curve fits, the following quantities are obtained:

$$\kappa = \sqrt{\Gamma_{\text{lifetime}(2d)} \Gamma_{\text{lifetime}(3d)}} \quad (13.11)$$

$$\Gamma_{2d} = \Gamma_{\text{lifetime}(2d)} + \Gamma_{\text{deph}(2d)} \quad (13.12)$$

$$\Gamma_{3d} = \Gamma_{\text{lifetime}(3d)} + \Gamma_{\text{deph}(3d)}. \quad (13.13)$$

To obtain a fourth equation needed to uniquely determine the four parameters, a transfer matrix calculation is used similar to the one presented in [figure 13.22](#), but with the structure of sample C. At resonance, there is a difference of tunnelling rates given by

$$\Gamma_{\text{lifetime}(3d)} - \Gamma_{\text{lifetime}(2d)} = 1.41 \text{ meV}. \quad (13.14)$$

This allows us to solve for all four parameters, with the results given in [table 13.1](#). Also stated in table 13.1 are the same parameters determined for sample B, which showed a weak signature of interference. From these numbers, we see that the most important difference between samples was an approximately 25% reduction in the dephasing rates for sample C, presumably a result of the increased width of the left quantum well. This reduction in dephasing makes a dramatic difference

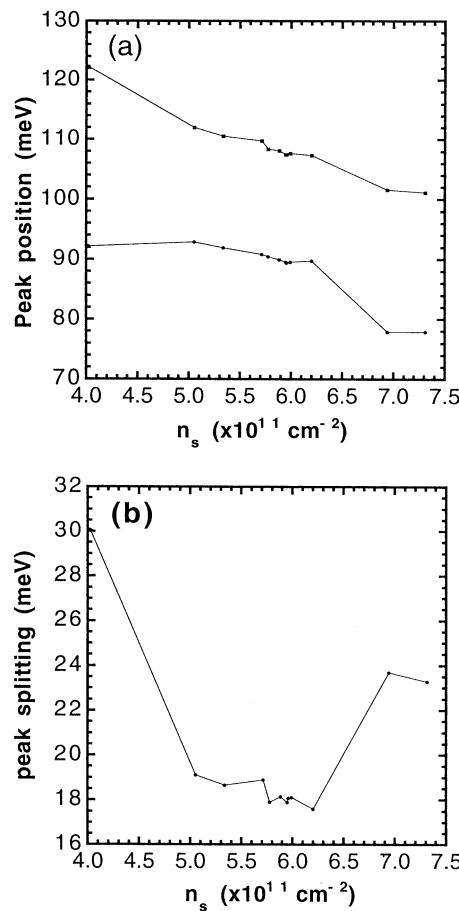


Figure 13.30. (a) Peak positions and (b) peak splitting extracted from curve fits to the interference formula for sample C.

in the appearance of the absorption profiles at resonance. Further reductions in dephasing would allow an even clearer observation of interference effects and would be important for fabricating devices based on this effect.

13.6 Conclusions

The decades-old promise of light emission from intersubband transitions in semiconductors is beginning to reach its potential. The past ten years have seen major advances in growth technology which have enabled the design of emission sources operating at wavelengths ranging from a few micrometres to

Table 13.1. Comparison of decay rates and interference ratio for samples B and C. (All numbers are in units of meV.)

Variable	Sample B	Sample C
Γ_{2d}	8.93	8.42
Γ_{3d}	12.42	11.97
κ	3.15	6.40
$\Gamma_{\text{lifetime}(2d)}$	5.05	5.74
$\Gamma_{\text{lifetime}(3d)}$	6.46	7.15
$\Gamma_{\text{depth}(2d)}$	3.88	2.68
$\Gamma_{\text{depth}(3d)}$	5.96	4.82

a few hundred micrometres. The advantages of using bandgap engineering to tailor emission frequencies for intersubband devices is clear, but many interesting challenges remain, particularly at longer wavelengths (20–100 μm). Characterization and understanding of the physical mechanisms at work in intersubband relaxation processes will certainly continue to play a key role in the further development of intersubband emission devices.

In our work with PQWs, we have shown that the far-infrared emission occurs at the bare harmonic oscillator frequency even though the wells are heavily populated with electrons, in accordance with the generalized Kohn's theorem. Compared to previous electrically pumped FIR emitters, this source has the attractive features that the emission spectrum is concentrated at a single frequency and the emission is obtained without the need for a magnetic field. Emission from a PQW using resonant vertical injection was demonstrated and may be promising for further improvements in the output power of this type of device. Optical excitation of a PQW with an ultrafast laser pulse has produced a few cycles of terahertz emission, and these results are consistent with the electrical excitation results.

We have also explored intersubband absorption in asymmetric coupled quantum well systems. In the case of purely bound states, anti-crossing behaviour is observed between the spatially direct and spatially indirect transitions. The linewidths extracted from the absorption data of bound systems are useful in designing samples in which the excited states are quasi-bound and coupled to a continuum of states by a thin tunnelling barrier. Through appropriate bandgap engineering, these structures exhibit a strong reduction of the absorption strength at an energy between the two bare state transition energies. This reduction of absorption relative to emission may enable the design of an intersubband laser which does not require a population inversion.

Acknowledgments

We have benefitted greatly from collaborations and discussions with many colleagues. In particular, it is a pleasure to thank H Schmidt, A Imamoglu, and S J Allen at the University of California Santa Barbara, and K Unterrainer, R Kersting, and E Gornik at Technische Universität Wien.

This work was supported by AFOSR under grant number F49620-94-1-0158 and by the NSF Center for Quantized Electronic Structures (QUEST).

References

- [1] Heyman J N, Unterrainer K, Craig K, Williams J, Sherwin M S, Campman K, Hopkins P F, Gossard A C, Murdin B N and Langerak C J G M 1996 *Appl. Phys. Lett.* **68** 3019
- [2] Helm M, Colas E, England P, DeRosa F and Allen S J Jr 1988 *Appl. Phys. Lett.* **53** 1714
- [3] Xu B, Hu Q and Melloch M R 1997 *Appl. Phys. Lett.* **71** 440
- [4] Gornik E and Tsui D C 1976 *Phys. Rev. Lett.* **37** 1425
- [5] Gornik E 1991 *Landau Level Spectroscopy* ed E I Rashba and G Landwehr (Amsterdam: North-Holland) p 911 and references therein
- [6] Gornik E 1972 *Phys. Rev. Lett.* **29** 595
- [7] Kobayashi K L I, Komatsubara K F and Otsuka E 1973 *Phys. Rev. Lett.* **30** 702
- [8] Andronov A A, Belyantsev A M, Gavrilenko V I, Dodin E P, Krasilnik Z F, Nikonorov V V and Pavlov S A 1984 *Sov. Phys.–JETP Lett.* **40** 989
- [9] Unterrainer K, Kremser C, Gornik E, Pidgeon C R, Ivanov Yu L and Haller E 1990 *Phys. Rev. Lett.* **64** 2277
- [10] Helm M, England P, Colas E, DeRosa F and Allen S J Jr 1989 *Phys. Rev. Lett.* **63** 74
- [11] Faist J, Capasso F, Sivco D L, Hutchinson A L and Cho A Y 1994 *Science* **264** 553
- [12] Sirtori C, Faist J, Capasso F, Sivco D L, Hutchinson A L and Cho A Y 1996 *Appl. Phys. Lett.* **69** 2810
- [13] Lyubomirsky I, Hu Q and Melloch M R 1998 *Appl. Phys. Lett.* **73** 3043
- [14] Rochat M, Faist J, Beck M, Oesterle U and Illegems M 1998 *Appl. Phys. Lett.* submitted
- [15] Brown E R, McIntosh K A, Nichols K B and Dennis C L 1995 *Appl. Phys. Lett.* **66** 285
- [16] Verghese S, McIntosh K A and Brown E R 1997 *Appl. Phys. Lett.* **71** 2743
- [17] Brey L, Johnson N F and Halperin B I 1989 *Phys. Rev. B* **40** 10 647
- [18] Kohn W 1961 *Phys. Rev.* **123** 1242
- [19] Wixforth A, Sundaram M, Ensslin K, English J H and Gossard A C 1991 *Phys. Rev. B* **43** 10 000
- [20] Gossard A C, Sundaram M and Hopkins P H 1994 *Epitaxial Microstructures* ed A C Gossard (Boston, MA: Academic) p 153
- [21] Strasser G, Bochter K, Witzany M and Gornik E 1991 *Infrared Phys.* **32** 439
- [22] Hirakawa K, Grayson M, Tsui D C and Kurdak C 1993 *Phys. Rev. B* **47** 16 651
- [23] Sen S, Capasso F, Gossard A C, Spah R A, Hutchinson A L and Chu S N G 1987 *Appl. Phys. Lett.* **51** 1428
- [24] Sha W, Smirl A L and Tseng W F 1995 *Phys. Rev. Lett.* **74** 4273

- [25] Kersting R, Unterrainer K, Strasser G, Gornik E and Kauffmann H F 1997 *Phys. Rev. Lett.* **79** 3038
- [26] Kersting R, Heyman J N, Strasser G and Unterrainer K 1998 *Phys. Rev. B* **58** 4553
- [27] Unterrainer K, Kersting R, Heyman J N, Strasser G, Maranowski K D and Gossard A C unpublished
- [28] Imamoglu A and Ram R J 1994 *Opt. Lett.* **19** 1744
- [29] Harris S E, Field J E and Imamoglu A 1990 *Phys. Rev. Lett.* **64** 1107
- [30] Harris S E 1989 *Phys. Rev. Lett.* **62** 1033
- [31] Imamoglu A and Harris S E 1989 *Opt. Lett.* **14** 1344
- [32] Boller K J, Imamoglu A and Harris S E 1991 *Phys. Rev. Lett.* **66** 2593
- [33] Zibrov A S, Lukin M D, Nikonov D E, Hollberg L W, Scully M O, Velichansky V L and Robinson H G 1995 *Phys. Rev. Lett.* **75** 1499
- [34] Padmabandu G G, Welch G R, Shubin I N, Fry E S, Nikonov D E, Lukin M D and Scully M O 1996 *Phys. Rev. Lett.* **76** 2053
- [35] Campman K L, Schmidt H, Imamoglu A and Gossard A C 1996 *Appl. Phys. Lett.* **69** 2554
- [36] Kroemer H 1994 *Quantum Mechanics: for Engineering, Material Science, and Applied Physics* (Englewood Cliffs, NJ: Prentice-Hall)
- [37] Faist J, Capasso F, Sirtori C, West K W and Pfeiffer L N 1997 *Nature* **390** 589
- [38] Schmidt H and Imamoglu A 1996 *Opt. Commun.* **131** 333
- [39] Schmidt H, Campman K L, Gossard A C and Imamoglu A 1997 *Appl. Phys. Lett.* **70** 3455

PART 5

SPIN-DEPENDENT PROPERTIES

Chapter 14

Spin-dependent properties of magnetic III–V semiconductors

Hideo Ohno

Laboratory for Electronic Intelligent Systems, Research Institute of Electrical Communication, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan
E-mail: ohno@riec.tohoku.ac.jp

Spin-dependent properties of a new magnetic semiconductor based on a III–V compound, $(\text{Ga},\text{Mn})\text{As}$, are reviewed. It is now possible to prepare epitaxial thin films of an alloy between non-magnetic GaAs and magnetic ions such as Mn that exhibit ferromagnetism with a transition temperature as high as 110 K by low-temperature molecular beam epitaxy ($< 300^\circ\text{C}$). Exchange interactions between localized Mn spins and conduction carriers manifest themselves in a variety of transport properties, and from these one can determine the magnitude of the interaction. $(\text{Ga},\text{Mn})\text{As}$ can readily be incorporated in the existing $(\text{Al},\text{Ga})\text{As}/\text{GaAs}$ semiconductor heterostructures, allowing one to explore new fields in semiconductor physics and technology, where both semiconducting and magnetic properties play critical roles.

14.1 Introduction

Modern information technology is based on the charge and spin of electrons. Semiconductor devices utilize the charge of electrons in semiconductors, whereas magnetic storage takes advantage of the spin of electrons in magnetic materials. It is then quite natural to ask oneself whether one can use both the charge and spin of electrons at the same time, especially in semiconductors, to enhance the device performance and to create new functionality that enriches further the already flourishing semiconductor electronics. Semiconductors used in electronics such as Si and GaAs are non-magnetic; that is, they contain no magnetic ions. This

makes properties of Si and GaAs relatively insensitive to the spin orientation of carriers, which is the reason why most of the spin-dependent phenomena could be virtually ignored in semiconductor device operations.

There are ways to enhance the magnitude of spin sensitivity. When magnetic ions are introduced in non-magnetic semiconductors, exchange interaction between band electrons and electrons localized at the magnetic ions modifies the properties of semiconductors and the spin-dependent phenomena are magnified. We can then incorporate such materials in heterostructures to control and utilize the magnified spin-dependent phenomena. Study of magnetic semiconductors based on non-magnetic semiconductors, called diluted magnetic semiconductors (DMSs), and its heterostructures had been centred mostly on II–VI based DMSs, such as CdTe and ZnS [2], although II–VI compounds are relatively less attractive for applications. An approach compatible with the present-day electronic materials is to make non-magnetic III–V semiconductors such as GaAs magnetic by introducing a high concentration of magnetic ions. GaAs and its heterostructures are currently being widely used for electronic well as optoelectronic devices, and has been the test bench for new physics and new device concepts. The introduction of magnetic III–V's based on GaAs, therefore, opens up the possibility of using a variety of magnetic and/or spin-dependent phenomena, not present in the conventional non-magnetic III–V's, in the optical and electrical devices already established.

This paper reviews the spin-dependent properties of a magnetic III–V ($(\text{Ga,Mn})\text{As}$). Section 14.2 describes the preparation of $(\text{Ga,Mn})\text{As}$ and its general properties. Section 14.3 presents the magnetic properties and spin-dependent magnetotransport properties. Section 14.4 discusses the possible origin of ferromagnetism observed in $(\text{Ga,Mn})\text{As}$. Section 14.5 presents the spin-dependent properties of $(\text{Ga,Mn})\text{As}$ -based heterostructures.

14.2 Preparation and basic properties of $(\text{Ga,Mn})\text{As}$

In order to observe pronounced spin-dependent phenomena, one needs to introduce a sizable amount of magnetic elements into non-magnetic semiconductors, which requires high doping of magnetic elements beyond its solubility limit in the case of III–V's. Munekata *et al* showed by the combination of InAs and Mn that low-temperature MBE (LT-MBE, growth temperature $< 300^\circ\text{C}$) could produce epitaxial layers exceeding the solubility limit and suppressing the Mn segregation phenomena [1], leading to successful epitaxial growth of $(\text{In,Mn})\text{As}$. Subsequent discovery of hole-induced ferromagnetic order in p-type $(\text{In,Mn})\text{As}$ [2] encouraged researchers to investigate GaAs-based systems [3] and lead to the successful growth of ferromagnetic $(\text{Ga,Mn})\text{As}$ [4]. Currently a number of groups are working on the MBE growth of III–V based DMS's to advance the understanding of this new material as well as heterostructures based on it [4–9].

14.2.1 Molecular beam epitaxy

Molecular beam epitaxial (MBE) growth of (Ga,Mn)As is carried out by using solid source MBE with elemental sources Ga, Mn, In, Al and As. Mn provides both localized spins and carriers (holes) due to its acceptor nature. Once calibrated, Mn content x in the $(\text{Ga}_{1-x}\text{Mn}_x)\text{As}$ films grown on GaAs substrates can be determined by x-ray diffraction measurements (XRD). Reflection high-energy electron diffraction (RHEED) patterns are used to monitor the surface reconstruction during growth, which was always carried out under As-stabilized conditions. The RHEED pattern of a GaAs buffer layer growth at substrate temperature T_S of 250 °C shows a (1×1) pattern (no reconstruction). The (Ga,Mn)As growth can be started by simply commencing the Mn beam during the low-temperature GaAs growth. The properties of grown (Ga,Mn)As depend on growth parameters such as As overpressure and T_S [10, 11]. However, as long as the established growth procedure is followed, the properties of (Ga,Mn)As films are reproducible; for a given Mn concentration x , the ferromagnetic transition temperature T_C is always in the range of $2000x \pm 10$ K ($x < 0.05$). The surface reconstruction of (Ga,Mn)As is (1×2) during and after growth. When the Mn flux or the substrate temperature or both were too high, the MnAs (NiAs (hexagonal)-structure) second-phase formation occurred. Maximum x is about 0.07, above which second-phase formation of the growth front occurs even at low-growth temperature.

14.2.2 Lattice constant and local lattice configuration

The relationship between the composition x and the lattice constant a of (In,Mn)As and (Ga,Mn)As determined by XRD is shown in figure 14.1 [4]. The relaxed lattice constant, a , of (Ga,Mn)As is calculated from the diffraction peak position in XRD spectra assuming that (Ga,Mn)As films are fully strained and that (Ga,Mn)As had the same elastic constant as GaAs. As can be seen from figure 14.1, a increases linearly with x following Vegard's law ($a = 0.566(1 - x) + 0.598x$ (nm)). The extrapolated lattice constant for hypothetical zincblende MnAs (0.598 nm) is in good agreement with the MnAs lattice constant determined from the extrapolation from the (In,Mn)As side (0.601 nm), which is close to a obtained by first-principle calculation [12].

Extended x-ray absorption fine structure (EXAFS) measurements of (Ga,Mn)As ($x = 0.005$ and 0.074) indicated that Mn was substitutionally incorporated into the Ga sublattice [13]. The Mn–As bondlength was found to be 0.249–0.250 nm, longer than the host Ga–As bondlength (0.244 nm) and shorter than the expected bondlength of Mn–As from the hypothetical zincblende MnAs (0.259 nm).

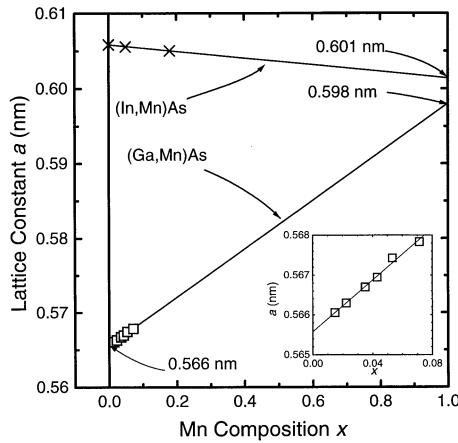


Figure 14.1. Relaxed lattice constant a versus Mn composition x in $(\text{Ga}_{1-x}\text{Mn}_x)\text{As}$ films. a was calculated from the x-ray diffraction peaks at room temperature, assuming that $(\text{Ga},\text{Mn})\text{As}$ layers are fully strained and that $(\text{Ga},\text{Mn})\text{As}$ has the same elastic constant as GaAs. The inset shows a magnified view of a of $(\text{Ga},\text{Mn})\text{As}$.

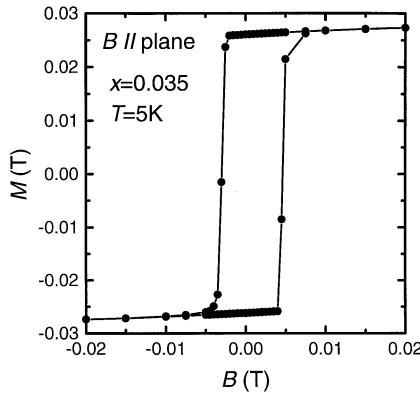


Figure 14.2. Magnetic field dependence of magnetization M at 5 K for a $(\text{Ga},\text{Mn})\text{As}$ film with Mn content $x = 0.035$. The field was applied parallel to the sample surface. The inset shows the temperature dependence of the remnant magnetization M_r of the same sample.

14.3 Magnetic and magnetotransport properties

14.3.1 Magnetic properties

Magnetization measurements with a SQUID (superconducting quantum interference device) magnetometer revealed sharp, square hysteresis loops in magnetization (M) versus magnetic field (B) curves at low temperatures as in

[figure 14.2](#), showing the presence of ferromagnetism in the (Ga,Mn)As films [2]. The ferromagnetic transition temperature T_C increases as $200x \pm 10$ K ($x < 0.05$); below 0.005, however, no ferromagnetism was detected. The magnetic easy axis was in-plane due to the compressive strain from the GaAs substrate; the easy axis can be made perpendicular with the reversal of the strain direction [14]. After saturation of the ferromagnetic part, a Brillouin function like paramagnetic responses were observed at high fields. Metallic films show only a small paramagnetic response, whereas insulating films show paramagnetic response as large as 50% of the total magnetization [15]. Whether this response is related to the spatial distribution of the ferromagnetic phase or intrinsic to ferromagnetism in (Ga,Mn)As is yet to be determined. The low-temperature saturation magnetization, M_S , of the (Ga,Mn)As films was consistent with the spin of Mn $S = 5/2$, although it is difficult to determine accurately S from these experiments alone, because of the error involved in determining x and the non-uniformity of x over the sample. The absence of remanence above T_C of (Ga,Mn)As confirmed that the concentration of MnAs ($T_C = 310$ K), if present, is very low. The susceptibility, χ , of (Ga,Mn)As films follows the Curie–Weiss form $\chi = C/(T - \theta)$, where C is the Curie constant and θ the paramagnetic Curie temperature. θ appears to fall close to T_C . In fully compensated (Ga,Mn)As samples however, θ becomes negative (-2 K) showing that the direct exchange among Mn is antiferromagnetic and the ferromagnetic interaction observed in these films are carried (hole) induced [16].

14.3.2 Magnetotransport properties

Resistivity as well as Hall resistivity measurements as a function of T as well as B have revealed rich information about the properties of (Ga,Mn)As. The Hall effect is dominated by the anomalous Hall effect in p-type (Ga,Mn)As, which made it possible to determine the magnetic properties from the magnetotransport measurements alone. The sheet Hall resistivity R_{Hall} (or the Hall resistivity ρ_{Hall}) in (Ga,Mn)As can be expressed as,

$$R_{\text{Hall}} = \frac{R_1}{d}B + \frac{R_S}{d}M \quad (\rho_{\text{Hall}} = R_0B + R_S M), \quad (14.1)$$

where R_0 is the ordinary (normal) Hall coefficient, R_S the anomalous Hall coefficient, d the sample thickness and M the magnetization of the sample. Separate measurements indicated that R_S is proportional to R_{sheet} in the present samples (skew scattering [17]) and thus $R_S/d = cR_{\text{sheet}}$, where c is a temperature-independent constant. The anomalous Hall term is the dominant term often up to room temperature. This allows one to determine M of the sample from $R_{\text{Hall}}(M \sim (1/c)R_{\text{Hall}}/R_{\text{sheet}})$. The ordinary Hall coefficient can be measured as the slope of the $R_{\text{Hall}} - B$ curve at low temperature under high-magnetic field, where M saturates, although the procedure is not as straightforward as it sounds because of the presence of negative magnetoresistance that extends to high fields.

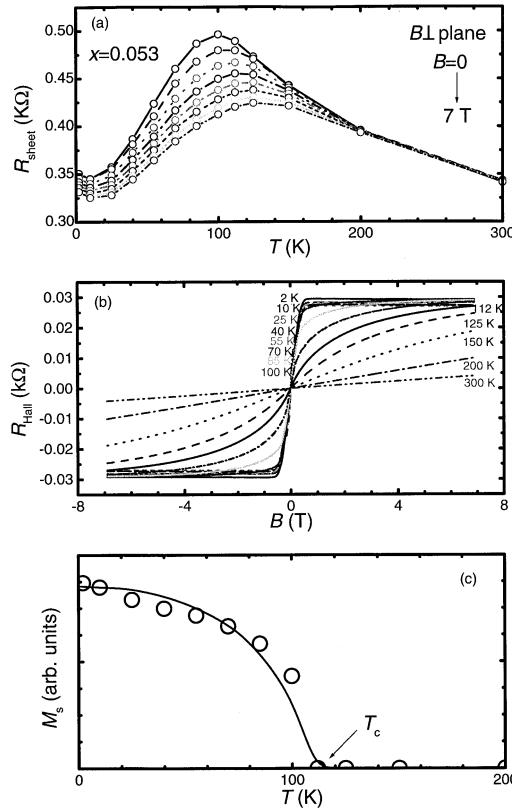


Figure 14.3. Magnetic field dependence of (a) the sheet resistivity R_{sheet} and (b) the sheet Hall resistivity R_{Hall} of $(\text{Ga},\text{Mn})\text{As}$ ($x = 0.053$) with temperature as a parameter. (c) Saturation magnetization determined from transport measurements, which shows that the transition temperature is 110 K.

Figure 14.3 shows the results of magnetotransport measurements of a 200 nm thick $(\text{Ga},\text{Mn})\text{As}$ layer grown on a $(\text{Al}_{0.9}\text{Ga}_{0.1})\text{As}$ buffer layer on (100) GaAs substrate ($x = 0.053$). R_{sheet} first increases as T decreases with an increase of negative magnetoresistance. The zero-field resistivity peaks at around T_C and then decreases (figure 14.3(a)). The T and B dependence of R_{Hall} reflects that of M (figure 14.3(b)). One can determine the ferromagnetic transition temperature T_C by the use of the Arrott plots [14, 18], in which $(R_{\text{Hall}}/R_{\text{sheet}})^2$ is plotted against $(B/(R_{\text{Hall}}/R_{\text{sheet}}))$ at each temperature to obtain a quantity proportional to M_S from the extrapolated intercept (M_S is zero when the intercept is at the origin). Note that $R_{\text{Hall}}/R_{\text{sheet}}$ is proportional to M . This way the T dependence of M_S and T_C can be determined from the transport results as shown in figure 14.3(c). The T dependence of M_S of the present $x = 0.053$ sample can be fitted

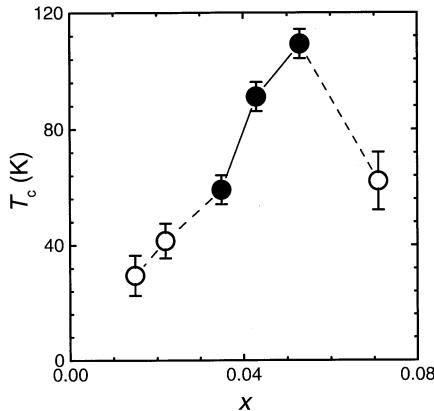


Figure 14.4. Ferromagnetic transition temperature T_C determined from magnetotransport measurements as a function of Mn composition x . Full circles show metallic samples; open circles show insulating samples.

with a standard Brillouin function as shown by the full curve in figure 14.3(c), which suggests that the ferromagnetism in (Ga,Mn)As may be understood in the framework of a mean field theory; in some of the samples, however, M_S shows rather linear T dependence. Since the anomalous Hall effect is dominant even at room temperature, and hence $\chi \approx (1/c)(R_{\text{Hall}}\mu_0/R_{\text{sheet}}B)$ holds for a wide temperature range, one can determine the paramagnetic Curie temperature by plotting $1/\chi$ (determined from transport measurements) as a function of T (see, for example, [19]). Thus, the determined θ is usually quite close to T_C . Figure 14.4 summarizes the x dependence of T_C for 200 nm (Ga,Mn)As grown on (Ga,Mn)As buffer layers ($x_{\text{Al}} = 0.9$).

The temperature T (2–300 K) dependence of R_{sheet} of six 200 nm thick (Ga,Mn)As epitaxial films grown on the $(\text{Al}_{0.9}\text{Ga}_{0.1})\text{As}$ buffer layer shows that the samples with intermediate Mn composition ($x = 0.035, 0.043$ and 0.053) are on the metal side of the metal–insulator transition, whereas low and high x samples are on the insulator side ($x = 0.015, 0.022$ and 0.071). At T_C , all samples show a hump in $R_{\text{sheet}} - T$ curves. This anomalous temperature dependence, especially in the metallic samples, is most probably due to the critical scattering. Critical scattering due to fluctuations of magnetic spins [20] is given by [21],

$$\rho = 3 \times 2\pi^2 \frac{k_F}{pe^2} \frac{m^2 J^2}{h^3} \frac{k_B T}{g^2 \mu_B^2} \chi(T, B), \quad (14.2)$$

where k_F is the Fermi wavenumber, m is the effective mass (we assume the heavy hole mass in GaAs, $m = 0.5m_0$, m_0 : free electron mass), J is the magnitude of exchange interaction between the carriers and the magnetic spins, k_B is the Boltzmann constant, e is elementary charge, h is the Plank constant, g is the

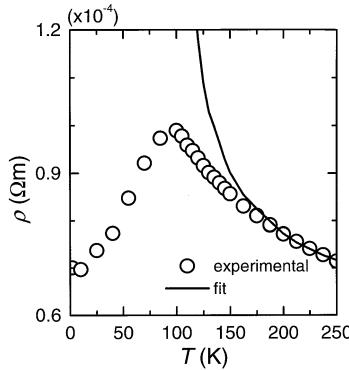


Figure 14.5. Circles show temperature dependence of resistivity ρ of a 200 nm sample with $x = 0.053$. Full curve is a fit of $\rho = a\chi T + b$ (a, b constants) to the high-temperature part, where χ is susceptibility. T dependence of χ was determined from the anomalous Hall effect. The functional form assumes the critical scattering for the temperature-dependent part.

Landé-factor of the Mn spins and μ_B is the Bohr magneton. The temperature dependence of ρ in the high-temperature region of metallic samples can be reproduced well by equation (14.2) as shown in figure 14.5, where the measured data can be fitted by χT , the characteristic temperature dependence of the critical scattering [21]. The deviation of the χT fit from experiment as T_C approaches is expected because the dc value of χ is used in the place of the wavevector (q) dependent χ , which is responsible for the scattering. Ordinarily, $\chi(q)$ is a decreasing function of q in ferromagnetic materials and thus the fit overestimates the scattering near T_C .

Negative magnetoresistance was observed in all the samples at low temperatures. Negative magnetoresistance of the metallic samples first increases as T decreases and peaks at T_C , whereas that of the insulating samples continues to increase as T decreases further below T_C and the magnetic field effect becomes quite pronounced at low temperatures. The negative magnetoresistance of metallic samples can also be understood as the reduction of spin-dependent scattering by aligning the spins by B described by equation (14.2). The fit of equation (14.2) to temperature and magnetic field dependence of ρ yields J (or $N_0\beta$) in the range of 1.3–1.5 eV. Previous estimates of $J(N_0\beta)$ using the expression for the high-temperature limit of equation (14.2) gives the upper limit of the interaction [18]. There still is, however, a number of factors one needs to take into account to properly understand J obtained from transport; for example, the effect of localization induced by disorder, which enhances the ferromagnetic interaction through electron-electron interaction, probably plays an important role here, and in the very vicinity of the metal-insulator transition, the description of spin-disorder scattering may also have to be modified. Large p-d exchange in

GaAs doped with Mn was reported by Szczytko *et al* [22] (2.5 eV, positive), who investigated magneto-optical properties. On the other hand, a smaller exchange of 1.0–1.2 eV (negative) was inferred from the photoemission experiments on (Ga,Mn)As [23]. Typical p–d exchange ($N_0\beta$) in II–VI DMSs is about 1 eV.

Very large negative magnetoresistance, which extends almost three orders of magnitude at low temperatures, has been observed in ‘reentrant’ insulating samples at high x [24]. This large negative magnetoresistance may be understood by considering either (or both) formation of bound magnetic polarons as was inferred in an earlier work on (In,Mn)As [6] or the increase of Fermi energy in a spin split band which reduces the localization length by reducing the energy difference between the mobility edge and the Fermi level [24]. Anisotropic conduction was also observed in these samples [25].

14.4 Origin of ferromagnetism

Since the magnetic interaction between Mn has been shown to be antiferromagnetic in semi-insulating fully compensated (Ga,Mn)As using Sn as a donor [16], the ferromagnetic interaction in magnetic III–V’s is most likely hole induced. There are currently two approaches to the understanding of the carrier-induced ferromagnetic interaction; one starts from the interaction between a carrier Fermi sea and localized magnetic moments, and the other starts from a d-band formed by the magnetic impurity, which is partially occupied because of the presence of carriers (holes).

One of the candidates of the former framework is a mean field theory developed by Dietl *et al* [26] developed taking into account the feedback mechanism between the magnetization polarization and the carrier polarization; magnetization polarization produces carrier polarization and it in turn produces magnetization polarization. This means field theory results in the same expression of T_C as that of the well-known Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction. The RKKY interaction is carrier-induced, sufficiently long ranged to account for the magnetic interaction in dilute systems, and has been put forward to explain the carrier (hole) induced ferromagnetism as well as spin-glass phase observed in a IV–VI compound PdSnMnTe [27].

Figure 14.6 shows T_C calculated by the mean field theory [28]. Here, T_C was calculated from the hole free energy, which was obtained by solving a (6×6) Luttinger Hamiltonian taking into account the presence of the p–d interaction. $x = 0.05$ and $|N_0\beta| = 1.0$ eV was assumed and the direct antiferromagnetic interaction among Mn was neglected. Figure 14.6 implies that $N_0\beta = 1.3$ eV to obtain $T_C = 110$ K for $p = 3.5 \times 10^{20}$ cm $^{-3}$, in rather good agreement with the range of $N_0\beta$ inferred from experiments. The mean field approach can also explain the absence of ferromagnetism in n-type materials; in n-type materials, the small effective mass together with the small $N_0\alpha$ (about 0.2 eV) makes it difficult for the ferromagnetic interaction to overcome the direct

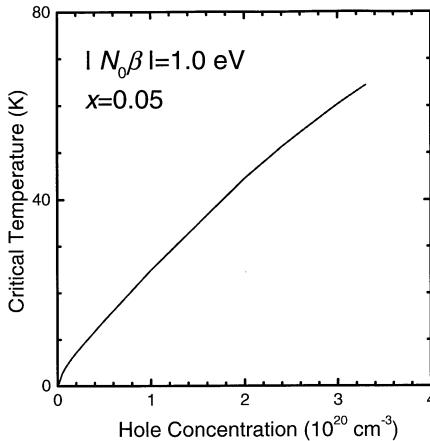


Figure 14.6. Ferromagnetic transition temperature T_C as a function of hole concentration calculated by a mean field theory. $x = 0.05$ and $|N_0\beta| = 1.0 \text{ eV}$ are assumed. T_C was obtained from the hole free energy calculated from a (6×6) Luttinger Hamiltonian with p-d exchange.

antiferromagnetic coupling among Mn. The mean field picture was also used to explain the hole-induced ferromagnetism in II–VI quantum wells [26, 29].

Ferromagnetism in insulating samples can be understood along the same line, if one considers the difference between the lengthscales associated with the transport and the magnetism. In insulating samples close to the metal–insulator transition, the localization length is smaller than the sample size (millimetres) but may still be significantly larger than the lengthscale of magnetic interactions (nanometres), thus making the mean field approach a good starting point, although localization phenomena would be increasingly important.

The other approach to understanding ferromagnetism in III–V DMS's starts from the d-band formed by transition metal impurities. Based on a first principle calculation of the ground-state energy of (In,Mn)As, which revealed that the ferromagnetic phase is indeed stable, Akai stressed the d-character of the holes and invoked a double-exchange picture for the observed ferromagnetism [30]. The energy gain of the ferromagnetic phase is proportional to the carrier concentration and therefore the calculation showed that compensating the holes results in destabilization of the ferromagnetic state. It is also interesting to point out that the calculation indicates that (In,Mn)As is half metallic, as was pointed out for (Ga,Mn)As using hypothetical ordered alloy and local spin density approximation [12].

14.5 Heterostructures

New physics such as fractional quantum Hall effects, have emerged from non-magnetic III-V semiconductor heterostructures. It has also been a test bench for a number of new device concepts, among which are quantum well lasers and high electron mobility transistors. Ferromagnetic III-V's can add a new dimension to the III-V heterostructure systems because they can introduce magnetic cooperative phenomena that were not present in conventional III-V materials.

Growth of semiconductor-based ferromagnetic/non-magnetic (Ga,Mn)As/GaAs multilayers was demonstrated by LT-MBE. XRD diffraction revealed high structural quality [19], and transport measurements showed that ferromagnetism can be retained down to 5 nm of the (Ga,Mn)As layer below which it became paramagnetic. The reason for this is not clearly understood; it could be due to Mn segregation at the initial stage of growth, which leads to depletion of Mn, or it could be due to some fundamental physics of ferromagnetism. (Ga,Mn)As quantum wells have also been realized and studied by magneto-optical means [31].

14.5.1 Ferromagnetic/non-magnetic/ferromagnetic trilayers

14.5.1.1 Interlayer coupling

A study was carried out to look into the magnetic interactions between two ferromagnetic (Ga,Mn)As layers separated by a non-magnetic semiconducting layer. In order to examine the presence and nature of the magnetic interaction, a series of structures with a 30 nm (Ga,Mn)As ($x = 0.04$) layer and a 30 nm (Ga,Mn)As ($x = 0.02$) layer separated by a non-magnetic (Al,Ga)As layer were grown on a 150 nm (Al_{0.9}Ga_{0.1})As buffer layer on (001) GaAs substrates [32]. Two sets of trilayer structures were prepared; one with GaAs as the non-magnetic layer with thickness d_{GaAs} ranging from 0 to 106 monolayers (ML) and the other with 10 ML (Al,Ga)As with two different Al compositions ($x_{\text{Al}} = 0.16$ and 0.20). From the measurements of the reference layers, T_C of (Ga,Mn)As $x = 0.02$ and 0.04 and 40 K and 80 K, respectively. R_{sheet} of the $x = 0.02$ layer was about three orders of magnitude higher than that of the $x = 0.04$ layer. This sheet resistivity difference allows one to probe only the properties of the $x = 0.04$ layer in the trilayer structures by the magnetotransport measurements.

Figure 14.7(a) shows $R_{\text{Hall}}/R_{\text{sheet}} - B$ curves of the trilayer structures with n MLs of the intermediary GaAs layer at 2 K with B perpendicular to the plane (the direction of the hard axis). Note that $R_{\text{Hall}}/R_{\text{sheet}}$ is approximately proportional to M perpendicular to the plane. As the thickness of the GaAs layer decreases (from $n = 106$ to 0), B_{sat} (at which $R_{\text{Hall}}/R_{\text{sheet}}$ saturates) of the top $x = 0.04$ layer decreases. The reduction of B_{sat} shows stronger coupling between the two magnetic layers, since the bottom (Ga,Mn)As layer ($x = 0.02$) has lower B_{sat} than that of the top layer (confirmed by separate measurements

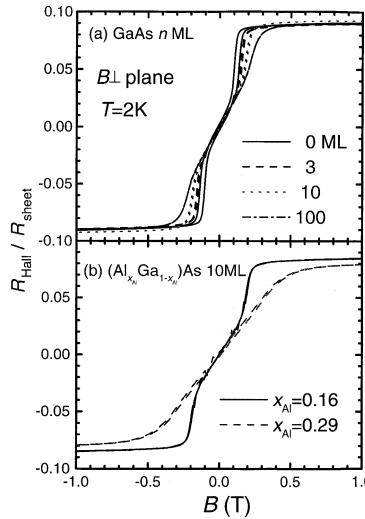


Figure 14.7. (a) $R_{\text{Hall}} / R_{\text{sheet}}$ versus B at 2 K for $(\text{Ga},\text{Mn})\text{As}/n$ ML's $(\text{Ga},\text{Mn})\text{As}$ ($n = 0, 3, 10, 100$) trilayer structures and (b) $(\text{Ga},\text{Mn})\text{As}/(\text{Al},\text{Ga})\text{As}/(\text{Ga},\text{Mn})\text{As}$ ($x_{\text{Al}} = 0.16, 0.29$) trilayer structures. Note that $R_{\text{Hall}} / R_{\text{sheet}}$ is, to first approximation, proportional to M perpendicular to the plane.

on reference samples). The results of B dependence of $R_{\text{Hall}} / R_{\text{sheet}}$ of the two $(\text{Ga},\text{Mn})\text{As}/10$ ML's $(\text{Al},\text{Ga})\text{As}/(\text{Ga},\text{Mn})\text{As}$ trilayer structures at 2 K are shown in figure 14.7(b) ($x_{\text{Al}} = 0.16$ and 0.29). As can be seen in figure 14.7(b), B_{sat} of $x_{\text{Al}} = 0.29$ is higher than that of $x_{\text{Al}} = 0.16$, showing weakening of coupling by increasing the barrier height in the valence band.

The coupling between the two $(\text{Ga},\text{Mn})\text{As}$ layers can be quantified in the following way. The total energy per unit volume of the single magnetic layer can be expressed as,

$$E = K \cos^2 \theta_K (1 - \alpha \cos^2 \theta_K) - (MB/\mu_0) \cos \theta - J \cos \theta_J, \quad (14.3)$$

where the first term is the anisotropy energy, the next the Zeeman energy, and the third the coupling energy between the two $(\text{Ga},\text{Mn})\text{As}$ layers; K is the anisotropy energy density, θ_K is the angle between M and the hard axis, θ is the angle between M and B , θ_J is the angle between M at the top of $(\text{Ga},\text{Mn})\text{As}$ and M at the bottom of $(\text{Ga},\text{Mn})\text{As}$, α (~ 0.35) is an anisotropy constant introduced to describe the magnetization jump around $B \sim 0.2$ T, μ_0 is the magnetic permeability of vacuum and J is the coupling energy density. Because B_{sat} of the $x = 0.02$ layer is lower than B_{sat} of $x = 0.04$, we assume that M of the $x = 0.02$ layer is along B and fully saturated when the $x = 0.04$ layer starts its reversal. We can then set $\theta = \theta_K = \theta_J$, since B , M of the bottom layer, and the direction of the hard axis are now all perpendicular to the sample plane. K

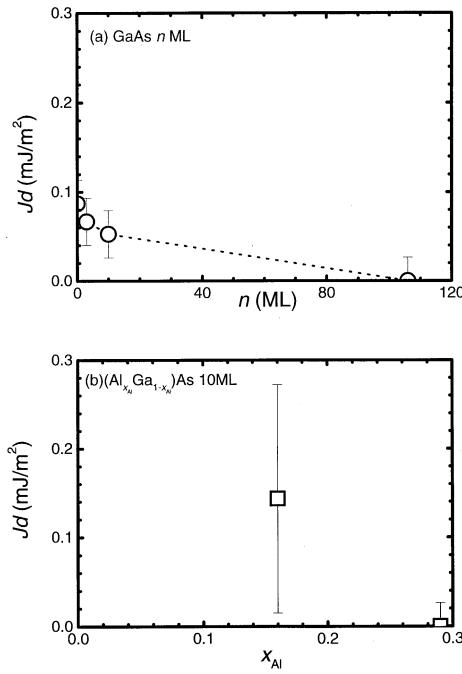


Figure 14.8. (a) Thickness and (b) Al content dependence of the coupling energy between two (Ga,Mn)As layers. d is the thickness of the magnetic layer. The magnetic coupling is consistent with the presence of carrier mediated magnetic interaction. As thickness becomes thicker or the barrier-height becomes higher, holes in the non-magnetic layer decrease and the magnetic coupling becomes weaker.

is the same for all the $x = 0.04$ layers, because all the $x = 0.04$ layers have the same thickness and the same geometry. For further analysis, magnetic coupling is assumed to be absent ($J = 0$) for the sample with non-magnetic 106 ML GaAs (too thick for any magnetic interaction) and for the $x_{\text{Al}} = 0.29$ sample (the two layers were shown to be decoupled experimentally by SQUID). Since non-zero J alters the magnetization process, we can deduce J from the difference in the magnetization curves.

Figure 14.8(a) and (b) shows J determined from figures 14.7(a) and (b) as a function of thickness or composition of the intermediary layer. J is always ferromagnetic because B_{sat} is always reduced when interaction sets in. The magnetic coupling is surprisingly long ranged. Both sets of results indicate the critical role of the holes in the intermediary layer on the magnetic coupling; the higher the hole concentration, the stronger the interaction. The results seem to indicate the magnetic interaction being mediated by holes in the non-magnetic intermediary layer. Recent calculation shows that the magnetic interaction should

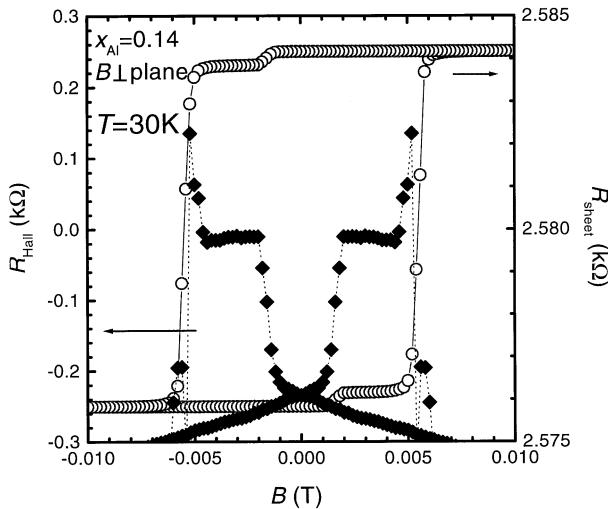


Figure 14.9. The magnetic field dependence of R_{Hall} and R_{sheet} of a 30 nm (Ga,Mn)As ($x = 0.05$)/10 ML (Al,Ga)As ($x_{\text{Al}} = 0.14$)/30 nm (Ga,Mn)As ($x = 0.03$) trilayer structure grown on a lattice relaxed (In,Ga)As buffer layer to tensile strain the structure. The step structure in R_{Hall} at ± 0.012 – 0.015 T shows that M of the two ferromagnetic layers are antiparallel in that magnetic field range; note that R_{Hall} is proportional to M and the easy axis is now perpendicular to the plane due to tensile strain. R_{sheet} shows a marked increase when the magnetization M in the two magnetic layers are antiparallel.

oscillate as a function of the hole density and thickness [33]. In order to clarify the mechanism responsible for the interaction more careful study is needed.

14.5.1.2 Spin-dependent transport

Figure 14.9 shows the result of magnetotransport measurements on a trilayer structure, where a clear increase in resistance is observed when M of the two layers becomes antiparallel [34]. Current flow is along the plane, i.e. the measurement configuration is current-in-plane (CIP). R_{Hall} in this case shows clear hysteresis reflecting M of the two layers because (a) tensile strain is built into the trilayer to make the easy axis perpendicular to the plane by the use of a (In,Ga)As buffer layer whose lattice constant is greater than the (Ga,Mn)As layers used, and (b) x of the two layers (0.03 and 0.05) were chosen in such a way that conductance of both layers are comparable at low temperature.

14.5.1.3 Tunnelling magnetoresistance

Since AlAs can produce a high potential barrier (550 meV) in the valence band with respect to the GaAs band, it is interesting to see whether one can observe

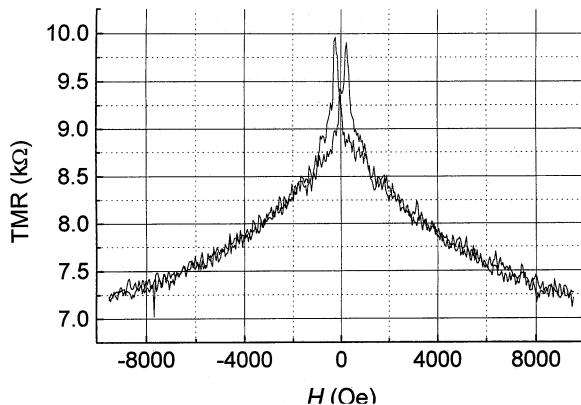


Figure 14.10. The magnetic field dependence of tunnelling resistance measured at 4.2 K on a trilayer structure with 200 nm (Ga,Mn)As/3 nm AlAs/200 nm (Ga,Mn)As ($x = 0.039$) [33].

tunnelling magnetoresistance (TMR) in all semiconductor ferromagnetic/non-magnetic structures. Figure 14.10 shows that one can indeed observe TMR in (Ga,Mn)As/AlAs/(Ga,Mn)As structures (4.2 K). The structure consists of 200 nm (Ga,Mn)As ($x = 0.039$) on both sides of 3 nm thick AlAs and H was applied in-plane. The TMR ratio defined by $(R_{\max} - R_{\min})/R_{\min}$ is 44.4% at 4.2 K [35].

14.5.2 Resonant tunnelling structures

When a semiconductor becomes ferromagnetic, spin splitting of the conduction as well as the valence bands is expected to occur due to s-d and p-d exchange interactions. Because (Ga,Mn)As can be grown pseudomorphically on non-magnetic GaAs-based heterostructures, one can fabricate resonant tunnelling diodes (RTD), a powerful tool for spectroscopy of electronic states, to investigate the spontaneous band splitting in (Ga,Mn)As. Temperature and magnetic field dependence of current–voltage (I – V) characteristics of p-type AlAs/GaAs/AlAs double barrier RTD's having (Ga,Mn)As as an emitter material were studied [36, 37]. Spontaneous splitting of resonant peaks in the absence of magnetic fields was observed upon lowering the measurement temperature below T_C , which is interpreted as the spin splitting of the valence band of ferromagnetic (Ga,Mn)As observed in tunnelling spectra. The structure studied consisted of (from the surface side) 150 nm (Ga_{0.97}Mn_{0.03})As/15 nm undoped GaAs spacer/5 nm undoped AlAs barrier/4 nm undoped GaAs quantum well/5 nm undoped AlAs barrier/5 nm undoped GaAs spacer/150 nm Be doped GaAs ($p = 5 \times 10^{17} \text{ cm}^{-3}$)/150 nm Be doped GaAs ($p = 5 \times 10^{18} \text{ cm}^{-3}$)/p⁺ GaAs substrates. All layers were grown at 650 °C except for the last (Ga,Mn)As layer which was grown at 250 °C.

14.6 Conclusion

The spin-dependent properties of the ferromagnetic III–V semiconductor (Ga,Mn)As and its heterostructures were described. Ferromagnetic (Ga,Mn)As prepared by LT-MBE exhibits ferromagnetism with T_C as high as 110 K. Magnetic and magnetotransport properties of (Ga,Mn)As epitaxial films showed the critical role played by holes and its exchange interaction with Mn local spins in ferromagnetism. Heterostructures based on (Ga,Mn)As, such as trilayer structures and RTD's, were shown to exhibit a number of new properties not present in non-magnetic structures. Ferromagnetic III–V's can thus be used to explore a new field of semiconductor physics and technology, where both semiconductor-heterostructure properties and magnetic/spin-dependent properties play critical roles.

Acknowledgments

The author thanks F Matsukura, A Shen, N Akiba, T Omiya, Y Ohno, A Oiwa, S Katsumoto, Y Iye, H Munekata, T Dietl, B Beschoten, P Crowell, D D Awschalom and A Twardowski for valuable collaboration and discussion. The work of Tohoku University was supported by the ‘Research for the Future’ Program (JSPS-RFTF97P00202) from the Japan Society for the Promotion of Science and by grant-in-aid on Priority Area ‘Spin Controlled Semiconductor Nanostructures’ (09244103) from the Ministry of Education, Japan.

References

- [1] Munekata H, Ohno H, von Molnár S, Segmüller A, Chang L L and Esaki L 1989 *Phys. Rev. Lett.* **63** 1849
- [2] Ohno H, Munekata H, Penney T, von Molnár S and Chang L L 1992 *Phys. Rev. Lett.* **68** 2664
- [3] De Boeck J, Oesterholt R, Van Esch A, Bender H, Bruynseraeede C, Van Hoof C and Borghs G 1996 *Appl. Phys. Lett.* **68** 2744
- [4] Ohno H, Shen A, Matsukura F, Oiwa A, Endo A, Katsumoto S, Iye Y 1996 *Appl. Phys. Lett.* **69** 363
- [5] Ohno H 1998 *Science* **281** 951
- [6] Hayashi T, Tanaka M, Nishinaga T, Shimada H, Tsuchiya H and Otsuka Y 1997 *J. Cryst. Growth* **175/176** 1063
- [7] Nishikawa Y, Satoh Y and Yoshino J 1997 *2nd. Symp. on Physics and Application of Spin Related Phenomena in Semiconductors* (Jan. 27–28, 1997, Sendai, Japan) p 122
- [8] Van Esch A, Van Bockstal L, De Boeck J, Verbanck G, van Steenbergen A S, Wellmann P J, Grietens B, Bogaerts R, Herlach F and Borghs G 1997 *Phys. Rev. B* **56** 13 103
- [9] Sadowski J, Domagala J, Bak-Misiuk J, Sqiatek K, Kanski J, Ilver L and Oscarsson H 1998 *Acta Phys. Polon. A* **94** 509

- [10] Matsukura F, Shen A, Sugawara Y, Omiya T, Ohno Y and Ohno H 1998 *Presented at The 25th Int. Symp. on Compound Semiconductors (October 12–16, Nara, Japan)*
- [11] Shimizu H, Hayashi T, Nishinaga T and Tanaka M 1999 *Appl. Phys. Lett.* **74** 398
- [12] Shirai M, Ogawa T, Kitagawa I and Suzuki N 1998 *J. Magn. Magn. Mater.* **177–181** 1383
See also Ogawa T, Shirai M, Suzuki N and Kitagawa I 1999 *J. Magn. Magn. Mater.* **196** 428
- [13] Shioda R, Ando K, Hayashi T and Tanaka M 1998 *Phys. Rev. B* **58** 1100
- [14] Ohno H, Matsukura F, Shen A, Sugawara Y, Oiwa A, Endo A, Katsumoto S and Iye Y 1996 *Proc. 23rd Int. Conf. on Physics of Semiconductors* ed M Scheffler and R Zimmermann (Singapore: World Scientific) pp 405–8
- [15] Oiwa A, Katsumoto S, Endo A, Hirawawa M, Iye Y, Ohno H, Matsukura F, Shen A and Sugawara Y 1997 *Solid State Commun.* **103** 209
- [16] Satoh Y, Inoue N, Nishikawa Y and Yoshino J 1997 *3rd Symp. on Physics and Application of Spin Related Phenomena in Semiconductors (Nov. 17–18, Sendai, Japan)* p 23
- [17] See, e.g., Chien C L and Westgate C R (ed) 1980 *The Hall Effect and Its Applications* (New York: Plenum)
- [18] Matsukura F, Ohno H, Shen A and Sugawara Y 1998 *Phys. Rev. B* **57** R2037
- [19] Shen A, Ohno H, Matsukura F, Sugawara Y, Ohno Y, Akiba N and Kuroiwa T 1997 *Japan. J. Appl. Phys.* **36** L73
- [20] von Molnár S and Kasuya T 1968 *Phys. Rev. Lett.* **21** 1757
- [21] Dietl T 1994 (Diluted) Magnetic semiconductors *Handbook of Semiconductors* vol 3B, ed S Mahajan (Amsterdam: North-Holland) p 1251
- [22] Szczytko J, Mac W, Stachow A, Twardowski A, Becla P and Tworzydlo J 1996 *Solid State Commun.* **99** 927
- [23] Okabayashi J, Kimura A, Rader O, Mizokawa T, Fujimori A, Hayashi T and Tanaka M 1998 *Phys. Rev. B* **58** R4211
- [24] Oiwa A, Katsumoto S, Endo A, Hirasawa M, Iye Y, Ohno H, Matsukura F, Shen A and Sugawara Y 1998 *Phys. Status Solidi b* **205** 167
- [25] Katsumoto S, Oiwa A, Iye Y, Ohno H, Matsukura F, Shen A and Sugawara Y 1998 *Phys. Status Solidi b* **205** 115
- [26] Dietl T, Haury A and Merle d'Aubigné Y 1997 *Phys. Rev. B* **55** R3347
- [27] Story T, Galazka R R, Frankel R B and Wolff P A 1986 *Phys. Rev. Lett.* **56** 777
- [28] Dietl T unpublished
- [29] Haury A, Wasiela A, Arnoult A, Cibert J, Tatarenko S, Dietl T and Merle d'Aubigné Y 1997 *Phys. Rev. Lett.* **79** 511
- [30] Akai H 1998 *Phys. Rev. Lett.* **81** 3002
- [31] Hayashi T, Tanaka M, Seto K, Nishinaga T and Ando K 1997 *Appl. Phys. Lett.* **71** 1825
- [32] Akiba N, Matsukura F, Shen A, Ohno Y, Ohno H, Oiwa A, Katsumoto S and Iye Y 1998 *Appl. Phys. Lett.* **73** 2122
- [33] Jungwirth T, Atkinson W A, Lee B H and MacDonald A H 1999 *Phys. Rev. B* **59** 9818
- [34] Akiba N, Nakata K, Matsukura F and Ohno H unpublished results
- [35] Hayashi T, Shimada H, Shimizu H and Tanaka M 1999 *J. Crystal Growth* **201/202** 689

- [36] Ohno H, Akiba N, Matsukura F, Shen A, Ohtani K and Ohno Y 1998 *Appl. Phys. Lett.* **73** 363
- [37] Akiba N, Matsukura F, Ohno Y, Shen A, Ohtani K, Sakon T, Motokawa M and Ohno H 1998 *Physica B* **256–258** 561

Chapter 15

Zeeman separation in GaAs quantum structures

Jong-Chun Woo¹ and Keon-Ho Yoo²

¹ *Department of Physics, Seoul National University, Seoul 151-742, Korea*

² *Department of Physics, Kyunghee University, Seoul 130-701, Korea*

15.1 Introduction

An electron has an associated magnetic moment arising from its spin and orbital motion. When the electron is placed in an external magnetic field B , the coupling between the magnetic moment and the magnetic field splits energy levels that are degenerate at zero magnetic field. The corresponding splitting in the spectral lines is called the Zeeman effect, which was first observed by P Zeeman in 1896.

For an atom of total angular momentum J , a Zeeman splitting is linear to B in a weak magnetic field $\vec{B} = B\hat{z}$:

$$\Delta E = g\mu_B B \Delta J_z, \quad (15.1)$$

where $\mu_B = e\hbar/2mc$ is the Bohr magneton. In the case of LS coupling, the Landé g factor g is given by [1]

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \quad (15.2)$$

Therefore the value of g experimentally obtained from the Zeeman spectroscopy presents a valuable information on the spin and angular momentum states of the atom. For the free electron whose $L = 0$ and $J = S = 1/2$, the g -value is 2.0 according to equation (15.2). However, the exact g -value is 2.00229 due to the vacuum fluctuation effect. In a strong magnetic field region, where the

magnetic energy $\hbar\omega_c = \hbar eB/mc$ becomes comparable to and larger than the LS coupling energy, equation (15.2) is inapplicable, since J is no longer a good quantum number.

The Zeeman effect in a semiconductor is more complicated than that of an atom due to the band dispersion relation and the degeneracy of electronic states. The value of g itself is far different from that of the free electron. It may also be anisotropic, and in some cases it may not be constant but B -dependent. In semiconductor quantum structures, the Zeeman effect is far more complicated because of the low-dimensional confinement introduced by the structures, and the observed data are often very difficult to interpret. Still, the Zeeman splitting measurements provide a very useful piece of information on the electronic state of the semiconductor, and serve a good test for the theoretical band structure.

In this paper, we review the results on Zeeman splitting of GaAs–AlGaAs quantum wells (QW), quantum wires (QWR) and quantum dots (QD). Also included are the results on GaAs and AlGaAs bulk states for comparison.

15.2 Zeeman effects of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ bulk

The Γ_6 conduction band state in a middle-gap semiconductor such as GaAs and AlGaAs is usually described by a parabolic band, and in this approximation the energy of an electron in an external magnetic field $\vec{B} = B\hat{z}$ is described by

$$E_c = E_{co} + (N + 1/2)\hbar\omega_c + g_e\mu_B S_{e,z}B \quad (S_{e,z} = \pm 1/2), \quad (15.3)$$

where N is the Landau quantum number, $S_{e,z}$ is the spin quantum number, and $\omega_c = eB/m_e^*c$ is the cyclotron frequency described with the electron effective mass m_e^* . If one isolates the spin Hamiltonian, it may be written as

$$H_e = g_e\mu_B \sum_{i=x,y,z} S_{e,i} B_i. \quad (15.4)$$

The Zeeman splitting is linear in B and the major contribution to the electron g -factor g_e comes from the spin-orbit coupling as calculated by the three- [2] or five-band [3] $\mathbf{k} \cdot \mathbf{p}$ theory.

The g_e value of GaAs measured by the spin quantum beat spectroscopy is -0.44 at 5 K and becomes less negative as the temperature increases [4]. In the case of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy, the electron g -factor increases as the aluminum composition x increases, and becomes positive after passing zero at $x \approx 0.12$ [5]. It has been reported that $g_e = +0.4$ at $x = 0.3$ and $+0.5$ at $x = 0.35$ [6].

The degenerate Γ_8 valence band state is usually described by the Luttinger (4×4) $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian [7], and its four (corresponding to $J_{h,z} = -3/2, -1/2, 1/2$ and $3/2$) Landau level spectra in a high magnetic field region are not linear in B and anisotropic. The hole spin Hamiltonian can be written as [8]

$$H_h = -2\mu_B \sum_{i=x,y,z} (\kappa J_{h,i} + q J_{h,i}^3) B_i. \quad (15.5)$$

For the interband optical transitions, the excitonic effect becomes dominant at low temperature. Theoretical descriptions of magnetoexcitons have been carried out in three magnetic field regions; low-, intermediate- and high-field regions depending on the ratio γ of the magnetic energy to the exciton binding energy. In the low-field regime ($\gamma \leq 0.4$, which is equivalent to $B \leq 1.5$ T in the case of GaAs), a perturbation scheme has been used [9, 10]. On the other hand, in the high-field regime ($\gamma \geq 2.5$; $B \geq 9.1$ T for GaAs), the adiabatic scheme works well [11]. In the intermediate regime, a phenomenological description was often employed [12] to cover the intermixing effect of low- and high-field behaviours. The eigenvalues in the low-field should contain the linear Zeeman and quadratic diamagnetic terms, while those of the high-field observation should contain the Landau level spectra.

If one focuses on the spin-dependent properties of the magnetoexcitons, the excitonic spin Hamiltonian can be written as [8]

$$H_{\text{ex}} = H_e + H_h + H_{e-h}, \quad (15.6)$$

where H_e and H_h are given in equations (15.4) and (15.5), and

$$H_{e-h} = - \sum_{i=x,y,z} (a S_{e,i} J_{h,i} + b S_{e,i} J_{h,i}^3). \quad (15.7)$$

In the jj coupling scheme, the electron of $S_e = 1/2$ and the hole of $J_h = 3/2$ make $J = 1$ and $J = 2$ excitonic states which are splitted by the electron-hole exchange interaction. In a magnetic field, the three Zeeman components of $J = 1$ state and the three out of five Zeeman components of $J = 2$ state are optically allowed [13].

All the six Zeeman components of the 1S exciton are observed in GaAs. [Figure 15.1](#) is the four reflection minima observed in the Faraday geometry [13]. The g_{ex} of the $J = 1$ exciton is estimated to be -1.9 ± 0.2 . The early report of Nam *et al* [12] had obtained the values of electron and hole Zeeman splitting factors as $g_e = -0.50 \pm 0.05$ and $\kappa = 1.0 \pm 0.2$, even though the assignment on the pair of Zeeman spectra was different from that which later becomes conventional. Hess *et al* [14] also observed a linear Zeeman splitting up to 20 T, and deduced $g_e = -0.44$ and $\kappa = 1.2$ corresponding to $g_{\text{ex}}(J = 1) = -1.97$. The observed Zeeman splitting is not sensitive to the crystal orientation.

15.3 GaAs/Al_xGa_{1-x}As quantum well

The QW system has an anisotropy and its response depends strongly on the direction of the magnetic field. The relevant energies related to the description of the magneto-optical properties are the magnetic energy, the exciton binding energy and the quantum confinement energy. The latter two strongly depend on the QW width and the barrier height. Due to this complexity, there seems to be no

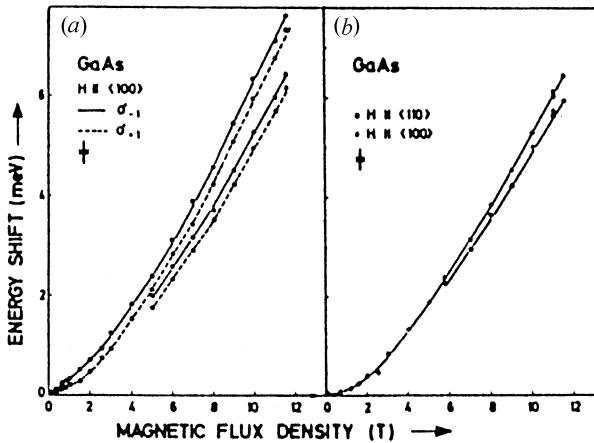


Figure 15.1. Energy shifts of the reflectance minima of the 1S excitons of GaAs bulk in (a) Faraday and (b) Voigt configurations. After [13].

unified theory that can analyse the Zeeman effects in all cases.. However, there are many approximate theories that are used to analyse specific experimental results.

The electron Zeeman splitting was investigated by the electron-spin-resonance (ESR) induced change of the magnetoresistivity [15, 16] in the modulation-doped GaAs/Al_xGa_{1-x}As ($x = 0.3$ or 0.35) QW. The splitting was found to be quadratic in B and the corresponding g_e factor was expressed as

$$g_e(B, N) = g_0 - c(N + 1/2)B, \quad (15.8)$$

where the two constants g_0 and c depend on the carrier density of the samples. Figure 15.2 shows the g_e factors measured in a 15 nm GaAs/Al_{0.35}Ga_{0.65}As modulation-doped QW and successfully described by equation (15.8).

The presence of zero-field splitting has been suggested theoretically [17]. Experimentally, the zero-field splitting seems to be present when the B -dependent Zeeman splitting had been fitted with a linear relationship to B [15], but it was found negligible in the quadratic fitting [16]. These experiments are explained successfully by an envelope function approach, and the B -dependent part in g_e is ascribed to be caused by k^4 non-parabolicity term [18]. The same approach showed that the inversion-asymmetry-induced k^3 non-parabolicity term gives rise to a small spin splitting at $B = 0$, which is -0.04 meV in a modulation-doped GaAs/Al_{0.3}Ga_{0.7}As heterostructure with an electron concentration $N_s = 10^{11} \text{ cm}^{-2}$ and a background charge density $N_D = 0.5 \times 10^{11} \text{ cm}^{-2}$ [19].

The electron g_e factor in GaAs/Al_{0.3}Ga_{0.7}As QWs with high hole concentrations was determined as a function of well width by measuring the Hanle effect [20]. For 20 nm wells g_e is close to the bulk value (-0.44), and increases for narrower wells passing through zero for well width close to 5.5 nm, as shown

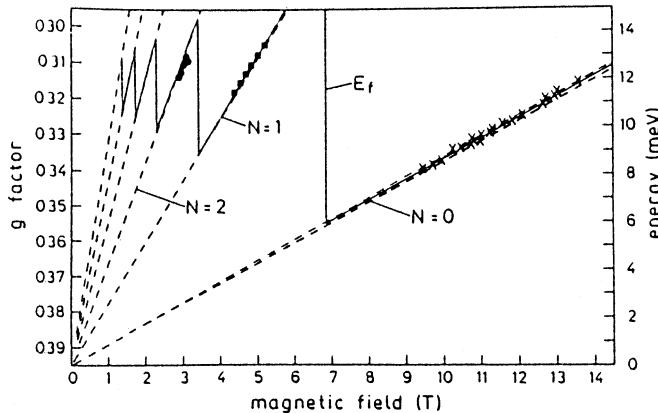


Figure 15.2. g_e factors measured in a 15 nm GaAs/Al_{0.35}Ga_{0.65}As modulation-doped QW. After [16].

in figure 15.3. These observations were explained by three-band $\mathbf{k} \cdot \mathbf{p}$ theory, including the allowance of conduction-band non-parabolicity and wavefunction penetration into the barriers. A similar well width dependence was observed in spin quantum beats in GaAs/Al_xGa_{1-x}As ($x = 0.35, 0.3$ and 0.27) with the well width from 1 to 20 nm [21]. The data are well explained by the one-band approximation using the Kane model [22] including the anisotropy of the g_e factor. The spin splitting measured by spin quantum beats in 25 nm GaAs QW is linear with $g_e = -0.390$ for $B < 8$ T and becomes sublinear for higher fields, as shown in figure 15.4 [23]. This deviation from the linearity was ascribed to the non-parabolicity.

The general properties of excitons in GaAs/Al_xGa_{1-x}As QWs in the presence of magnetic fields up to 10 T were calculated in an effective mass approximation neglecting the exchange effects [24]. In the high-field region the exciton states were described by expansion into Landau-level wavefunctions [25].

If the magnetic energy is smaller than the heavy- and light-hole splitting which is caused by the quantum confinement, the heavy- and light-hole states may be treated separately. For a (001)-grown QW, the heavy-hole exciton Hamiltonian may be written as [8]

$$H_{\text{ex}} = \sum_{i=x,y,z} [\mu_B(g_{e,i}S_{e,i} - g_{h,i}\tilde{S}_{h,i})B_i - c_i S_{e,i}\tilde{S}_{h,i}], \quad (15.9)$$

where the coefficient g_h is related to q and κ in equation (15.5), c is related to a and b in equation (15.7), and the effective heavy-hole spin \tilde{S}_h is defined such that $\tilde{S}_{h,z} = \pm 1/2$ correspond to $J_{h,z} = \pm 3/2$. Among the four basis states $|S_{e,z} = \pm 1/2, \tilde{S}_{h,z} = \pm 1/2\rangle$, the transitions between $|1/2, -1/2\rangle$ and $|-1/2, 1/2\rangle$ states are electric-dipole-allowed and the photons with σ^- and σ^+ polarizations

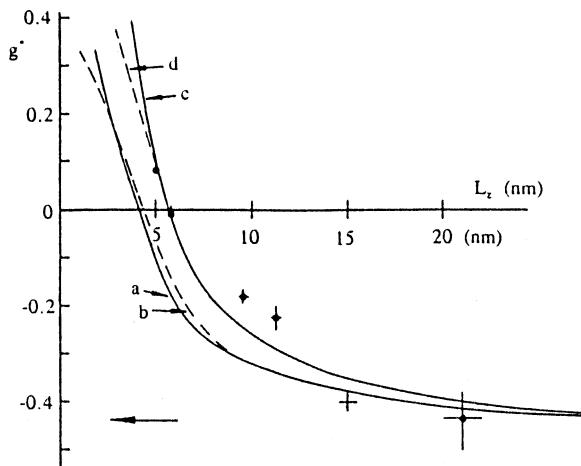


Figure 15.3. g_e factors as a function of well width, determined by the Hanle effect measurements in GaAs/Al_{0.3}Ga_{0.7}As QWs with high hole concentrations. After [20].

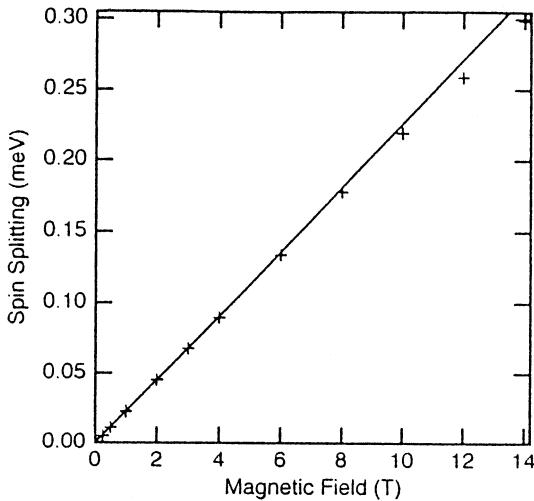


Figure 15.4. Dependence of electron spin splitting on the applied magnetic field. The splitting becomes sublinear for $B > b$ T. After [23].

are emitted, respectively. For magnetic field applied in the z -direction, the energy difference between the σ^+ and σ^- polarizations is

$$\Delta E = -(g_{e,z} + g_{h,z})\mu_B B, \quad (15.10)$$

and the excitonic g -factor may be written as

$$g_{\text{ex}} = -g_e - g_h. \quad (15.11)$$

Unfortunately there exist different definitions of g_{ex} and g_h . Some authors define g_{ex} in terms of the σ^- – σ^+ separation instead of the σ^+ – σ^- separation so that $g_{\text{ex}} = g_e + g_h$ [26, 27]. Since $|1/2, -1/2\rangle$ and $| -1/2, 1/2\rangle$ are $J_z = -1$ and $J_z = 1$ states, respectively, if we define g_{ex} -factor for $\Delta J_z = 1$ as in equation (15.1), the excitonic g_{ex} -factor should be half of that in equation (15.11) [28]. The g_h factor is also defined as

$$H_h = g_h \mu_B B J_{h,z} \quad (15.12)$$

rather than $H_h = -g_h \mu_B B \tilde{S}_{h,z}$ as in equation (15.9). In this case we get [29]

$$g_{\text{ex}} = -g_e + 3g_h. \quad (15.13)$$

In this paper, we use the definition in equation (15.11) if the translation is possible.

Clear σ^+ and σ^- splittings have been observed in magnetoreflectance measurements for a 7.9 nm GaAs/Al_{0.5}Ga_{0.5}As QW, and analysed by the theory of invariants to yield $g_e = -6.8$ and $\kappa = 1.37$ [30]. The excitonic g_{ex} factor was measured as a function of well width by the σ^+ – σ^- splitting in luminescence line below 2 T [26]. The g_{ex} factor is negative for wide QWs, but becomes positive as the well width decreases, passing through zero between 5 and 12 nm, as shown in figure 15.5. The observed splittings are linear in low field (<1 T), but become nonlinear as the field is increased [27]. The behaviour was analysed by an eight-band $\mathbf{k} \cdot \mathbf{p}$ theory and attributed to a spin-dependent field-induced admixture between the light- and heavy-hole valence bands.

Several experiments based on the coherent nonlinear optical measurements have been performed to measure the excitonic g -factor, g_{ex} , of GaAs/Al_xGa_{1-x}As QW. Among these, quantum beat measurements can determine the absolute value of $|g_{\text{ex}}|$ only. An absorption quantum beat study on a step multiple QW (MQW) composed of quantum well layers of (3 nm GaAs + 10 nm Al_{0.1}Ga_{0.9}As) and quantum barriers of 10 nm Al_{0.3}Ga_{0.7}As obtained $|g_{\text{ex}}| = 2.2$ at 4 T [29]. The spin–flip-induced hole burning measurements on a GaAs/Al_{0.3}Ga_{0.7}As MQW of 10 nm each in the widths showed that the excitonic splitting is quadratic in magnetic field up to 6 T, as shown in figure 15.6 [31]. The four-wave-mixing studies for a stepped (3 nm GaAs + 10 nm Al_{0.1}Ga_{0.9}As)/(10 nm Al_{0.3}Ga_{0.7}As) MQW and a (3 nm GaAs)/(15 nm Al_{0.3}Ga_{0.7}As) MQW gave $|g_{\text{ex}}| = 0.76 + 0.016B$ and $|g_{\text{ex}}| = 0.58 + 0.013B$, respectively [32]. The quantum beat spectroscopy was extended to measure the light-hole excitons [28]. The measurements on two GaAs/Al_{0.3}Ga_{0.7}As MQW whose well widths are 3 nm and 12 nm have shown that the heavy-hole splitting decreases with the well width, while that of the light hole exhibits an opposite trend. This study also showed that the Zeeman splitting of the magnetoexcitons related to the higher state heavy-hole is larger than that of the ground-state one.

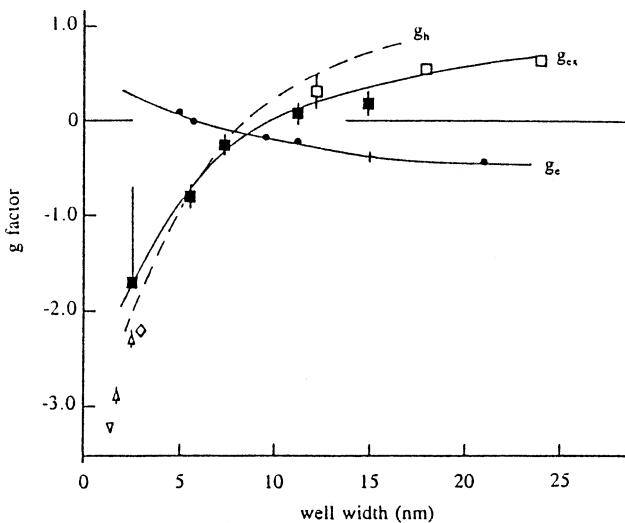


Figure 15.5. Well width dependence of the excitonic g_{ex} factors, after [27]. The sign of g_{ex} of this reference is opposite to equation (15.11). After [27].

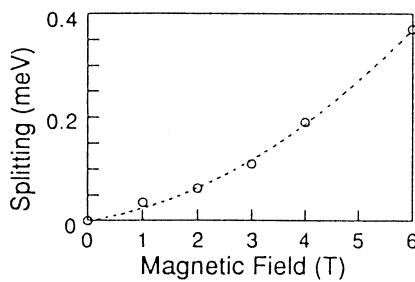


Figure 15.6. Magnetic field dependence of the exciton Zeeman splitting measured by spin-flip-induced hole burning on a (10 nm GaAs)/(10 nm Al_{0.3}Ga_{0.7}As) MQW. After [31].

15.4 GaAs/Al_xGa_{1-x}As quantum dot and quantum wire

The spin states of QD and QWR attract overwhelming interest with the recent progress on spin memory and spin-related quantum computation. However, a limited number of the reports are available on the spin-state study of these structures. A prime difficulty on the study is lack of samples with well-defined structure. In this respect, we have to pay attention to the method of sample preparation in order to understand the physical conditions of the experiment for a given sample. Furthermore, there is very little (if not none) work done on the

quantitative analysis on the QDs' Zeeman splittings. It is because the theory is difficult to formulate and the confining potential is not accurately known.

Most of the studies on GaAs-based QD have been performed on naturally formed samples, where the dots were formed by the width fluctuations in narrow GaAs/Al_xGa_{1-x}As quantum wells. Measurements of the Overhauser effect performed with naturally formed QD samples reported that $g_{\text{ex}} = 1.1$ for the excitonic Zeeman splitting for σ^+ excitation [33]. In this study, a single dot is selected by a 1.5 μm aperture in an aluminum mask in order to avoid the inhomogeneous broadening. In a micro-photoluminescence experiment on naturally formed QDs in 3.5 nm wide GaAs/Al_{0.35}Ga_{0.65}As QW in magnetic field up to 5 T, the Zeeman splittings for the ground-, first-excited and second-excited excitonic states were observed with g_{ex} factors 1.42, 2.55 and -2.04, respectively [34]. Another micro-photoluminescence experiment on a QD sample which is prepared by the laser-induced interdiffusion method of a GaAs/Al_xGa_{1-x}As QW shows that the exciton g_{ex} factor increases as the exciton peak energy increases [35]. This implies that the value of g_{ex} depends on the lateral geometrical size of QDs, since the peak energy is related to the geometrical size of individual dot. It should be noted that, in the QD systems described in the above, the lateral confinement is relatively weak.

An example of somewhat clearly defined zero-dimensionally confined electronic systems is InAs self-assembled QDs (SAQDs). In a magenetotunnelling spectroscopy of a single InAs QD grown within AlAs barrier, the electron g_e factors for the ground-state electron lie between 0.52 and 1.6, depending on the size of dots [36]. These values are quite different from the bulk value of -14.8. The linear Zeeman splitting with quadratic diamagnetic shift up to 10 T was observed in photoluminescence from single InAs/GaAs SAQDs, which was isolated by near-field optical microscope, as shown in figure 15.7 [37]. The observed splittings have no clear dependence on the emission peak energy in the GaAs/InAs/GaAs QDs, while the splittings tends to be larger for higher PL energy in the GaAs/InAs/Al_{0.36}Ga_{0.64}As QDs. The Zeeman splittings of the exciton and biexciton spectral lines were observed in micro-photoluminescence studies on self-assembled In_{0.60}Ga_{0.40}As QDs, which are shown in figure 15.8 [38]. The splittings from both excitons and biexcitons are the same as shown in the inset of figure 15.8, and their origins are interpreted from the exciton level splitting which has a g_{ex} value of 3.02. No detailed quantitative analysis on the Zeeman splittings of QDs is available yet, because the theory is difficult to formulate and the confining potential is not accurately known.

Even fewer reports are available on the Zeeman splitting and the spin state of quantum wires. Circular polarization experiments on a 1.5% compressively strained InAs_{0.48}P_{0.52}/InP QWR prepared by e-beam lithography, wet chemical etching and overgrowth on a single QW show that Zeeman splitting is highly nonlinear to B , as shown in figure 15.9 [39]. The nonlinearity is explained by anisotropic triaxial strain effect and the dimensional cross-over of strain. The Zeeman splittings of a GaAs/AlAs QWR-like superlattice grown on the

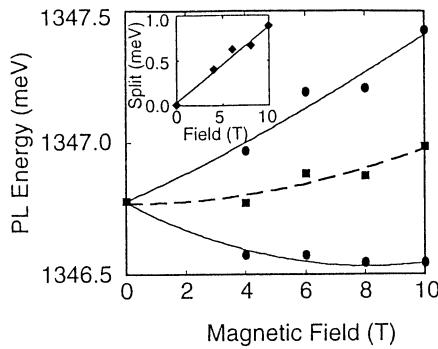


Figure 15.7. Magnetic field dependence of the photoluminescence peak energies from single InAs SAQD. After [37].

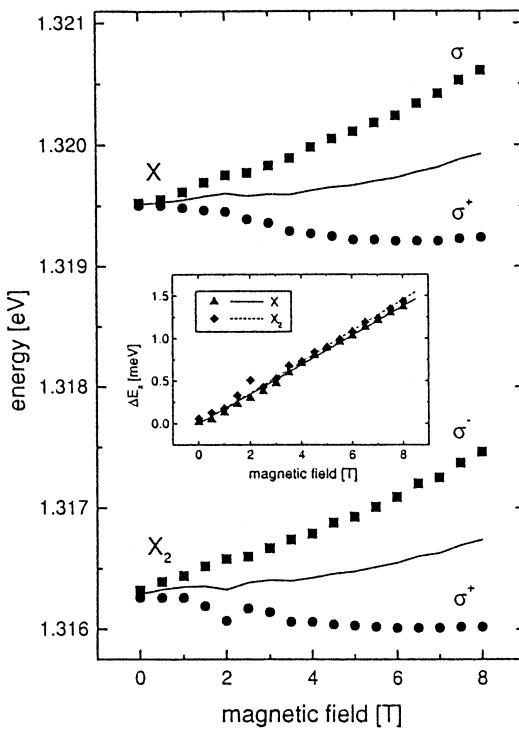


Figure 15.8. Exciton and biexciton energies as a function of magnetic field, as measured by photoluminescence on self-assembled $\text{In}_{0.60}\text{Ga}_{0.40}\text{As}$ QDs. The inset shows the Zeeman splittings of the exciton and biexcitons emission versus magnetic field. After [38].

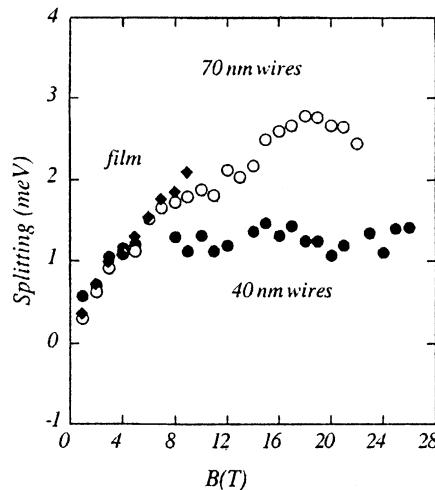


Figure 15.9. Zeeman splitting versus magnetic field on 40 and 70 nm wide InAs_{0.48}P_{0.52}/InP QWR. After [39].

stepped (311) GaAs substrate measured by spin quantum beats show an extreme anisotropy of the electron g_e factor [40]. The $|g_e|$ value changes from 0.56 along the wires to approximately zero perpendicular to the wires, and is consistent with an estimate based on $\mathbf{k} \cdot \mathbf{p}$ theory [22]. The Zeeman splitting becomes sublinear in strong magnetic field region ($B > 6$ T).

In the case of GaAs/Al_{0.5}Ga_{0.5}As QWR-like superlattice grown on a 1° tilted-cut (100) GaAs substrate, the Zeeman splitting has been measured by magneto-photoluminescence with the field applied along the growth axis. The experiments have been performed on two samples of very narrow QWR widths, $l_w \sim 8$ and 10.5 nm, whose Zeeman separations are summarized in figures 15.10(a) and (b), respectively. The two curves in figure 15.10(a) represent the data obtained at two different spots of the same geometry QWR sample. The field dependent Zeeman separation shows that it increases linearly to B at high field $B > B_c$ with an abrupt change of the slope near a critical field B_c . The Zeeman increases as B increases at low field $B < B_c$, but it is difficult to predict any regularity due to the resolution of the measurement. B_c corresponds to the field, in which the Landau orbital radius equals the QWR width, l_w . The Zeeman at the high-field region can be plotted with the relationship,

$$H_e = g_e \mu_B S_z (B_z + B_0) \quad (15.14)$$

where B_z is an applied field along the growth direction and B_0 is a constant field which implied the presence of an internal field (cf equation (15.4)). A typical example of an internal field is $\vec{L} \cdot \vec{S}$ coupling in an atom, and the B_0 related term in equation (15.14) may be indicative of the presence of the spin-Landau orbit coupling. The details on this subject are discussed elsewhere [41].

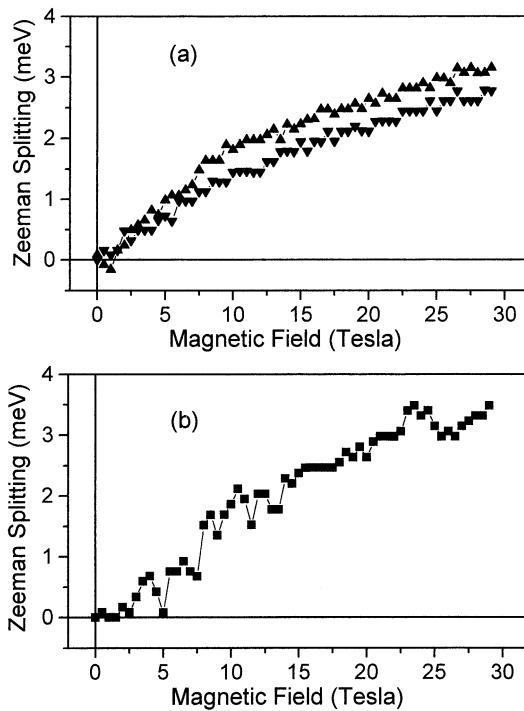


Figure 15.10. Zeeman splittings versus magnetic field on (a) 8 nm and (b) 10.5 nm wide GaAs/AlGaAs QWR superlattice.

15.5 Concluding remarks

Despite the long history on the study of semiconductor quantum structure, the study on the spin state is still in a preliminary stage. In this paper, we try to present a summary on the results obtained by others so far and the results we have obtained from a GaAs/Al_{0.5}Ga_{0.5}As QWR array sample. We have observed that the Zeeman separation increases in steps. The step-like shift is the result of the spin coupling to the Landau field, the field produced by Landau orbital motion.

References

- [1] White H H 1934 *Introduction to Atomic Spectra* (Tokyo: McGraw-Hill)
- [2] Roth L M, Lax B and Zwerdling S 1959 *Phys. Rev.* **114** 90
- [3] Hermann C and Weisbuch C 1977 *Phys. Rev. B* **15** 823
- [4] Oestreich M and Ruhle W W 1995 *Phys. Rev. Lett.* **74** 2315
- [5] Weisbuch C and Hermann C 1977 *Phys. Rev. B* **15** 816
- [6] Hermann C and Weisbuch C 1984 *Optical Reorientation* ed F Meier and B P Zakharchenya (Amsterdam: North-Holland)) p 463

- [7] Luttinger J M 1956 *Phys. Rev.* **102** 1030
- [8] van Kesteren H W, Cosman E C, van der Poel W A J A and Foxon C T 1990 *Phys. Rev. B* **41** 5283
- [9] Altarelli M and Lipari N O 1973 *Phys. Rev. B* **7** 3798
- [10] Cho K, Suga S, Dreybrodt W and Willmann F 1975 *Phys. Rev. B* **11** 1512
- [11] Altarelli M and Lipari N O 1974 *Phys. Rev. B* **9** 1733
- [12] Nam S B, Reynolds D C, Litton C W, Almassy R J, Collins T C and Wolfe C M 1976 *Phys. Rev. B* **13** 761
- [13] Willmann F, Suga S, Dreybrodt W and Cho K 1974 *Solid State Commun.* **14** 783
- [14] Hess K, Bimberg D, Lipari N O, Fischbach J U and Altarelli M 1976 *Proc. 13th Int. Conf. on Phys. of Semicon. (Marves, Rome)* ed G Fumi, p 142
- [15] Stein D, Klitzing K v and Weimann G 1983 *Phys. Rev. Lett.* **51** 130
- [16] Dobers M, Klitzing K v and Weimann G 1988 *Phys. Rev. B* **38** 5453
- [17] Marques G E and Sham L J 1982 *Surf. Sci.* **113** 131
- [18] Lommer G, Malcher F and Rössler U 1985 *Phys. Rev. B* **32** 6965
- [19] Lommer G, Malcher F and Rössler U 1988 *Phys. Rev. Lett.* **60** 728
- [20] Snelling M J, Flinn G P, Plaut A S, Harley R T, Tropper A C, Eccleston R and Phillips C C 1991 *Phys. Rev. B* **44** 11 345
- [21] Hannak R M, Oestreich M, Heberle A P, Rühle W W and Köhler K 1995 *Solid State Commun.* **93** 313
- [22] Ivchenko E L and Kiselev A A 1992 *Sov. Phys. Semicond.* **26** 827
- [23] Heberle A P, Rühle W W and Ploog K 1994 *Phys. Rev. Lett.* **72** 3887
- [24] Bauer G E W and Ando T 1988 *Phys. Rev. B* **37** 3130
- [25] Eric Yang S R and Sham L J 1987 *Phys. Rev. Lett.* **58** 2598
- [26] Lefebvre P, Gil B, Lascaray J P, Mathieu H, Bimberg D, Fukunaga T and Nakashima H 1988 *Phys. Rev. B* **37** 4171
- [27] Snelling M J, Blackwood E, McDonagh C J, Harley R T and Foxon C T B 1992 *Phys. Rev. B* **45** 3922
- [28] Traynor N J, Warburton R J, Snelling M J and Harley R T 1997 *Phys. Rev. B* **55** 15 701
- [29] Carmel O, Shtrikman H and Bar-Joseph I 1993 *Phys. Rev. B* **48** 1955
- [30] Bar-Ad S and Bar-Joseph I 1991 *Phys. Rev. Lett.* **66** 2491
- [31] Wang H, Jiang M, Merlin R and Steel D G 1992 *Phys. Rev. Lett.* **69** 804
- [32] Carmel O and Bar-Joseph I 1993 *Phys. Rev. B* **47** 7606
- [33] Brown S W, Kennedy T A, Gammon D and Snow E S 1996 *Phys. Rev. B* **54** R17 339
- [34] Heller W and Bockelmann U 1997 *Phys. Rev. B* **55** R4871
- [35] Bockelmann U, Heller W and Abstreiter G 1997 *Phys. Rev. B* **55** 4469
- [36] Thornton A S G, Ihn T, Main P C, Eaves L and Henini M 1998 *Appl. Phys. Lett.* **73** 354
- [37] Toda Y, Shinomori S, Suzuki K and Arakawa Y 1998 *Appl. Phys. Lett.* **73** 517
- [38] Kuther A, Bayer M, Forchel A, Gorbunov A, Timofeev V B, Schäfer F and Reithmaier J P 1998 *Phys. Rev. B* **58** R7508
- [39] Notomi M, Hammersberg J, Zeman J, Weman H, Potemski M, Sugiura H and Tamamura T 1998 *Physica B* **249** 171
- [40] Oestreich M, Heberle A P, Rühle W W, Nötzel R and Ploog K 1995 *Europhys. Lett.* **31** 399
- [41] Woo J C *et al* to be published

PART 6

DEVICE APPLICATIONS

Chapter 16

Self-assembling of silicon quantum dots and its application to novel nanodevices

M Hirose

Hiroshima University, Higashi-Hiroshima 739-8527, Japan

E-mail: hirose@sxsys.hiroshima-u.ac.jp

Nanometre-scale silicon quantum dots have been fabricated on ultrathin SiO_2 layers by controlling the early stages of low-pressure chemical vapour deposition from a silane gas. The average size and areal density of quantum dots can be controlled by deposition temperature. The optical absorption edge exhibits a blue-shift as the average dot size decreases and the corresponding shift of the subband energy due to quantum confinement has been confirmed.

A single quantum dot covered with 1 nm thick SiO_2 formed on a silicon substrate offers a double barrier structure consisting of a conducting probe of atomic force microscope/ SiO_2 /a Si dot/ SiO_2 /Si multilayer. Clear negative resistance has been observed at 300 K for the double barrier structures with dot sizes of 2.7–2.8 nm. Quantum-dot floating gate MOS memory devices have been designed and fabricated. Analysis of the observed I – V and C – V characteristics has revealed that single electron charging to each of quantum dots as a floating gate can be achieved.

16.1 Introduction

Nanometre-sized silicon structures have been employed to develop novel Si-based functional devices such as resonant tunnelling diodes [1–3], one-dimensional transport devices [4] and quantum dot memory devices [5–7]. A crucial issue for the room-temperature operation of such Si quantum effect devices is to fabricate well-defined Si structures with a feature size below 3 nm as predicted from the quantization energy or coulombic charging energy. For fabricating Si quantum dot structures without defects and damage, a self-assembling process is thought

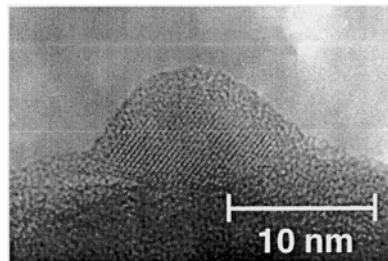


Figure 16.1. Cross-sectional TEM micrograph of a Si dot formed by 36 s deposition on as-grown SiO_2 at 600°C .

to be suitable to achieve a good size uniformity and a high areal density. It has been reported that Si nanocrystallites can be spontaneously grown on SiO_2 by controlling the early stages of low-pressure chemical vapour deposition (LPCVD) [3, 8] and exhibit a blue-shift of the optical absorption edge which is attributable to the quantum confinement effect [9]. This paper describes the formation mechanism of self-assembled Si dot, the quantum confinement effect, and the device applications.

16.2 Self-assembling of silicon quantum dots

The formation of a cap-shaped SiO dot on thermally grown SiO_2 was confirmed at deposition temperatures above 525°C by high-resolution TEM as shown in figure 16.1. It is interesting to note that the lattice images for $\text{Si}(111)$ planes were always observed at a tilt by ~ 30 degrees from the substrate surface normal. This indicates that the growth of Si dots proceeds so that a $\text{Si}(100)$ plane becomes parallel to the substrate surface to minimize the surface free energy.

Figure 16.2 shows Arrhenius plots of areal Si-dot density as evaluated by AFM. At the early stages of the dot formation, the density dramatically increases as the deposition temperature rises to 570°C , with an activation energy of ~ 5.4 eV until reaching a value of $2 \times 10^{11} \text{ cm}^{-2}$, at which it tends to saturate because of the coalescence growth as confirmed by TEM observations. Considering that the $\text{Si}-\text{O}$ bond energy is estimated to be 4.85 eV from the atomization energy of SiO_2 [10], the $\text{Si}-\text{O}$ bond breaking might play a role in the creation of nucleation sites. The size distribution of obtained Si dots evaluated from AFM images can be fitted to a log-normal function [11] as indicated in figure 16.3. It is obvious that the higher temperature deposition provides the broader size distribution. To get an insight into the growth mechanism, the average dot diameter or dot height was determined at each deposition temperature by fitting a measured size distribution to a log-normal function, and the activation energy of dot formation was assessed from the Arrhenius plot as shown in figure 16.4. The dot diameter increases as the deposition temperature

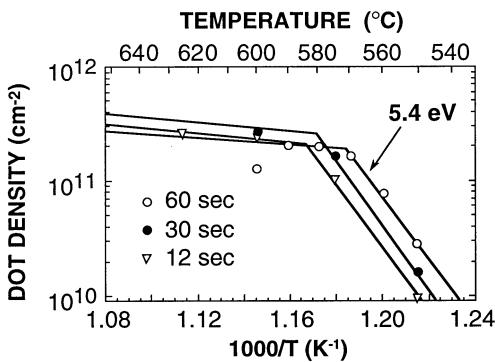


Figure 16.2. Deposition temperature dependence of areal Si-dot density on as-grown SiO_2 . The deposition time was varied from 12–60 s.

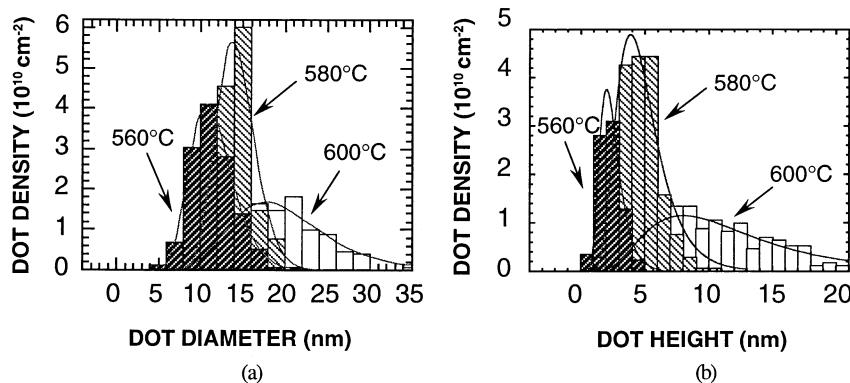


Figure 16.3. Distributions of (a) Si-dot diameters and (b) dot heights obtained by AFM images for samples prepared at different deposition temperatures for 60 s. The full curves denote log-normal functions well fitted to the measured distributions.

increases with an activation energy of 0.82 eV which corresponds to the SiH_4 decomposition energy on c-Si [12]. On the other hand, the activation energy for the dot height is found to be 2.2 eV, being almost identical to the Si cohesive energy of 2.14 eV [13] or 2.45 eV [14]. These results imply that the dot diameter is controlled by the thermal decomposition process of SiH_4 on the Si nucleation sites created on SiO_2 and the height is rate-limited by the cohesive reaction of absorbed precursors. In addition, it is likely that reactive species diffusing on the surface mainly contribute to the dot growth as described later. In a thermodynamic concept for nucleation, it is energetically favourable for clusters to exceed a critical size to keep the further growth, while clusters smaller than the critical size decompose. According to a simple nucleation theory [15], the critical radius

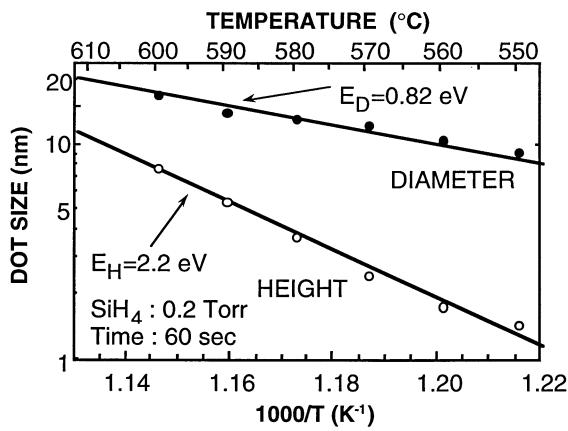


Figure 16.4. Deposition temperature dependences of average Si-dot diameter and dot height on as-grown SiO_2 . The deposition time was 60 s.

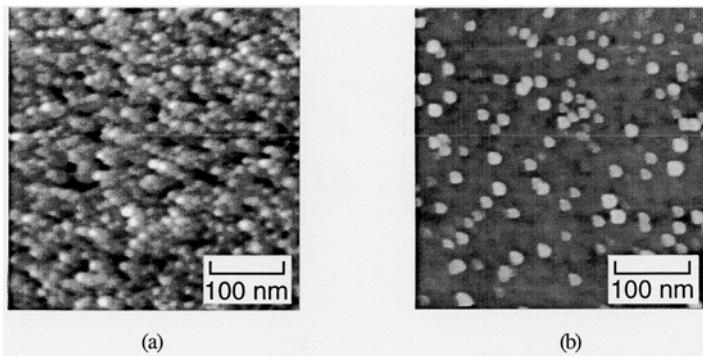


Figure 16.5. AFM images of Si dots formed at 565°C for 60 s on SiO_2 surfaces (a) with and (b) without HF treatment.

for a stable Si cluster can be estimated to be 0.23 nm that is slightly smaller than the size of the crystalline silicon unit cell.

Significant enhancement of a real Si-dot density and resulting decrease in dot size are caused by treating the SiO_2 surface in a 0.1% HF solution before LPCVD as shown in figures 16.5 and 16.6. It should be noted that the size distribution becomes appreciably narrower compared with the case of growth on as-grown SiO_2 as indicated in figure 16.6. This implies that the uniform nucleation proceeds on the HF-treated SiO_2 surface. The Fourier transform infrared attenuated total reflection (FT-IR-ATR) spectrum of the HF-treated SiO_2 exhibits absorption bands centred at ~ 3450 and $\sim 3250 \text{ cm}^{-1}$, indicating that the surface is terminated by OH bonds. Taking into account the fact that the surface

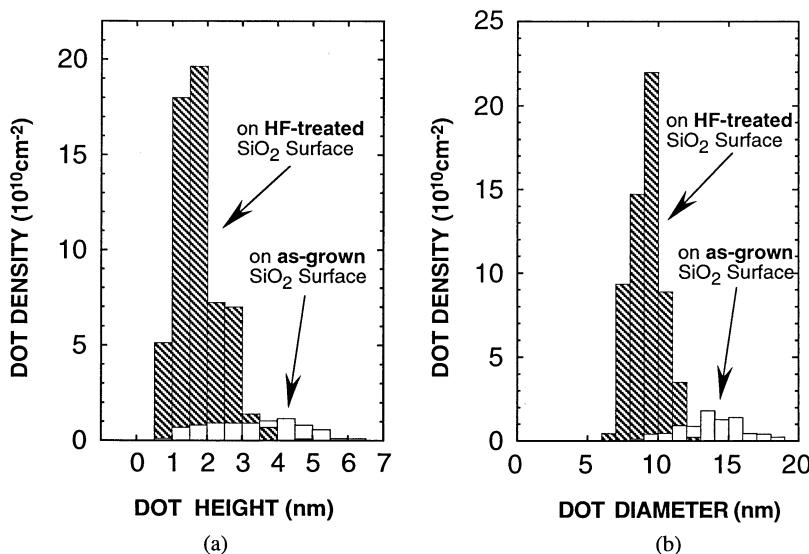


Figure 16.6. Distributions of (a) Si-dot heights and (b) dot diameters for samples shown in figure 16.5.

silanol groups ($\equiv \text{Si}-\text{OH}$) can act as nucleation sites for reactive species such as SiH_2 , the result of figure 16.6 can be explained by the OH termination of the SiO_2 surface with HF treatment.

16.3 Quantum confinement in silicon dot

The subband formation in nanometre Si dots has been suggested by the blue-shift of the optical absorption edge with decreasing average dot size [8, 9, 16]. However, the dot size dependence of the energy bandgap as determined by the onset of absorption spectrum has not been well explained by theories [17–22]. For a better understanding of the quantum size effect in Si dots, direct measurement of the subband energy is needed. By using high-resolution x-ray photoelectron spectroscopy (XPS), the size dependence of the valence band maximum for self-assembled Si quantum dots formed on SiO_2 was measured.

16.3.1 Valence band spectra and charging effect

The high-resolution XPS measurements were carried out by ESCA-300 (Scienta Instruments), using monochromatized Al Ka radiation (1486.6 eV) with an acceptance angle of 3.3° . The instrument energy resolution defined by the full width at half maximum (FWHM) of Ag Fermi edge is 0.3 eV for 75 eV pass energy.

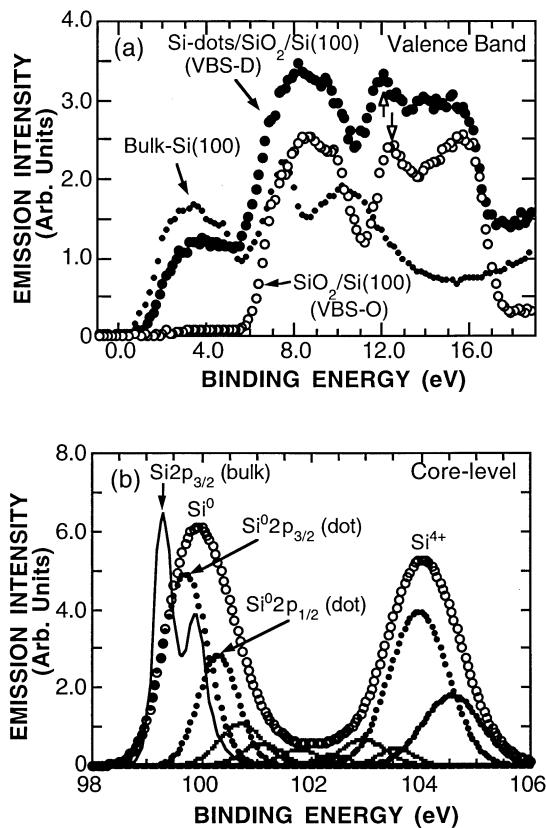


Figure 16.7. The valence band spectra for bulk-Si, Si dots/5.5 nm SiO₂/Si and 5.5 nm SiO/Si structures for a take-off angle of 10° (a), and the Si2p core level spectra for bulk-Si (full curve) and the Si-dots/SiO₂/Si structure (open circles) where the suboxide signals are also indicated (b). The spectra were not corrected for charging effect. The Si dots were deposited at 580 °C.

The valence band and Si2p core level spectra for a Si-dots/5.5 nm SiO₂/Si(100) structure and a hydrogen-terminated bulk-Si(100) surface are compared in figure 16.7. The valence band spectrum for a 5.5 nm SiO₂/Si structure (denoted by VBS-O) is also shown in figure 16.7(a). By comparing the valence band spectra for the Si dots/SiO₂/Si structure (denoted by VBS-D) with the VBS-O, the feature of the VBS-D below 5 eV is referred to the energy band arising from Si3p states of Si dots. The Si substrate signal through 5.5 nm thick SiO₂ is negligible at a photoelectron take-off angle of 10°. The valence band edge of the Si-dots exhibits a shift of about 0.6 eV to the higher binding energy from that of bulk-Si. This shift arises not only from the dot charging effect

but also from the subband formation in the dot. It should be noted that the VBS-O and VBS-D in the energy range from 11–17 eV originate from the Si3p–O2p and Si3s–O2p bonding orbitals in the oxide layers [23]. The valence band peak position at about 12 eV for the VBS-D is located at slightly lower binding energy side than the case of the VBS-O as indicated by the arrows near 12 eV. The reason for this is discussed below.

The Si2p spectrum shown in figure 16.7(b) for the Si-dots/SiO₂/Si structure is composed of the two strong components for the chemically shifted Si⁴⁺ signal from the SiO₂ layer and for metallic Si⁰ signal from the dots. Note that a contribution to the Si⁰ peak from the Si substrate through the 5.5 nm thick SiO₂ layer is negligible at take-off angles of 5° and 10°. The Si⁰2p_{3/2} signal for the dots is broad because of the size distribution and shifted to higher binding energy by about 0.5 eV in figure 16.7(b) with respect to the Si2p_{3/2} bulk signal mainly due to the charging effect of silicon dots. Because of the dielectric properties of the oxide layer and the undoped Si-dots, the emission of photoelectrons during XPS measurements creates positive charges in the SiO₂ layer and presumably also in the dots, which induce the corresponding spectral shift to the higher binding energy side. The induced positive charge density in the oxide surface layer and the dots should be kept constant under steady-state conditions because the stable spectra were obtained, but it depends on the x-ray power density on the surface. As for the Si dots/SiO₂/Si structure, in which the SiO₂ layer are differentially charged during the measurements. Therefore, the average charge density in the SiO₂ layer of the Si dots/SiO₂/Si structure is smaller than that in the same thick SiO₂ layer of the SiO₂/Si structure under equivalent measurement conditions. This is the reason for the smaller shift of the valence band spectrum for the Si-dots/SiO₂/Si structure (VBS-D) with respect to that for the SiO₂/Si structure (VBS-O) as shown in figure 16.7(a).

In order to reveal the quantum size effect in nanometre Si dots, we focus attention on the valence band edge and the Si2p core-level peak for the Si-dots. Both the charging effect and the quantum confinement in the Si dots result in an energy shift of the valence band spectrum for the dots. We have obtained the same energy difference between the valence band maximum (VBM) and Si2p peak energy for the bulk-Si(100) and for the Si(100) substrate of a 3.2 nm SiO₂/Si(100) structure in which there exists the oxide charging effect, indicating that the influence of the escape depth difference between the photoelectrons from the valence band and Si2p core level on the energy shift is negligible. Therefore, it is reasonable to conclude that the charging effect causes almost the same amount of energy shift for both the valence band top and Si2p core level of the dots. Consequently, the dot charging effect for the Si-dots/SiO₂/Si structure can be eliminated by employing the Si2p core level for the dots as a reference energy. Unlike the valence band edge energy, the core level energy is not influenced by the quantum size effect in Si dots because the wavefunctions are strongly localized. To keep the dot charging effect identical during the XPS measurements of the Si2p core level and valence band spectra, each set of the two spectra was

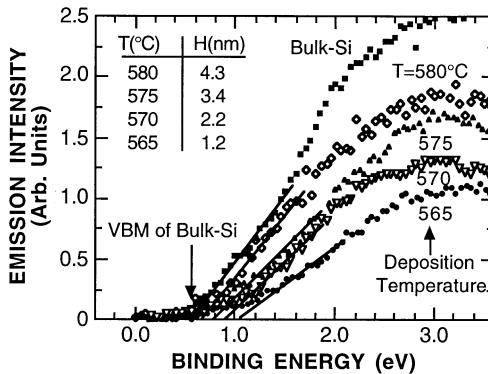


Figure 16.8. The valence band edge spectra for bulk Si and Si quantum dots measured at a take-off angle of 5°. The valence band spectra are aligned by taking $\text{Si}2\text{p}_{3/2}$ peaks for bulk Si and Si dots as energy references. The deposition temperature (T) and the corresponding average dot height (H) are tabulated in the inset.

obtained without any change in the measurement conditions. Since the $\text{Si}2\text{p}$ core level photoelectron peak energy from the Si dots was used as a reference, the average Coulombic charging effect of the dots could be cancelled out as a difference between the $\text{Si}2\text{p}$ reference energy and the valence band edge. Thus, the measured energy shift of VBM with respect to that in bulk-Si can be attributed to the quantum size effect.

In order to determine the energy separation between the VBM and core level in the Si dots, we encountered the problem that the Si^02p spectral shape for dots is significantly broader than that for bulk-Si. This arises from the differential charging effect associated with the dot size distribution. Similar to the $\text{Si}2\text{p}$ for bulk-Si, all the $\text{Si}2\text{p}$ signals for Si dots on SiO_2 were deconvoluted into the $\text{Si}2\text{p}_{3/2}$ and $\text{Si}2\text{p}_{1/2}$ spin-orbit doublet which are separated by 0.61 eV with the $\text{Si}2\text{p}_{1/2}/\text{Si}2\text{p}_{3/2}$ intensity ratio of 0.5. The chemical shifts for suboxides signals Si^{1+} , Si^{2+} , Si^{3+} and silicon dioxide Si^{4+} with respect to the metallic Si state Si^0 were assumed to be the same as a previous study [24]. Taking the binding energy of $\text{Si}2\text{p}_{3/2}$ for bulk Si as the energy reference for all the other $\text{Si}2\text{p}$ signals from Si-dot/ SiO_2 structures, the energy scale of the valence band spectra has been corrected by shifting each spectrum by a value given by the difference between the respective $\text{Si}^02\text{p}_{3/2}$ peak position for the dots and that for bulk Si to eliminate the charging effects in Si dots. Thus, the valence band spectra for Si-dots/ SiO_2 structures with different dot sizes can be directly compared with the bulk Si signal by eliminating the charging effect.

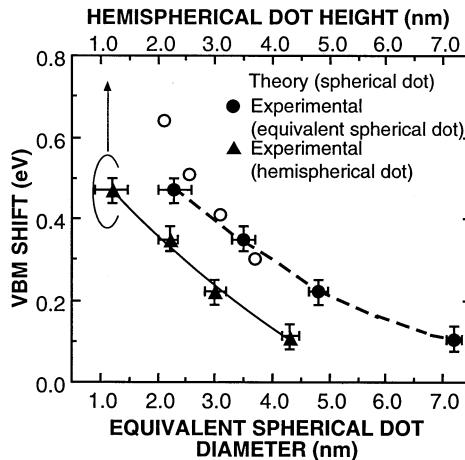


Figure 16.9. A comparison between the measured energy shift of the valence band maximum in Si quantum dots and theoretical result in [22]. The equivalent diameter of the hemispherical dot is defined as the diameter of the sphere with the same volume as the dot. The broken curve should be compared with the theory (open circles).

16.3.2 Comparison with theory

As shown in figure 16.8, the valence band spectrum shifts towards the higher binding energy by decreasing the average dot height from 4.3 nm to 1.2 nm. By linearly extrapolating the leading edge of the valence band signal to zero intensity, the VBM positions were determined for bulk Si and Si dots with different average heights. The extrapolated VBM energy for Si quantum dots is plotted as a function of the dot height as shown in figure 16.9. The height of hemispherical dot is an important dimension to assess the quantum size effect, while theoretical calculations were made mostly for spherical dots. We therefore introduce an equivalent sphere diameter for the hemispherical dot, which is defined as the diameter of a sphere with the same volume as the dot to preserve the total number of electronic states. Thus, the VBM energy for the corresponding equivalent dot diameter is also represented in figure 16.9. The non-uniform charging effect will occur in the Si dots because of the size distribution. This results in broadening of the Si2p spectral shape and the tailing of the valence band spectrum. Therefore, only the average values as measured at the core level peak positions can be used as an energy reference for data processing. As a matter of fact, the uncertainty of peak position (± 20 meV) due to the non-uniform charging effect is rather small compared to the magnitude of VBM shifts (≥ 100 meV) that could be attributed to the quantum confinement. Almost the same results as shown in figures 16.8 and 16.9 are obtained at the take-off angle of 10° , indicating that the charging effect in dot and SiO_2 has been properly eliminated.

Theoretical studies on the electronic structures and optical properties of Si quantum dots are not very consistent with each other [17–22]. Based on these calculations, the enlargement of the bandgap (ΔE) in a Si quantum dot with respect to the energy bandgap of bulk-Si could be expressed as: $\Delta E \propto d^{-a}$, where d is the diameter of a spherical dot and the exponent a varies from 1 to 2, depending on calculation methods [17, 20, 21]. Wang and Zunger [22] have compared their calculated result as obtained by empirical pseudopotential plane wave theory with other calculations [18–21], giving a fitted related of ΔE (eV) = $88.34d^{-1.37}$ (d is in Å) for Si quantum dots. Changes in the VBM energy for spherical dots with different sizes have also been calculated in [22] as indicated in figure 16.9. There exists a significant difference between the dependence of the calculated VBM energy shift on spherical dot diameter and that of experimental VBM shift on hemispherical dot height, while the difference becomes small if replacing the hemispherical dot height by the diameter of an equivalent sphere as shown by the broken curve in figure 16.9. Therefore, the experimental VBM shifts are consistent with the theoretical values if considering the difference between real dot shape and the theoretical model. To obtain the quantum shift for the conduction band edge, other methods, such as x-ray absorption spectroscopy [25] could be employed. It should be noted that almost the same energy shift of the valence band edge and the conduction band edge in silicon quantum dots has been predicted by the calculation in [22].

16.4 Resonant tunnelling through a Si quantum dot

The resonant tunnelling devices consisting of Si/SiO₂ double-barrier structures could be integrated on a silicon integrated circuit chip in future to add a new function to the system. Better understanding of tunnel current through ultrathin SiO₂ thermally grown on Si and the controlled formation of nanometre-scale Si quantum dots will enable us to realize the resonant tunnelling devices operating at 300 K.

16.4.1 Device fabrication

Silicon quantum dots were fabricated on *cz*, n⁺Si(100) substrates with a resistivity of 1×10^{-3} – 1×10^{-2} Ω cm. After the Si wafer cleaning in an NH₄OH : H₂O₂ : H₂O = 0.15 : 3 : 7 solution at 80 °C followed by dilute HF treatment, a 3 nm thick SiO₂ layer was grown at 900 °C in 2% O₂ + N₂ gas mixture. The hemispherical Si quantum dots were spontaneously grown on the SiO₂ layer by LPCVD using pure silane at a pressure of 0.2 Torr.

For fabricating SiO₂/Si-dot/SiO₂ double-barrier structures, the 3 nm thick SiO₂ layer on n⁺Si was uniformly etched-back to 1 nm in thickness by 0.1% HF with an etch rate of 0.35 nm min⁻¹ [26]. The SiO₂ layer thickness was determined by analysing the Si2p x-ray photoelectron spectrum. After the Si dot formation, the 1 nm thick SiO₂ as a top barrier layer was formed by oxidation of the dot

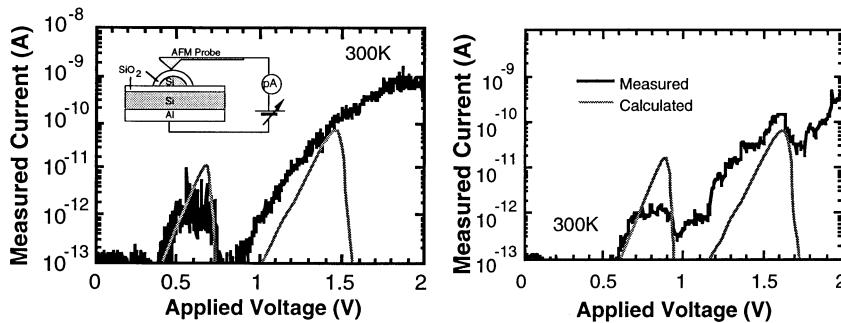


Figure 16.10. Measured I - V characteristics for Si/SiO₂ double-barrier structures with a Si dot height of (a) 2.7 nm and (b) 2.8 nm, respectively. The broken curves indicate the calculated characteristics for corresponding devices structures.

at room temperature or 800 °C to complete Au-coated AFM tip/1 nm SiO₂/a Si dot/1 nm SiO₂/n⁺Si double-barrier structures.

16.4.2 Resonant tunnelling characteristics

The typical I - V characteristics of the double-barrier structures with a single quantum dot are shown in figure 16.10, where voltage was swept with a step of 2 mV in (a) and 5 mV in (b) for every 0.1 s,. A gold-coated Si tip with a spring constant of 0.22 N m⁻¹ was used as a top electrode as illustrated in the inset of the figure. The drift of the AFM image was less than 5 nm min⁻¹. For reproducible I - V measurements, an AFM tip apex with a diameter of 100 nm was employed. Since the average distance among the Si dots is about 15 nm, tunnel current through a single Si dot can be measured within a few minutes before the tip touches the neighbouring Si dots as a consequence of the drift. An n⁺-Si substrate was negatively biased to avoid the anodic oxidation of the Si dot surface by physisorbed water on SiO₂ [27]. In figure 16.10(a), the current bump or the negative conductance are clearly observable around 0.6 V with a peak current-to-valley ratio as high as 10 although the second peak expected at 1.55 V is not clear. Increases in the leakage current through the 1 nm thick SiO₂ barrier sometimes smear out the resonance current for the second peak. There exist the SiO₂/Si interface states and traps in the SiO₂ barrier layers which might also modify the leakage current. On the other hand, in figure 16.10(b) the current bumps and negative conductance are observable at both 0.75 and 1.55 V. The grey curves in the figure indicate the calculated I - V characteristics for the double-barrier structures with a Si dot height of (a) 2.7 nm or (b) 2.8 nm. The resonant tunnelling current has been calculated by using a transfer matrix approach [28, 29], in which the tunnelling electron effective mass, the electron barrier height at the SiO₂/Si interface and that at the Au/SiO₂ interface are set to be 0.3m₀ [30], 3.25 eV [30]

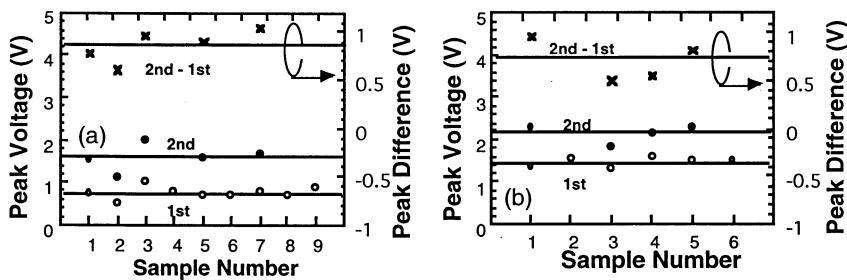


Figure 16.11. Measured voltages where the first and second current peaks or bumps appear and the voltage differences between the second and first peak for double-barrier structures with dot height of 2–3.5 nm. The top oxide layer is native oxide in (a), while it is thermal oxide in (b). The full curves correspond to the calculated peak voltages.

and 4.1 eV [31], respectively. For fitting the calculated I – V curve to the measured one, a flat-band voltage shift due to the workfunction difference and the oxide charge was taken into account. The calculated current level might include an error of an order of magnitude or more because of the fluctuation in the SiO_2 thickness and the probe-apex contact area.

The voltages at which the resonant tunnelling current peaks or shoulders are observed, are summarized in figure 16.11 together with the voltage difference between the first and second peak for double-barrier structures whose dot heights are in the range 2–3.5 nm. The top SiO_2 layer for the nine different double barrier structures (sample number 1–6) in figure 16.11(b) is thermally grown SiO_2 . The peak voltage for the devices with a top native oxide barrier exhibits a shift toward the lower voltage compared with the case of thermal oxide presumably because the native oxide layer contains positive charges. The small variations in the peak voltage for each device structure could be attributed to the difference in the dot size and the influence of very few interface states. The capacitance of $\text{Au}/1\text{ nm } \text{SiO}_2/\text{Si-dot}/1\text{ nm } \text{SiO}_2/\text{n}^+/\text{Si}(100)$ is estimated to be 1.4 aF, yielding the charging energy of about 60 meV that is too small to reveal the Coulomb staircase in I – V curves.

16.5 Silicon quantum dot memory

MOS devices with an array of nanometre silicon dots as a floating gate have been fabricated, and their memory operations associated with electron charging to the dots have been demonstrated at room temperature [6, 7, 32, 33]. A unique hysteresis in capacitance–voltage (C – V) characteristics of the MOS structures has been interpreted in terms of electron charging to the Si quantum dot floating gate [33]. A key question is how many electrons can be stably stored in one quantum dot under the memory operation. In order to answer this question

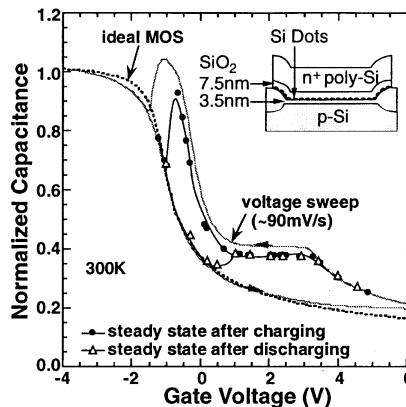


Figure 16.12. $C-V$ characteristics of a Si quantum dot floating gate MOS capacitor with a gate area of $2.5 \times 10^{-3} \text{ cm}^2$. The measurement frequency was 100 kHz. The steady-state $C-V$ curve was obtained from the transient response under constant bias.

charged states of a Si quantum dot (QD) floating gate in the MOS structure have been studied by analysing the steady state and transient current–voltage ($I-V$) and $C-V$ characteristics [34].

16.5.1 Device fabrication

Single-crystalline Si quantum dots were self-assembled on 3.5 nm thick SiO_2 by LPCVD of pure SiH_4 at 580°C [35]. After the first-layer Si dot array formation a ~ 1 nm thick oxide layer was grown and the second dot array was deposited under the same conditions. The total dot density was $\sim 5 \times 10^{11} \text{ cm}^{-2}$ without any significant change in the dot size distribution as measured by AFM. The average dot height and diameter were 5 and 10 nm, respectively. The surface of the second Si dot layer was covered by ~ 1 nm thick oxide. Subsequently, a 3.3 nm thick amorphous Si layer was grown over the Si quantum dots by LPCVD at 440°C and fully oxidized in dry O_2 at 1000°C to form a 7.5 nm thick control oxide. No significant change in the surface morphology was observed in this oxide layer fabrication. Finally, 300 nm thick n^+ poly-Si gates were fabricated. The cross sectional view of a Si QD floating gate MOS capacitor is illustrated in the inset of figure 16.12.

16.5.2 Memory characteristics of Si quantum dot floating gate MOS structures

The measured $C-V$ characteristics of a Si QD floating gate MOS capacitor are shown in figure 16.12. When the gate bias was swept over the range between -4 V and $+6$ V, the $C-V$ curve showed the unique hysteresis which arises

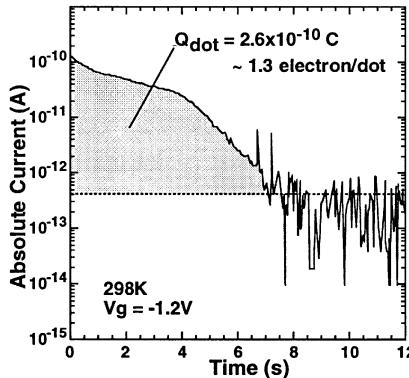


Figure 16.13. Temporal change in discharge current from the charged Si dots. The bias was settled from 0 to -1.2 V at time zero. The number of electrons retained in one dot is evaluated to be ~ 1.3 .

from electron charging to the dots through the 3.5 nm thick oxide. The steady state $C-V$ curve after electron charging at 6 V or after discharging at -4 V was determined from the transient capacitance as described in a previous paper [33]. The charged and discharged states are maintained longer than 10^3 s at gate voltages between -0.7 V and $+0.7\text{ V}$ (figure 16.1) and write/erase cycles can be repeated, indicating that electrons are stably stored in the Si dots. By analysing the flat-band voltage shift it is estimated that about 1.2 electrons per dot exist as the floating gate charge. At gate voltages higher than $+0.7\text{ V}$ the steady state $C-V$ curve after charging is identical to that after discharging. This is because the tunnel injection current from the substrate to the dots balances the dots-to-substrate backward tunnelling current or the dots-to-gate emission current at a given gate voltage, and the number of stored electrons per dot determines the capacitance value. On the other hand, at gate voltages lower than -0.7 V , electrons in the Si dots are completely released and the device stays in the discharged state or the memory charge is erased.

To evaluate the total charge stably stored in QDs around zero bias, the external gate bias was kept at 0 V after writing operation and switched to -1.2 V (erased state), and the discharge current was measured as shown in figure 16.13. The non-exponential current decay might be due to the coulombic interaction among electrons stored in the quantum dots. By integrating the transient current, the total dot charge Q_{dot} is obtained to be $2.6 \times 10^{-10}\text{ C}$, which corresponds to the number of retained electrons per dot being ~ 1.3 . This is consistent with the value estimated from the flat-band voltage shift of the $C-V$ curve when the charging energy of a quantum dot is taken into account.

A memory transistor with a quantum dot floating gate has been fabricated

by using the similar fabrication process as the MOS capacitors, and the detailed memory operation will be reported elsewhere.

16.6 Summary

A self-assembled Si quantum dot layer on SiO_2/Si substrates has exhibited a shift of the valence band edge due to carrier confinement, and the extent of the confinement is a function of the average dot size. Resonant tunnelling through a single quantum dot is confirmed at 300 K. If the dot size control becomes possible, a quantum dot array with a common gate electrode will give the similar resonant tunnelling characteristics as the case of a single quantum dot. Memory characteristics of Si quantum dot floating gate MOS structures have indicated that each quantum dot could act as a memory site where a single electron is stably stored. It is likely that the storage of more than one electron per dot may be achieved by properly designing the dot size and the oxide layer thickness of the MOS structure. The size and position control of silicon dot is certainly an important challenge for the future.

Acknowledgment

Part of this work has been supported by the Core Research for Evolution Science and Technology (CREST) of the Japan Science and Technology Corporation (JST).

References

- [1] Hirose M, Morita M and Osaka Y 1977 *Japan. J. Appl. Phys.* **16** Suppl. 16-1, 561
- [2] Yuki K, Hirai Y, Morimoto K, Inoue K, Niwa M and Yasui J 1995 *Japan. J. Appl. Phys.* **34** 860
- [3] Fukuda M, Nakagawa K, Miyazaki S and Hirose M 1997 *Appl. Phys. Lett.* **70** 2291
- [4] Morimoto K, Hirai Y, Yuki K and Morita K 1996 *Japan. J. Appl. Phys.* **35** 853
- [5] Yano K, Ishii T, Hashimoto T, Kobayashi T, Murai F and Seki K 1994 *IEEE Trans. Electron Dev.* **41** 1628
- [6] Tiwari S, Rana F, Hanafi H, Hartstein A, Crabbè E F and Chan K 1996 *Appl. Phys. Lett.* **68** 1377
- [7] Welses J J, Tiwari S, Rishton S, Lee K Y and Lee Y 1997 *IEEE Electron Dev. Lett.* **18** 278
- [8] Nakajima A, Sugita Y, Kawamura K, Tomita H and Yokoyama N 1996 *Japan. J. Appl. Phys.* **35** L189
- [9] Shiba K, Nakagawa K, Ikeda M, Kohno A, Miyazaki S and Hirose M 1997 *Japan. J. Appl. Phys.* **36** L1279
- [10] Sanderson R T 1976 *Chemical Bonds and Bond Energy* ed E M Loeb (New York: Academic) p 135

- [11] Irani R R and Callis C F 1963 *Particle Size Measurement: Interpretation and Application* (New York: Wiley)
- [12] Joyce B A, Brandley R R and Booker G R 1967 *Phil. Mag.* **15** 1167
- [13] Philips J C 1973 *Bonds and Bands in Semiconductors* ed A M Alper, J L Margrave and A S Nowick (New York: Academic) p 52
- [14] Harrison W A 1980 *Electric Structure and the Properties of Solids—The Physics of the Chemical Bond* ed P Renz and K Sargent (San Francisco, CA: Freeman) p 171
- [15] Jones D W, Kaldas E, Mlavsky A I, Shaw D W and Wolff G A 1974 *Crystal Growth Theory and Techniques* ed C H L Goodman (New York: Plenum) p 55
- [16] Yamada Y, Orii T, Umezawa I, Takeyama S and Yoshida T 1996 *Japan. J. Appl. Phys.* **35** 1361
- [17] Takagahara T and Takeda K 1992 *Phys. Rev. B* **46** 15 578
- [18] Ren S Y and Dow J D 1992 *Phys. Rev. B* **45** 6492
- [19] Proot J P, Delerue C and Allan G 1992 *Appl. Phys. Lett.* **61** 1948
- [20] Delley B and Steigmeier E F 1993 *Phys. Rev. B* **47** 1397
- [21] Hirao M 1994 *Mater. Res. Soc. Symp. Proc.* **358** 3
- [22] Wang L W and Zunger A 1996 *Semiconductor Nanoclusters* ed P V Kamat and D Meisel (Amsterdam: Elsevier Science) p 161
- [23] Alay J L and Hirose M 1997 *J. Appl. Phys.* **81** 1606
- [24] Himpsel F J, McFeely F R, Taleb-Ibrahimi A and Yarmoff J A 1988 *Phys. Rev. B* **38** 6084
- [25] He J, Klug D D, Tse J S, Ratcliffe C I and Preston K F 1997 *Appl. Phys. Lett.* **71** 3149
- [26] Fukuda M, Bjorkman C H, Yamazaki T, Miyazaki S and Hirose M 1994 *Proc. 2nd Int. Symp. on Ultra-Clean Processing of Silicon Surfaces* (Leuven: Academische Coöpperatief) p 297
- [27] Yasutake M, Ejiri Y and Hattori T 1993 *Japan. J. Appl. Phys.* **32** L1021
- [28] Tsu R and Esaki L 1973 *Appl. Phys. Lett.* **22** 562
- [29] Kalotas T M and Lee A R 1991 *Am. J. Phys.* **59** 48
- [30] Yoshida T, Imafuku D, Alay J L, Miyazaki S and Hirose M 1995 *Japan. J. Appl. Phys.* **34** L903
- [31] Sze S M 1981 *Physics of Semiconductor Devices* (New York: Wiley) p 278
- [32] Guo L, Leobandung E and Chou S Y 1996 *Technical Digest of 1996 Int. Electron Device Meeting (San Francisco)* p 955
- [33] Kohno A, Murakami H, Ikeda M, Miyazaki S and Hirose M 1997 *Extended Abstracts of the 1997 Int. Conf. on Solid State Devices and Materials (Hamamatsu)* p 566
- [34] Kohno A, Murakami H, Ikeda M, Nishiyama H, Miyazaki S and Hirose M 1998 *Extended Abstracts of the 1998 Int. Conf. Solid State Devices and Materials (Hiroshima)* p 174
- [35] Nakagawa K, Fukuda M, Miyazaki S and Hirose M 1997 *Mater. Res. Soc. Symp. Proc.* **452** 243

Chapter 17

Quantum devices based on III–V compound semiconductors

H Hasegawa¹, H Fujikura and H Okada

*Research Center for Interface Quantum Electronics (RCIQE)
and Graduate School of Electronics and Information
Engineering, Hokkaido University, Kita-13, Nishi-8, Sapporo
060-8628, Japan*

¹ *E-mail: hasegawa@rciqe.hokudai.ac.jp*

Scaling down of feature sizes into the nanometre range is a common trend in advanced silicon and compound semiconductor devices, and the progress of the ‘nanofabrication’ technology along this line has opened up exciting possibilities of constructing novel quantum devices whose operations are directly based on the quantum mechanics. This paper discusses the present status and key issues of the compound semiconductor quantum devices, introducing and reviewing recent results obtained by the author’s group at the Research Center for Interface Quantum Electronics (RCIQE). In quantum devices, wave-particle motions of individual electrons are controlled by artificial quantum structures such as quantum wells, quantum wires, quantum dots and single and multiple tunnelling barriers so as to realize devices with new functions and higher performances. Since electrons manifest predominantly as either wave-nature or particle-nature depending on their environments, one can conceptually envisage two kinds of nanoelectronics in the quantum regime, i.e. ‘quantum wave electronics’ and ‘single electronics’. After giving a brief overview on quantum devices, the paper discusses specific details of III–V quantum structure formation by molecular beam epitaxy (MBE), formation and control of III–V nanostructure surfaces and interfaces, novel Schottky in-plane gate quantum wave devices and, finally, novel Schottky in-plane gate and wrap-gate single electron devices. A particular emphasis is paid on single electronics because of its promising prospects.

17.1 Introduction

Scaling down of feature sizes into the nanometre range seems to be a common trend in advanced semiconductor devices including silicon and compound semiconductor devices [1]. For example, production-level realization of the minimum feature size of 70 nm is targeted at the year 2010, according to the ‘Roadmap’ for the Si LSI industry. Compound semiconductor electronic devices such as GaAs MESFETs, HEMTs and HBTs based on either GaAs-based or InP-based heterostructures have also made remarkable progresses by miniaturization into the nanometre range, providing higher speeds with lower power consumption than Si devices at the same miniaturization level. Currently, the InP-based HEMT device with the feature sizes in the submicrometre range is the fastest device of all the semiconductor devices, approaching THz operation frequencies.

Such progress of the ‘nanofabrication’ technology will soon make, or partially has already made in test devices, the dimensions of scaled-down transistors comparable with the Fermi wavelength of electrons in semiconductors. This may raise some concerns on the performance of traditional devices based on semiclassical principle, but it also opens up exciting possibilities of constructing novel quantum devices directly based on quantum mechanics [2]. In quantum devices, quantum mechanical wave-particle motions of individual electrons are controlled by artificial quantum structures such as quantum wells, quantum wires, quantum dots and single and multiple tunnelling barriers so as to realize devices with new functions and higher performances.

The purpose of this paper is to discuss the present status of research on the quantum devices based on III–V compound semiconductors, introducing and reviewing the recent results obtained by the author’s group at the Research Center for Interface Quantum Electronics (RCIQE). Most of the results have been obtained under a multiuniversity national project dedicated to single electron devices (the ‘SED’ Project) which is currently ongoing (head investigator: H Hasegawa, RCIQE, period: April 1996–March 2000) in Japan involving about 50 professors from all over Japan. In this project, an emphasis is placed on the ‘quantum mechanical aspects’ of the single electron devices and intensive collaborative efforts are being made to find ways for next generation electronics through the vast open fields of the quantum regime where no ‘roadmaps’ exist.

After giving a brief overview on quantum devices in section 17.2, the paper discusses specific details of III–V quantum structure formation by molecular beam epitaxy (MBE) in section 17.3, formation and control of III–V nanostructure surfaces and interfaces in section 17.4, novel Schottky in-plane gate quantum wave devices in section 17.5 and, finally, novel Schottky in-plane gate and wrap-gate single electron devices in section 17.6 with a brief conclusion in section 17.7, respectively. A particular emphasis is paid on single electronics because of its promising prospects.

17.2 Brief general overview on compound semiconductor quantum devices

17.2.1 Wave-particle duality and devices

When the device feature sizes go into the decanometre region or below, new physical effects appear. Among various effects, those which may be utilized for device applications include the following:

- (1) formation of new quantum states in single and coupled quantum structures such as quantum wells, quantum wires and quantum dots and the formation of artificial mini-bands in superlattices;
- (2) tunnelling and resonant tunnelling in single and multiple barriers;
- (3) propagation of phase-coherent guided electron waves along one-dimensional quantum wires, and their wave interactions; and
- (4) Coulomb blockade effect in small tunnel junctions.

All these effects except (4) originate from quantum mechanics.

Since electrons manifest predominantly as either wave-nature or particle-nature depending on their environments, one can conceptually envisage two kinds of devices in the quantum regime, i.e. ‘quantum wave devices’ and ‘single electron devices’. In the ‘quantum wave devices’, electrons are put into various phase-coherent structures and their wave properties are utilized to realize various new functions [2]. In ‘single electron devices (SEDs)’ [3], electrons are put into ‘dot’ structures and their particle-nature is emphasized so that the number of electrons in the dot can carry the information.

17.2.2 Quantum wave devices

17.2.2.1 Devices proposed

The quantum wave devices utilizing the above effects (1)–(3), which have been so far proposed and partially fabricated, include; (1) quantum wire transistors, (2) electron lenses, reflectors and deflectors, (3) electron Bragg reflectors and superlattice devices, (4) resonant tunnelling diodes and transistors and (5) quantum inter-interference devices and Y-switch devices using waveguide structures. Most of these devices have their optical and microwave device analogues, since they rely on the general properties of waves.

17.2.2.2 Materials

III–V compound semiconductor materials are featured by precisely controlled crystal growth capabilities down to monolayer level by advanced MBE and metal organic vapour phase epitaxy (MOVPE) techniques, wide selection varieties of materials with different properties, availability of high quality hetero-interfaces with designable band alignments and negligible interface defect states, superb

electron transport with low probabilities of scattering, and availability of self-organization and selective growth modes. Owing to these features, most of the research on quantum wave devices has been done so far on III–V compound semiconductor materials such as GaAs/AlGaAs and InAlAs/InGaAs/InP systems.

At RCIQE, extensive efforts are being made to produce artificial quantum structures such as quantum wires and dots by MBE and MOVPE. Recent results based on MBE are described in section 17.3 of this paper, and recent efforts based on MOVPE is described by a separate paper by Motohisa and Fukui [4] in these proceedings.

17.2.2.3 Device structures and key issues

From the structural point of view, the device may be classified into two types, i.e. the vertical multithin film type structure where electrons are transported perpendicular to the surface of the epitaxial multilayer structures, and the lateral waveguide type structure where electrons are transported along the waveguide structure defined by some means such as electron beam (EB)-lithography and etching or some selective growth during multilayer growth. Of the former type devices, the resonant tunnelling based vertical devices can operate at room temperature and appear to be most promising. Thus, substantial work is going on particularly in the United States and Japan to investigate their feasibility as basic devices for next-generation electronics.

For high-density planar integration, lateral waveguide type devices are obviously much more suitable. As electron wave-versions of microwave and millimetre-wave integrated circuits, complex analogue types of new signal processing functions may be realized from phase-coherent interactions in guided electron waves propagating on planar substrates. Unfortunately, not enough work has been performed up to now as regards the feasibility of these devices operating at acceptably high temperatures.

Previous GaAs 2DEG-based gated quantum wires and wire-based coupler devices mostly utilized the so-called split gate structure. This structure unfortunately provides only weak electron confinement, and thus the devices could be operated only at low temperatures in millikelvin to a few kelvin range. In order to achieve higher temperature operation, one needs a gate structure which provides stronger confinement of electrons so that the subband spacing in the electron wave mode becomes much larger than kT .

Another important consideration is that the phase coherent length, ℓ_c , along the wire which should be longer than the channel length. This length is again expected to be strongly temperature dependent. According to a latest measurement done using InGaAs/InP triple barrier resonant tunnelling diode [5], ℓ_c is 90 nm at 4.2 K; and 55 nm at 77 K. Such small values impose severe limitations on the feasibility of devices based on electron–wave interaction. Another important issue of the waveguide type device is the fact of the Fermi–Dirac distribution force trade-off between coherence and available current level,

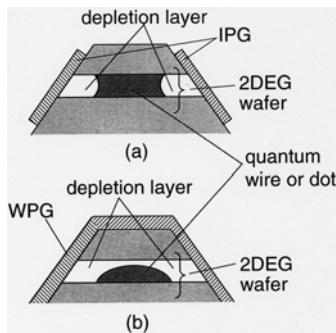


Figure 17.1. Schematic illustration of (a) Schottky in-plane gate (IPG) and (b) Schottky wrap gate (WPG) structures.

resulting in poor drivability.

The issues mentioned above seem to indicate that these devices are less attractive for application as Boolean switches where room-temperature operation with a large current drivability is an essential requirement. However, this may not be the case when the device is used in low-power, high-performance, complex signal processing analogue applications including quantum computing.

17.2.2.4 Approaches at RCIQE

At RCIQE, we are interested in analogue device applications such as quantum computing applications of lateral waveguide type devices. To provide stronger electron confinement than that in previous split gate devices, we have proposed and tested two kinds of new Schottky gate structures. They are the Schottky in-plane gate (IPG) and the Schottky wrap gate (WPG) structures, schematically shown in figure 17.1. Using these gate structures, we have fabricated quantum wire (QWR) transistors [6–9], gated Aharonov–Bohm (A–B) ring devices [10] and IPG-based GaAs wave coupler devices [11].

17.2.3 Single electron devices

17.2.3.1 Devices proposed and their prospects

In ‘single electron devices (SEDs)’, electrons are put into ‘dot’ structures and their particle-nature is emphasized so that the number of electrons in the dot can carry the information. Single electron devices so far proposed include single electron transistors (SETs), single electron memories [12], dot-based cellular automata [13], binary decision devices [14] etc.

As is well known, a SET consists of a dot, two tunnel junctions and a control gate. Its dot-based structural simplicity, small size nature and low power capability of single electron devices are extremely attractive and make

them candidates for next-generation ultrahigh-density main-stream information processing devices co-existent with or beyond ultimately scaled-down Si CMOS devices.

Any serious candidate for the next generation electronics should be potentially capable of surmounting the limitation of the existing technology. The present LSI is mainly based on the Boolean logic architecture which is implemented by an array of transistors which operate only as simple, small and robust switches of lumped nature. Limiting factors anticipated for the hitherto so successful technology include; (1) material and processing related limitation (doping fluctuation, avalanche breakdown, MOS interface instability, electromigration, stress-migration, interface reaction), (2) power limitation, (3) wiring limitation, (4) quantum mechanical limitation (quantum fluctuation, failure of device and device isolation due to tunnelling), and (5) system architecture limitation.

SEDs that allow manipulation of individual electrons, can be regarded as ultimate forms of the electron device. Their potential integration level is obviously extremely high due to its small size, being 10^{10} – 10^{11} Trs cm $^{-2}$. Extremely low operation power solves some of the instability and reliability problems. The speed power product of SETs is predicted to lie close to the quantum limit set by Heisenberg's uncertainty principle [15]. Thus, SEDs are among the fastest switching devices that are allowed to exist in the real world.

From these simple consideration, SEDs seem to be potentially capable of overcoming some of the important limitations, if they can be operated at room temperature, and if the technology for their high-density integration is made available.

17.2.3.2 Classical or quantum?

Early SETs utilizing metal dots were based on the classical Coulomb blockade of tunnelling. The first semiconductor SET was a GaAs-based one, utilizing large dots formed by split gate configuration [16]. Its operation was based on the purely classical Coulomb blockade and operated only in the millikelvin range.

Such SET devices may be viewed as classical devices [17]. However, an important difference emerges between metal SETs and semiconductor SETs, as one reduces the dot size for higher temperature operation. This is because of the difference in Fermi wavelength. Namely, the Fermi wavelength of metal is typically 0.1 nm whereas that of a semiconductor can be as large as 50 nm. Thus, within the foreseeable future with the technological miniaturization capability of down to 1 nm, quantum effects can be ignored in metal SETs, whereas it becomes vitally important in semiconductor SETs. It has recently been shown using a side-gated vertical III–V resonant tunnelling diode structure [18] that the resultant nanometre-scale quantum dot behaves like an artificial atom having shell structures whose electron spin filling obeys Hund's rule. Strongly temperature-dependent peak heights of conductance spectra can also be explained only with

a quantum mechanical consideration [19]. A simple calculation based on an effective mass approximation tells us that the quantum confinement energy will become larger than the static charging energy, when the dot size becomes a few tens of nanometres. Addition of the quantum confinement energy on top of the charging energy makes obviously higher temperature operation easier.

Interdot coupling in various dot arrays also provides the chance of forming artificial molecules. Such arrays will have voltage tunable coupling and voltage tunable internal electron occupancy or internal states. This seems to provide an extremely broad freedom for realization of new functional devices.

Thus, various quantum effects expected in semiconductor SEDs seem to provide the chance for room-temperature operation as well as unexplored new possibilities for new functional devices and new architecture LSIs.

Therefore, an important key issue for semiconductor SEDs is the development of a suitable theory for device analysis and design based on quantum mechanics. Unsolved issues include level structures and wavefunctions of quantum states in dots, many body effects, nature and rates of single electron tunnelling, phase coherence, tunnelling time, co-tunnelling, current leakage mechanisms, effects of external electromagnetic environments, charge fluctuation and off-set charge, current fluctuation and noise etc. For semiconductor coupled dots, one also needs a theory for quantum mechanical analysis and design of such artificial molecules and their arrays.

17.2.3.3 Materials for SEDs

What is the most suitable material for SED LSIs? A brief comparison of various material technologies is given in [table 17.1](#). Among various material technologies, the metal SED technology is the oldest one. Room-temperature operation of metal-based SETs was demonstrated recently [20].

As for the comparison between Si and compound semiconductor SEDs, one naturally pays attention to a recent rapid progress toward room-temperature operation that has been demonstrated in Si SED technology. Conductance oscillation at room temperature has been reported in ultranarrow Si wire based SETs [21, 22]. However, these devices are in the stage of discrete devices. For high-density integration of SED devices, a technology for high-density uniform dot array formation is required, and in this respect, the III-V material technology is most advanced, as the examples given in section 17.3 will demonstrate.

Recently, as a more short-term application, room-temperature single electron memory devices using Si nanocrystal dots [23] or ploy-Si granular structure [24] are being intensively developed, realizing a 128 Mb memory [25]. Since the operation principle of these devices is simple and very similar to the flash memory devices, III-V versions utilizing advanced heterostructures are obviously possible as has been partially demonstrated [26].

Thus, it is still possible that III-V SED devices will become the main stream devices, if they can be operated at room temperature. Even, if Si SED

Table 17.1. Comparison of various material technologies for SET formation.

	Metal-dot	Si-dot	III-V-dot
RT operation demonstration	◎ STM oxidation	○ EB ₊ oxidation	×
Dot array with position and size control	○	×	◎
Availability of ultrathin tunnelling barrier	○	◎	III-V hetero ◎ Insulator ×
V_{th} control	×	○	○

technology becomes eventually the main stream devices, there will be ample room for III-V SEDs for high-speed/low-power memory and logic applications in mobile communication and computation co-existent with conventional III-V devices such as MESFETs, HEMTs, HBTs, LDs and LEDs. Signal processing in the terahertz region as well as quantum computation is also an interesting application field for III-V SEDs. Highly sensitive electrometers, single photon generators and detectors and measurement standards are other application areas.

17.2.3.4 System architecture

As for the system architecture of SED LSIs, one can think of the following three alternatives; (1) conventional Boolean logic based on von Neumann type architecture implemented with conventional SET switch-array device architecture with improved switching performance, (2) conventional Boolean logic system architecture with a new device architecture utilizing SEDs with increased functionality than switches, and (3) non-von Neumann architecture with new device architecture utilizing SEDs with increased functionality.

One serious difficulty expected in the single electronics is the stochastic nature of quantum processes, and the selection of the device/system architecture should be made with a serious consideration of this problem.

Quantum dot systems seem to offer unique advantages for the realization of second and third alternatives. As an example, in one-dimensional and two-dimensional arrays of dots with tunnelling junctions, electrons propagate like solitons [3], and this may be used as wiring or to implement certain types of special information processing. It has been shown theoretically that cellular automata can be constructed by the chain of sets of five single electron dots [13].

Recently, binary decision device arrays [14] and Boltzmann neuron machines [27] based on more conventional tunnel junctions have been proposed.

Approaches at RCIQE

At RCIQE, we are very much interested in developing device technologies for GaAs-based and InP-based SEDs. New Schottky in-plane (IPG) and wrap gate (WPG) structures already shown in figure 17.1 have also been applied to GaAs/AlGaAs etched wires and selective MBE grown InGaAs wires, and have resulted in higher temperature operation at 30–50 K of III–V based SETs. Conductance oscillation and Coulomb blockade characteristics were observed for a Schottky IPG SET formed on a GaAs-based etched 2DEG bar [28] and an InP-based WPG SET formed on a selective MBE grown InGaAs quantum wire [29], respectively. The appearance of grouped peaks in the latter strongly indicated the existence of shell filling in these lateral devices. The conductance oscillation characteristics observed in a three-dot WPG SET formed on a GaAs-based etched 2DEG bar suggested artificial molecule formation due to dot coupling in this device [30].

17.3 Formation of GaAs- and InP-based quantum structure by molecular beam epitaxy

17.3.1 Approaches for formation of quantum structures

For fabrication of integrated circuits based on quantum devices, technologies for forming high-density arrays of quantum structures, such as quantum wires and dots, are required. Representative approaches so far tried include the following:

- (1) the direct fabrication of nanostructures by applying standard Si ULSI processing technologies, such as EB lithography, dry etching, oxidation, to silicon or silicon-on-insulator (SOI) wafers;
- (2) the formation of nanostructures on III–V multilayer epitaxial wafers by EB lithography and etching;
- (3) the selective depletion of two-dimensional electron gas (2DEG) in III–V wafers by Schottky split gate, IPG and WPG geometry etc;
- (4) the direct fabrication of nanostructures by scanned probe-induced atom-manipulation and surface reaction;
- (5) the realization of wire or dot structures using the self-organizing growth mechanism in MBE or MOVPE growth of III–V materials; and
- (6) the formation of ultrafine particles by CVD processes and various molecular reactions.

Among these, approaches (1) and (2) are most practical if they work sufficiently well. However, achievable sizes, size uniformity, size controllability and interface smoothness are not sufficient yet. Process-induced damages

are serious in some cases. In approach (3), dots are formed far from the surface and deep in the high quality crystal in a gate controlled fashion, and the isotropic feature of Coulomb force quickly smears the ruggedness of the surface. Weak confinement potential in the original split gate structure can be significantly increased by IPG and WPG geometries, as already mentioned. Approach (4) is very good for research on discrete level devices [20], but utterly impractical at present for high-density integration. Examples of approach (5) include selective growth on mesa-patterned substrates, area-selective growth using patterned insulator windows, self-organized growth utilizing step bunching on vicinal substrates, Stranski–Krastanow (SK) mode driven self-assembly of dots, strain-driven self-organized disk formation on high-index substrates, etc. Using approach (6), small and uniform silicon nanoparticles have been realized with observation of Coulomb staircases at room temperature [31].

Among these various approaches applied to various materials, the most advanced technology of uniform nanostructure formation with control of position, shape and size is the selective MBE and MOVPE growth technique as applied to GaAs-based and InP-based materials.

At RCIQE, we are taking the approaches of (2) and (5) for formation of GaAs-based and InP-based quantum structures. Namely, with approach (2), etched wires are made by applying EB lithography and wet chemical etching to MBE grown heteroepitaxial wafers. Such wires are then subject to device processing steps to produce IPG and WPG controlled quantum wire transistors and single electron devices. The other approach (6) is carried out by the selective MBE growth on patterned substrates. For patterning of substrates, we use again approach (2). We believe that this approach is one of the most promising methods for the formation of high-density nanostructure arrays with position- and size-control capabilities beyond present lithography limits. This type of technology is essentially important for ultralarge scale integration of quantum devices such as quantum wire transistors and single-electron transistors. However, only a few attempts have been made so far for actually constructing quantum devices on such selectively grown III–V quantum nanostructure arrays. Recent results on selective MBE growth of InGaAs nanostructures are presented in the next section.

17.3.2 Selective MBE growth of InGaAs quantum wires, dots and wire-dot coupled structures

As mentioned above, formation of high-density arrays of uniform and high-quality quantum wires and dots is one of the key issues for high-density integration of quantum devices. From such a viewpoint, we have recently been studying selective MBE growth of InP-based InGaAs/InAlAs on various patterned InP substrates. As compared with the GaAs/AlGaAs system, this material system is attractive for realizing high-temperature operating quantum devices due to its large conduction band discontinuity and superb electron transport.

We have so far succeeded in forming high-quality quantum wires [32–34],

quantum dots [35, 36] and quantum wire-dot coupled structures [37], using $\langle \bar{1}10 \rangle$ -oriented mesa-stripe and pedestal patterns. Using this direction, wires and dots can be formed in a reproducible and controllable fashion. A minimum wire width of about 30 nm has been achieved. More recently, such $\langle \bar{1}10 \rangle$ -oriented InGaAs quantum wires were successfully applied to the fabrication of quantum wire transistors [38, 39] and single electron transistors [29]. On the other hand, a disadvantage of $\langle 110 \rangle$ - or $\langle \bar{1}10 \rangle$ -oriented wire growth is that the wire can be grown only in one direction. Although almost all the selectively grown quantum wires reported so far by us and other groups [40, 41] are such wires, this lack of flexibility becomes a serious disadvantage in constructing high-density quantum circuits. This problem can be overcome by growing quantum wires along the $\langle 100 \rangle$ -direction, where wires can be formed in two orthogonal and equivalent $\langle 100 \rangle$ -directions [42]. In order to test the feasibility of such growth, we have performed some basic growth experiments of InGaAs wires on $\langle 100 \rangle$ -oriented mesa-stripes [35, 43].

17.3.2.1 Principle of formation of wires and dots

The principle of growth of InGaAs quantum wire on InAlAs ridge structure is schematically shown in [figure 17.2](#). Here, the wires are formed by utilizing the difference in the shape of the ridge top between the InGaAs and InAlAs ridges. If one can find a growth condition forming a sharp InGaAs ridge and a round InAlAs ridge as shown in figures 17.2(a) and (b), respectively, an InGaAs ridge wire as shown in figure 17.2(c) can be formed by combining these growths. Similarly, an InGaAs quantum dot can be formed on an InAlAs pyramidal-shaped structure, if one can find growth condition forming an InGaAs pyramid having a sharp top corner and an InAlAs pyramid having a rounded top corner.

The actual shape of the ridge and pyramid top may be given by well-defined crystalline facets or simple smooth curve described by a radius of curvature, depending on the crystalline orientation and growth conditions. Actual examples are given below.

17.3.2.2 Preparation of patterned substrates

Three different kinds of patterned (001) InP substrates as shown in [figure 17.3](#) were prepared. Arrays of mesa-stripes shown in figures 17.3(a) and (b) were used for the formation of InGaAs ridge quantum wires along the $\langle \bar{1}10 \rangle$ -direction and $\langle 100 \rangle$ -direction, respectively. On the other hand, rather complicated pattern containing an array of square mesa pedestal connected each other with $\langle \bar{1}10 \rangle$ -oriented mesa-stripes was used for formation of quantum wire-dot coupled structures. The $\langle \bar{1}10 \rangle$ -oriented mesa-stripes shown in figure 17.3(a) were prepared by a standard photolithography and wet chemical etching, whereas the patterns for formation of the $\langle 100 \rangle$ -oriented wire and the quantum wire-dot coupled structures were formed using EB-lithography and wet chemical etching.

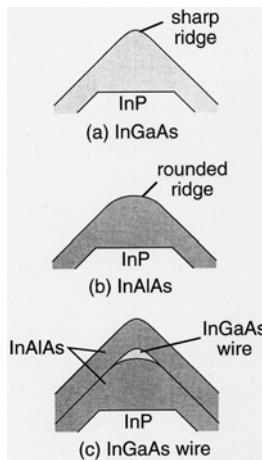


Figure 17.2. Principle of formation of a ridge quantum wire.

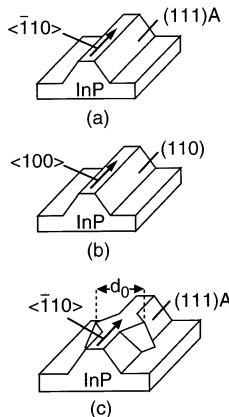


Figure 17.3. Patterned InP substrates used for selective MBE.

17.3.2.3 MBE growth

A standard solid-source MBE system was used in our study. In all the growth runs, the samples were carefully cleaned in organic solvents, then etched slightly by $1\text{H}_2\text{SO}_4 : 5\text{H}_2\text{O}_2 : 4\text{H}_2\text{O}$ for 30 s and dipped in HF for 1 min. Then, substrates were immediately introduced into the MBE chamber. For growth monitoring, reflection high-energy electron diffraction (RHEED) pattern observations were made on unpatterned planar (001) InP substrates put beside the patterned substrates. The growth rates for InGaAs and InAlAs on the planar substrates were both set at 600 nm h^{-1} , and a growth temperature of $T_g = 500\text{--}580^\circ\text{C}$ were

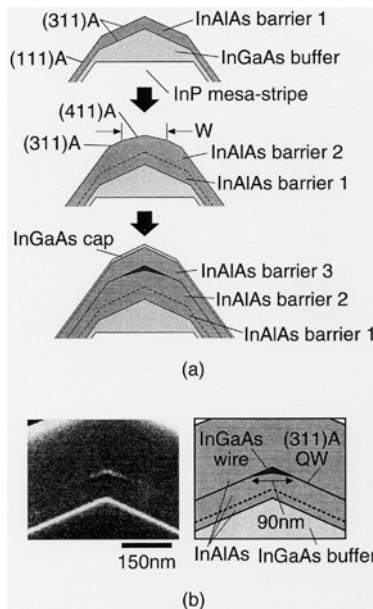


Figure 17.4. (a) Preparation sequence and (b) cross-sectional SEM image of the $\langle\bar{1}10\rangle$ -oriented InGaAs ridge quantum wire.

used. The As₄ back pressure was chosen so as to maintain As stabilized surfaces during the growth.

17.3.2.4 Growth and characterization of $\langle\bar{1}10\rangle$ -oriented InGaAs quantum wires

The formation sequence of the $\langle\bar{1}10\rangle$ -oriented InGaAs ridge quantum wire is shown in figure 17.4(a). On the substrate shown in figure 17.3(a) an InGaAs buffer layer, an InAlAs barrier layer 1 and an InAlAs barrier layer 2, a thin InGaAs wire layer, an InAlAs barrier layer 3 and a thin InGaAs cap layer were grown successively. The principle of ridge wire formation is as follows [32]. First, the growth of the InGaAs buffer layer at 550 °C and InAlAs barrier layer 1 at 500 °C forms a (311)A ridge on each mesa-stripe, as shown in figure 17.4(a). This (311)A ridge becomes unstable if the growth of InAlAs barrier layer 2 is carried out at a higher growth temperature of 580 °C and gradually changes into a (411) ridge with width W . Then, subsequent growth of InGaAs at 500 °C leads to the selective formation of InGaAs wires with width W surrounded by (311)A facets and (411)A facets. Finally, the wires are covered by the thick InAlAs barrier layer 3 and the InGaAs cap layer. Figure 17.4(b) shows an example of the cross-sectional SEM image of the InGaAs ridge quantum wire. An arrow-head shaped InGaAs wire with a width of 90 nm was clearly seen.

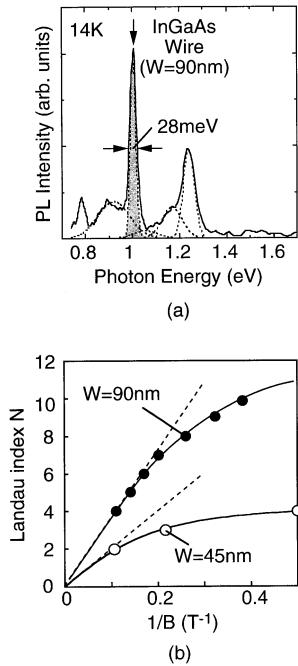


Figure 17.5. Results of (a) PL and (b) magnetoresistance measurements for the $\langle\bar{1}10\rangle$ -oriented InGaAs ridge quantum wires.

The width of the wire, W , can be controlled by changing the growth time of the InAlAs layer 2, since the width of the (411)A facet region is proportional to the growth time of the InAlAs layer 2. At present, a minimum wire width of about 30 nm has been achieved. More specifically, the growth of the bottom InAlAs layer 2 at 580 °C for $t = 3, 5$ and 10 min resulted in the formation of the wire with widths, W , of 35, 46 and 90 nm, respectively. Moreover, the wire width W and the total thickness of the bottom barrier layer can be adjusted independently as one likes by adjusting the thickness of the first InAlAs layer grown at 500 °C properly.

Figure 17.5(a) shows a PL spectrum of the wire with the width W of 90 nm. A dominant narrow PL peak at 1.0 eV was assigned as an emission from the InGaAs ridge quantum wires themselves by a spatial and energy resolved cathodoluminescence (CL) observation. Additionally, this energy position agrees well with the theoretical value obtained by numerically solving the Schrödinger equation using the observed wire shape and size and assuming lattice-matching of the InGaAs wires to InP. On the other hand, the side peaks at higher and lower energy sides come from the parasitic quantum wells formed on the (311)A sidewalls and at the bottom of the mesa, respectively. The observed narrow PL

peak width of 28 meV indicates that the present $\langle\bar{1}10\rangle$ -oriented InGaAs ridge quantum wire has the best uniformity of all the InP-based selectively grown InGaAs quantum wires and has comparable or better uniformity than most of the GaAs-based selectively grown quantum wires [34].

Magnetroresistance measurements were made for the InGaAs ridge quantum wire having Si-doped top InAlAs barrier layer 3. Clear Shubnikov-de Haas (SdH) oscillations were observed, indicating that the crystal and the interface of the present wires are free from serious scattering centres. Landau plots obtained for the InGaAs ridge quantum wires having widths of 90 and 45 nm are shown in figure 17.5(b). Nonlinearity observed in the Landau plots indicates these quantum wires act as high quality one-dimensional electron wave guides. The value of the subband spacing was estimated to be 10 and 13 meV for these wires by an analysis based on the harmonic potential approximation. These values are much larger than those realized in the conventional 2DEG-based split-gate type wires (<3 meV). Moreover, these values are the largest of all the subband spacings measured by electrical methods for selectively grown quantum wires [44, 45].

17.3.2.5 Growth and characterization of quantum wire-dot coupled structures

The InGaAs quantum wire-dot coupled structure was formed by applying the similar selective growth procedure for formation of the $\langle\bar{1}10\rangle$ -oriented InGaAs quantum wires to the patterned substrate as shown in figure 17.3(c). Growth included that of a bottom InGaAs buffer layer and that of an InAlAs/InGaAs/InAlAs structure. Growth of a bottom InGaAs buffer layer and an InAlAs layer on the pattern resulted in formation of pyramids having (521) facets with a rounded top corner on the square mesa region and (411)A/(311)A facet ridges on the stripes, respectively. Then, subsequent growth of a thin InGaAs layer led to selective growth of InGaAs quantum dot and quantum wires on the rounded top of the pyramid and on the (411)A facet region of the ridges, respectively, as shown in figure 17.6(a). Additional narrow quantum wires were also formed on the rounded $\langle\bar{1}10\rangle$ -oriented ridges of the pyramid. Since these additional wires provide potential barriers due to quantum confinement caused by reduction of width and thickness, a coupled wire-dot structure having a potential barrier in between is realized as shown in figure 17.6(a). Finally, the overall structure was covered by a thick InAlAs barrier layer.

A plan-view SEM and panchromatic CL images of high-density InGaAs coupled wire-dot array are shown in figure 17.6(b). As shown in the SEM image, the (521) pyramids were connected smoothly with the (311)A ridges. Formation of the InGaAs quantum dot at the top of the (521) pyramid as well as that of the InGaAs ridge quantum wires on the (311)A ridges are confirmed by the CL image shown in figure 17.6(b). Figure 17.6(c) shows results of spatial and energy resolved CL observations for the area containing a single dot. The CL image taken at 1.0 eV clearly indicated successful formation of quantum wires and dot at the top of the (311)A ridges and that of the (521) pyramid, respectively.

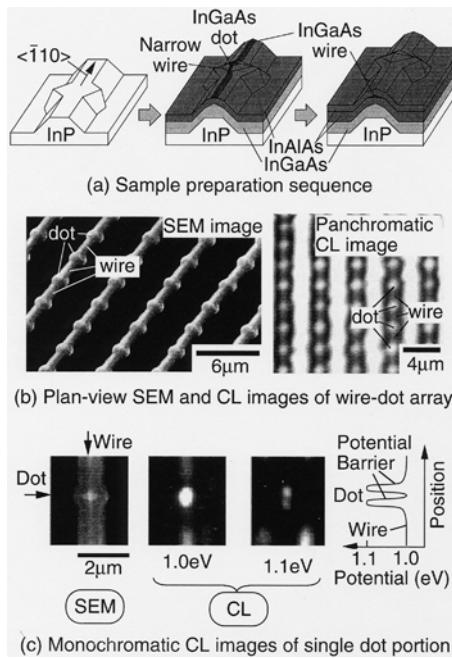


Figure 17.6. Direct formation of InGaAs quantum wire-dot coupled structure arrays by selective MBE growth.

On the other hand, double spots seen at the centre of the CL image taken at 1.1 eV revealed the presence of potential barriers between the dot and the wires as schematically drawn at the right-hand side of figure 17.6(c). By comparing the SEM and CL images, positions of the double spots were found to almost coincide with those of the $\langle\bar{1}10\rangle$ -oriented ridge of the pyramid, confirming the realization of the potential barrier by the narrow quantum wires along the $\langle\bar{1}10\rangle$ -oriented ridge of the pyramid. Since the main part of such a potential modulation is most likely to come from the difference in the quantum size effect for electrons in the conduction band, the height of the potential barriers for electron is roughly estimated at 0.1 eV from the CL data.

As for the size controllability of the dot and the potential barrier, they can be controlled by the initial pattern geometry and the growth conditions. At present, minimum dot width and lateral barrier width of about 100 nm and 80 nm, respectively, were achieved in a controlled fashion, indicating the applicability of the present approach to the formation of actual SETs.

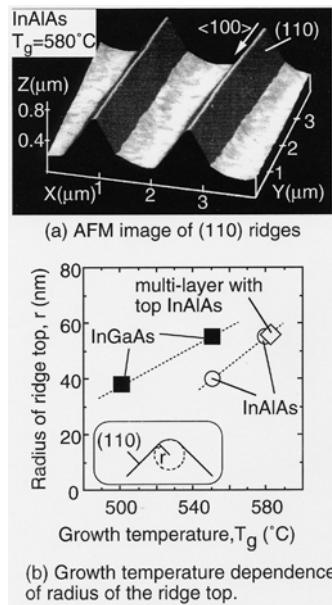


Figure 17.7. Results of AFM observation for the $\langle 100 \rangle$ -oriented ridges.

17.3.2.6 Growth of $\langle 100 \rangle$ -oriented InGaAs quantum wires

In order to find appropriate growth conditions for formation of the InGaAs quantum wires on the $\langle 100 \rangle$ -oriented mesa-stripes, single InGaAs layers, single InAlAs layers as well as InGaAs/InAlAs multilayers were grown on the patterned substrates shown in figure 17.3(b).

The AFM study showed that all the growth produced ridge structures having two (110) sidewalls on the mesa-terraces. An example of the AFM image is shown in figure 17.7(a) for the single InAlAs layer grown at 580 °C. All the ridges possessed round top shapes as shown in the inset of figure 17.7(b) which can be well represented in terms of the radius of curvature, r . The observed radius r for each sample is summarized in figure 17.7(b). The radius r was found to increase with the growth temperature for different types of the ridge. In addition, the InGaAs ridge possessed larger radius r than that of the InAlAs ridge grown at the same temperature. The multilayered sample having an InAlAs top layer was found to possess the same value of r as that of the single InAlAs layer. Moreover, in both the single layered and multilayered samples, the value of r was found to saturate to a characteristic value for a given growth temperature, and an additional growth did not change the value of r any more. The difference in the initial terrace width of InP mesa-stripe also had no effect on the radius r .

All of the above phenomena indicated that the radius r is determined by a growth kinetics-related dynamic balance between the desorption and

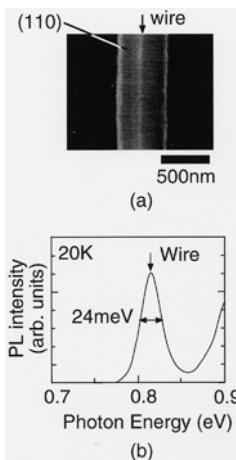


Figure 17.8. (a) Plan-view SEM image and (b) PL spectra of the (100)-oriented InGaAs ridge quantum wire.

incorporation rates of group-III atoms at the ridge top as in the case of the self-limited selective MOVPE growth of GaAs quantum dot [46]. Namely, by increasing the growth temperature, desorption rate of group-III atoms from the ridge top increases much more drastically than that from the (110) sidewalls probably due to weak bonding energy at the ridge top, leading to rounding of the ridge shape. On the other hand, stronger bonding of the Al atoms to the surface than that of Ga atoms reduces the desorption at the top of the ridge, resulting in the formation of the observed sharper ridge shape on InAlAs than on InGaAs.

Based on the data shown in figure 17.7, the (100)-oriented InGaAs quantum wire was formed by growing a thin InGaAs wire layers (2.5 nm) at 500 °C on the multilayered ridge having a top InAlAs layer grown at 580 °C. The wires were finally covered by thick InAlAs barrier layers. From the observed difference in the radius between the InGaAs ridge and the InAlAs ridge, formation of crescent-shaped wires having a width of 78 nm and a lateral thickness at the centre of 7 nm were expected.

Figure 17.8(a) shows a plan-view SEM image of the (100)-oriented InGaAs wire. As seen here, the uniformity along the wire is fairly good. The PL spectra of the wires taken at 20 K is shown in figure 17.8(b). An intense and narrower PL peak was observed at 0.82 eV. From the spatially and energy resolved CL observations, this peak was found to come from the InGaAs ridge wire itself. However, the peak energy value was found to be smaller than that expected for the wire having the sizes mentioned above. A numerical analysis solving Schrödinger's equation for the expected wire shape and size with the assumption of perfect lattice match gave a peak position of 0.91 eV. This large deviation of the peak positions can be caused either by the difference in the size or by that in the

composition or both. Namely, one possible explanation is the formation of wires with the vertical thickness larger than the expected value of 7 nm due to growth of the InGaAs layers with a certain vertical thickness on the (110) sidewalls. Another possibility is the increase of InAs alloy composition of the InGaAs wire portion as reported by other groups for $\langle\bar{1}10\rangle$ -oriented InGaAs/InAlAs ridge quantum wires [47]. However, a further investigation is necessary to clarify this point.

17.4 Control of surfaces and interfaces of nanostructures

17.4.1 A key issue for nanofabrication

One major nanofabrication issue lies in the control of surfaces and interfaces of quantum structures. This is because, homo-, hetero-, MOS and Schottky interfaces, which are essential constituent elements in the present-day semiclassical devices, and control classical motions of carriers, should now control the wave-particle properties of individual electrons in quantum devices. Due to reduced dimensions and to different roles of the interfaces, interfaces play far more important and delicate roles in the quantum devices. Thus, the interfaces of quantum structures should be perfect in the interface atom arrangements and be capable of producing desired potential profiles required for quantum device operation.

Additionally, the interface region should be free of ionized impurities and trapping defects such as surface states, interface states and discrete deep levels. This is particularly important because any stray charge near the dot produces additional polarization charge or offset charge in the dot, and shifts the threshold voltage for SET in a random fashion. This famous and serious problem of ‘offset charge’ appears to be much easier to cope with in Si and III–V SETs than in metal dots embedded in amorphous insulators owing to ample experience of strict charge control near the Si–SiO₂ interface or epitaxial heterointerfaces in practical FET devices. However, it is still a very difficult technological problem which one has to solve for success.

Particularly, for III–V materials, the control of surface states by a suitable passivation technique is vitally important, even if dots and tunnelling barriers are made by potential engineering within III–V materials. This is because SEDs will be placed near the surface for planar integration.

In this connection, the author’s group at RCIQE has recently succeeded in removing Fermi level pinning and surface state effects from AlGaAs/GaAs near surface quantum wells [48,49] and InAlAs/InGaAs/InP quantum wires [50], using a silicon interface control layer (Si-ICL) based surface passivation technique [51], which is shown schematically in [figure 17.9\(a\)](#). In this structure, an ultrathin Si layer grown by MBE is inserted between the compound semiconductor and the outer thick Si-based dielectric layer. The basic idea is to achieve coherent termination of surface bonds of the compound semiconductor by the Si-ICL and subsequently to passivate Si-ICL by Si-based dielectric. [Figure 17.9\(b\)](#) shows the

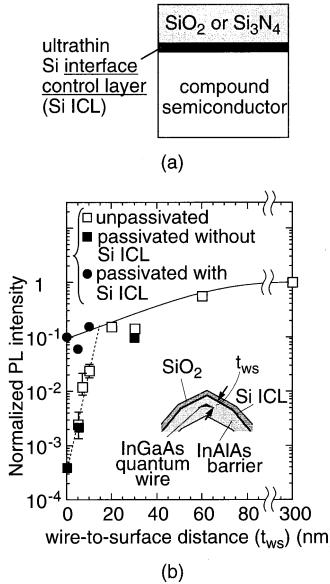


Figure 17.9. (a) Si ICL-based passivation structure. (b) Effect of Si ICL-based passivation on PL properties of the near surface InGaAs ridge quantum wires.

measured PL data for the $\langle\bar{1}10\rangle$ -oriented InGaAs ridge quantum wire with and without surface passivation by the Si-ICL technique, demonstrating the powerful passivation effect of the present method. Such a technology may eventually lead to the introduction of pinning-free ultrathin tunnelling insulator films to III-V SEDs.

17.4.2 Formation technology of damage-free Schottky gates for quantum structures

17.4.2.1 Formation of metal-semiconductor interfaces for quantum devices

Control electrodes are necessary in any quantum devices as in present-day devices. Requirements for the formation technology of metal electrodes in quantum devices are very severe. It should be highly controllable in adjusting the barrier height to match the energy requirements in quantum devices. It should not produce damage to the semiconductor, since quantum devices are extremely sensitive to defects.

In this section, Schottky contact formation by an *in situ* electrochemical process is described. This work is motivated by the following features of the process. Electrochemical deposition of metals is a low-temperature process and an extremely low-energy process. Heat of condensation and bombardment by electrons or ions, that are encountered in the conventional metal deposition

processes in vacuum, do not exist. Furthermore, it has also been demonstrated recently that an oxide-free intimate contact can be formed by a pre-deposition etching of semiconductor surface carried out *in situ* in the same electrolytes [53].

As for the Schottky barrier heights (SBH), they are known, particularly on III-V materials, to be very weakly dependent on the metal workfunction due to the strong Fermi level pinning. According to the metal induced gap state (MIGS) model [54, 55], which is currently the most prevailing model for the Fermi-level pinning, strong pinning is an intrinsic property of an intimate metal semiconductor interface, because it is due to a high-density MIGS continuum produced by the penetration of the evanescent wave portion of the metal wavefunction into semiconductor. However, we have found that SBHs become much more strongly dependent on the metal workfunction by using the *in situ* electrochemical process.

17.4.2.2 *In situ* electrochemical process for metal deposition

The *in situ* electrochemical process was done using the set-up schematically shown in figure 17.10(a). The electrolyte bath contains three electrodes, i.e. a semiconductor electrode onto which metal is deposited, a Pt counterelectrode and a reference saturated calomel electrode (SCE). The overpotential of the semiconductor electrode with respect to the SCE reference was controlled by a potentiostat with a pulser.

In this study, metal films of Pt, Ni, Co, Ag and Sn were electrodeposited on n-type (001) GaAs epitaxial layers grown by MBE on n⁺-GaAs substrates and n-type (001) InP substrates. GeAu/Ni alloyed ohmic contacts were formed on the back of the substrates after annealing in N₂ for 5 min at 350 °C. Selective electrodeposition of metal into circular areas with a diameter of 400 μm were realized by using a photoresist mask defined by photolithography. Prior to the electrochemical process, both of the GaAs and the InP surfaces were chemically etched in order to remove native oxide.

The electrolytes containing various metal ions shown in table 17.2 were used. The electrochemical process used here consisted of controlled anodic etching of the semiconductor surface and subsequent *in situ* cathodic deposition of metal in the same electrolyte. Electrical supply waveforms used for etching and metal deposition are shown in figures 17.10(b) and (c). The anodic etching of GaAs was carried out under illumination by a tungsten lamp at zero applied potential with respect to SCE. On the other hand, avalanche pulse etching made using an electrical supply waveform as shown in figure 17.10(b) was used for anodic etching of the InP surface. Deposition conditions were optimized by investigating current–voltage characteristics in the electrolyte so as to minimize H₂ generation. Both dc and pulsed modes shown in figure 17.10(c) were used for metal deposition. The electrical conditions for metal deposition are also summarized in table 17.2.

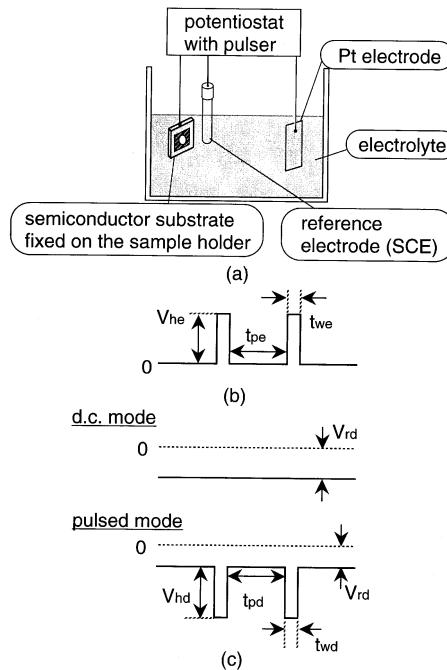


Figure 17.10. (a) Set-up of the novel *in situ* electrochemical process, and pulse waveforms used for (b) etching and for (c) metal deposition.

17.4.2.3 Electrical properties of Schottky contacts

The electrical properties of the Schottky diodes were investigated by I – V and C – V measurements. For the analysis of the I – V characteristics, a standard thermionic emission model was used, using the effective Richardson constant of 8.16 and $9.4 \text{ Acm}^{-2} \text{ K}^{-2}$ for n-GaAs and n-InP, respectively. The SBH values were obtained from the intercept of the $\log I$ versus V plot. To determine SBH values from the C – V measurement, conventional $1/C^2$ – V plots were used to obtain the values of the built-in potential (V_{bi}).

Typical forward I – V characteristics of the electrochemically produced n-GaAs and n-InP Schottky contacts are shown in figures 17.11(a) and (b), respectively, together with the values of the SBH and the ideality factor. The GaAs and InP Schottky contacts showed nearly ideal thermionic emission characteristics of ideality factors close to unity. In particular, for GaAs, the thermionic emission model seems to be applicable over a wide range of 5–8 orders of magnitude of the current. The SBH values from the I – V method for n-GaAs and n-InP Schottky diodes showed excellent agreements with those from the C – V measurements as also shown in table 17.2. This means that the interfaces are free from near-surface band modification, indicating formation of intimate interfaces

Table 17.2. Preparation conditions and electrical properties of GaAs and InP Schottky diodes.

Semiconductor	Metal	Plating condition	ϕB_n^{I-V} (eV)	n-value	ϕB_n^{C-V} (eV)
GaAs	Sn	dc $V_{rd} = 0.68$ V	0.69	1.06	0.70
GaAs	Co	pulse $V_{rd} = 0.5$ V; $V_{hd} = 0.7$ V $t_{pd} = 400 \mu s$; $t_{wd} = 50 \mu s$	0.78	1.03	0.80
GaAs	Ni	dc $V_{rd} = 0.9$ V	0.81	1.05	0.79
GaAs	Pt	pulse $V_{rd} = 0.4$ V; $V_{hd} = 1.0$ V $t_{pd} = 400 \mu s$; $t_{wd} = 10 \mu s$	0.92	1.08	0.91
InP	Ag	pulse $V_{rd} = 0.3$ V; $V_{hd} = 0.8$ V $t_{pd} = 400 \mu s$; $t_{wd} = 10 \mu s$	0.35	1.15	—
InP	Co	pulse $V_{rd} = 0.4$ V; $V_{hd} = 1.0$ V $t_{pd} = 400 \mu s$; $t_{wd} = 10 \mu s$	0.48	1.07	0.45
InP	Ni	pulse $V_{rd} = 0.4$ V; $V_{hd} = 1.0$ V $t_{pd} = 400 \mu s$; $t_{wd} = 10 \mu s$	0.53	1.12	0.50
InP	Pt	pulse $V_{rd} = 0.4$ V; $V_{hd} = 1.0$ V $t_{pd} = 400 \mu s$; $t_{wd} = 10 \mu s$	0.86	1.08	0.84

without any interlayers.

In figure 17.12, the observed SBHs of various metals on n-GaAs and n-InP contacts are plotted versus metal workfunction. In the case of Pt/InP diodes, SBH could be changed over 340 meV just by changing the electrochemical conditions which changed the average size of Pt particles. The result is compared with a recent theoretical prediction by Mönch [52] based on the MIGS model. This prediction gave the following formula for the slope factor defined by $S =$

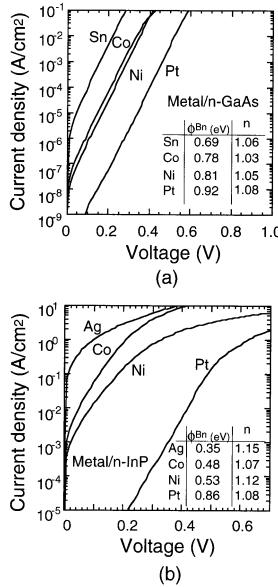


Figure 17.11. Forward I – V characteristics of various Schottky diodes for (a) n-GaAs and (b) n-InP.

$d(\text{SBH})/d(\text{work function})$.

$$1/S - 1 \approx 0.29(\varepsilon_\infty - 1)^2/\varepsilon_i, \quad (17.1)$$

where ε_i is the relative permittivity of the interface and ε_∞ is the permittivity of semiconductor. Using $\varepsilon_i = 4$ and $\varepsilon_\infty = 10.9$ for GaAs, $S = 0.12$ is obtained as plotted in figure 17.12(a). Similarly, $S = 0.16$ is obtained for InP as plotted in figure 17.12(b) using $\varepsilon_i = 4$ and $\varepsilon_\infty = 9.5$. It is seen in figures 17.12(a) and (b) that the experimentally observed metal workfunction dependence for both GaAs and InP deviates largely from the theoretical predictions by the MIGS model.

17.4.2.4 Surface/interface studies

In order further to understand the underlying mechanism for the observed large dependence of SBH on metal workfunction, various surface/interface studies were made. The surface morphology of the plated surfaces were investigated by the AFM and SEM. The chemical composition and profile of the interface was studied by XPS. Raman spectroscopy measurements of the plated surfaces were also made.

First, the SEM study showed that the electrodeposition in both dc and pulse modes proceeded with the formation of nanometre sized metal particles as shown in figures 17.13(a) and (b) taken at the initial stage of Pt electrodeposition on

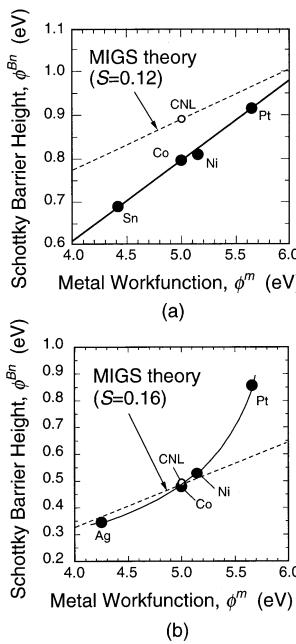


Figure 17.12. Metal workfunction dependence of SBHs of (a) n-GaAs and (b) n-InP.

n-GaAs and n-InP. Further deposition did not increase the particle size, but increased its number, leading to uniform coverage of the semiconductor surfaces by nanoparticles after a long plating time. $I-V$ measurements for a single Pt dot/n-GaAs contact and single Pt dot/n-InP contact using an AFM system with a conductive cantilever indicates that each Pt nanoparticle forms a well-behaved Schottky barrier with a high SBH and very low leakage currents both on the n-GaAs surface and on the n-InP surface. Details of the $I-V$ measurement for a single Pt dot/n-GaAs contact will be given in section 17.6.3.

The chemical properties of the etched surfaces and the Schottky interfaces were investigated by XPS. XPS spectra from the electrochemically produced Pt/n-GaAs contacts are compared with those by electron beam (EB) deposition in figure 17.14(a). The interfaces prepared by EB deposition indicated the presence of signals of Ga_2O_3 and elemental Ga in the Ga_{2p} spectrum. It is evident that an oxide layer and a decomposed GaAs layer coexist at the EB-deposited Schottky interface. On the other hand, neither oxide nor elemental Ga peaks were observed at the metal–semiconductor interface formed by the *in situ* electrochemical process. Exactly the same result was obtained for InP. This indicates that oxide-free intimate interface is realized by the novel *in situ* electrochemical process.

The Raman spectra obtained from the Pt deposited n-GaAs surface are

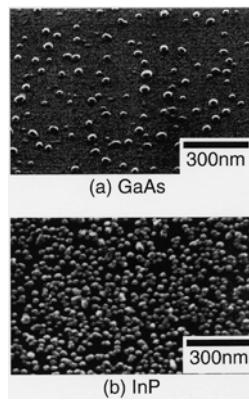


Figure 17.13. SEM images of nanometre-sized self-assembled Pt dots on (a) GaAs and (b) InP.

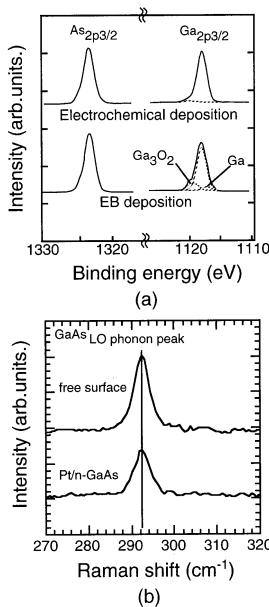


Figure 17.14. (a) XPS and (b) Raman spectra from a Pt/n-GaAs interface.

shown in figure 17.14(b). It is seen from figure 17.14(b) that no appreciable Raman peak shift and broadening of the GaAs LO phonon peak took place after electrochemical deposition. Since Raman peak shifts and broadening are related to the presence of stress, the result shows that the Pt deposited surface is remarkably free from stress. Exactly the same result was obtained for InP [53].

Thus, the observed strong dependence of SBH on the metal workfunction in GaAs and InP takes place on stress-free and oxide-free intimate contacts realized by uniform coverage with metal nanoparticles. According to the MIGS model, such an ideal interface should lead to strong pinning which is exactly opposite to the present observation. We believe that the result can be explained in terms of the DIGS model [56] where the interface disorder produces a DIGS continuum and pins the Fermi level at the charge neutrality level E_{HO} . Namely, the present electrochemical process proceeds with gentle precipitation of metal nanoparticles at an extremely low processing energy (several hundred millivolts) and at a low temperature. This results in the formation of coherent interfaces free from defects and random stress with a very much reduced DIGS density and leads to pinning-free Schottky barriers as demonstrated here.

17.5 Compound semiconductor quantum wave devices

17.5.1 Various lateral quantum wave devices investigated at RCIQE

At RCIQE, we were interested in lateral waveguide type devices for quantum computing and other analogue signal processing applications. To provide stronger electron confinement, we have proposed and tested two kinds of new Schottky gate structures which provide stronger confinement. They are Schottky in-plane gate (IPG) [6, 7] and Schottky wrap gate (WPG) [30] structures as already shown in figure 17.1. Using these gate structures, we have fabricated quantum wire transistors (QWTrs) [8] shown in figure 17.15(a) and gated Aharonov–Bohm (A–B) ring devices [10]. A GaAs 2DEG-based Schottky IPG QWTr having a channel length of 1 μm has shown the existence of first plateau of conductance quantization up to 100 K [9], demonstrating much increased subband spacing and much longer phase coherent length in these devices. In addition, heights of conductance quantization observed in such QWTrs have been found to show temperature dependence [11] that seems to be consistent with the recent theory based on the Tomonaga–Luttinger (TL) liquid, and, if this is really the case, it may further provide possibilities for new types of device applications through an artificial modulation to the wire [57]. IPG A–B ring devices [10] also showed the existence of electron wave interference at much higher temperature than in previous devices. Being encouraged by these results, we have recently investigated the feasibility of an IPG-based GaAs wave coupler device as discussed in the next section.

17.5.2 Schottky IPG single and coupled quantum wire transistors

17.5.2.1 Device structures and fabrication process

Schematic plan-views of the single and coupled Schottky IPG QWTr are shown in figures 17.15(a) and (b), respectively. In these structures, a one-dimensional channel is formed by depleting the 2DEG bar laterally by negative IPG bias.

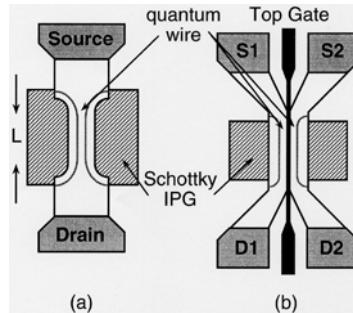


Figure 17.15. Schematic plan-views of (a) single and (b) coupled quantum wire transistors.

$\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ heterostructures were grown by standard MBE. Hall mobility and sheet carrier density of 2DEG were $9.3 \times 10^4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $6.4 \times 10^{11} \text{ cm}^{-2}$ at 77 K, respectively. For fabrication of both devices, standard electron beam (EB) lithography and wet chemical etching were applied to form the quantum wire. Two Pt Schottky IPG electrodes were formed by the *in situ* electrochemical process [6–9] on both sides of the channel after formation of source and drain Ge/Au/Ni ohmic contacts. The wire length L was determined from Pt IPG length and samples with various L from 0.6 to 1.6 μm were fabricated. For the coupled QWTr, an additional Cr/Au top gate was formed on the top of the quantum wire by the EB lithography and lift-off process.

17.5.2.2 Gate control in single QWTrs

The fabricated single and coupled quantum wire transistors showed good field effect transistor characteristics both at room and low temperatures. At low temperatures, the drain saturation current, I_{DS} , decreased linearly with increasing negative IPG voltage for single QWTr indicating ballistic transport [8]. Single QWTr also showed clear Shubnikov–de Haas (SdH) oscillation at low temperature and a nonlinear Landau plot of the IPG controlled QWTr was obtained. At high magnetic fields, the Landau index N linearly changed with inverse magnetic field B^{-1} , showing the same slope for the 2DEG wafer itself. This result indicated that the IPG gate controls the effective channel width W_e without changing the sheet carrier density, n_s according to the following formula, as first expected by Petrosyan and Shik [58],

$$W_e = W_{e0} - \frac{4\varepsilon V_{IPG}}{qn_s}, \quad (17.2)$$

where W_{e0} is the geometrical wire width.

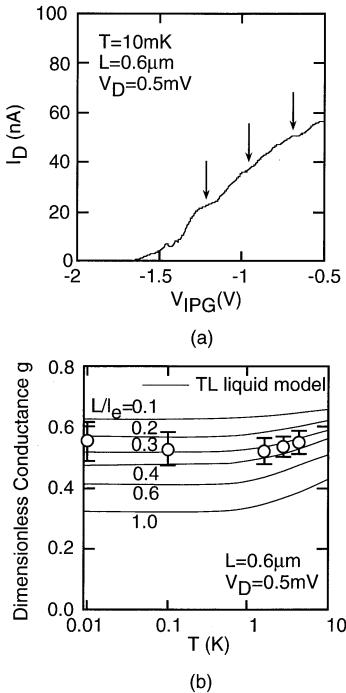


Figure 17.16. Conductance quantization in single QWTr. (a) I_D – V_{IPG} characteristics. (b) Temperature dependence of the value of the first quantized conductance plateau.

17.5.2.3 Conductance quantization in single QWTr

Figure 17.16(a) shows the behaviour of the drain current near pinch-off observed by applying negative IPG voltage at low temperature. The conductance of the channel was quantized and three plateaux were seen. The first plateau was visible up to 60 K. This means the IPG structure realizes strong electron confinement and is consistent with the large subband separation of about 10 meV derived in the same IPG structure [8]. The value of the quantized conductance at low temperature was found to be smaller than the ideal value $2e^2/h$ and the value gradually increased with increasing temperature. The observed temperature dependence of the values of the quantized conductance is shown in figure 17.16(b) with open circles. The full curves in the figure are calculated dependences based on the Tomonaga–Luttinger (TL) liquid model [59]. This model predicts the collapse of the conductance quantization at low temperatures due to electron–electron interaction. Thus, the height of each conductance step decreases gradually from its ideal quantized value down to a fractional value with decreasing temperature. In figure 17.16(b), typical calculated dependences are shown for $L/\ell_e = 0.1, 0.2, 0.3, 0.4, 0.6$ and 1.0 . The calculated curve with $L/\ell_e = 0.3$

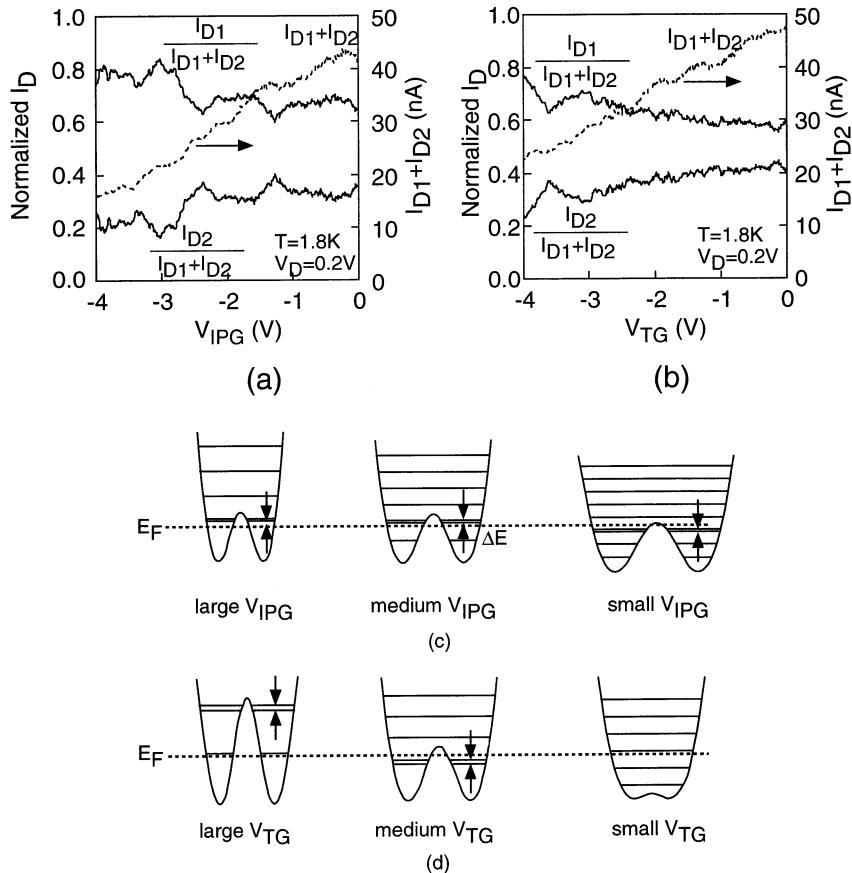


Figure 17.17. The I - V characteristics of the coupled QWTr at 1.8 K taken under (a) V_{IPG} swing with constant $V_{TG} = -1\text{ V}$ and (b) V_{TG} swing with constant $V_{IPG} = -2\text{ V}$. (c) and (d) showing explanations of observed behaviours in V_{IPG} swing and V_{TG} swing, respectively.

and $K_\rho = 0.7$ well represents observed values. Here L , ℓ_e and K_ρ are channel length, the mean free path of the electron and correlation exponent, respectively. The parameter K_ρ is related to the electron-electron interaction strength, and the derived value of 0.7 means that the interaction is repulsive. From the L/ℓ_e value of 0.3, ℓ_e is estimated to be about $2\text{ }\mu\text{m}$ which is consistent with the expected value from the Hall effect measurement. A similar temperature dependence of the quantized conductance was previously reported by Tarucha *et al* for a different wire structure [60]. The present result seems to be consistent with their report and shows the agreement with TL liquid based theory at a higher temperature range.

17.5.2.4 I–V characteristics of coupled QWTr

[Figure 17.17\(a\)](#) shows I – V characteristics of the coupled QWTr measured at 1.8 K. Here, two drain currents, I_{D1} and I_{D2} , which flow from electrodes S1 to D1, and from S1 to D2 in [figure 17.15\(b\)](#), respectively, were measured as a function of the IPG bias V_{IPG} . In figure 17.17(a), these two drain current were normalized by total drain current $I_{D1} + I_{D2}$ which is also plotted. During the measurement, the negative bias voltage on the top gate, V_{TG} , was kept constant. The drain current I_{D1} decreased with an increase of negative IPG bias and periodic peaks and dips appeared in its V_{IPG} dependence. At the V_{IPG} bias where I_{D1} showed a peak, I_{D2} showed a dip, and *vice versa*, whereas no such large variations took place in the total current ($I_{D1} + I_{D2}$). Three pairs of these appeared periodically. The measured dependence of the two drain currents on V_{TG} for a constant V_{IPG} is shown in [figure 17.17\(b\)](#). Here, only one pair of a peak and dip was observed.

In order to understand the observed behaviour, a computer simulation of the potential distribution and quantized energy levels was carried out for coupled QWTrs. Since this was not a self-consistent calculation taking account of potential modification due to electron occupation, exact gate voltage dependence could not be reproduced. However, it was useful to get insights into the physical situation. A symmetric mode and an antisymmetric mode are formed near the top of the barrier with an energy separation of ΔE as expected. The coupling length, L_T , or the length required to transfer electron from one electron waveguide to the other is given simply by

$$L_T = \frac{\pi \hbar v_F}{\Delta E} \quad (17.3)$$

where v_F is the Fermi velocity. As an example, $\Delta E = 0.5$ meV gives $L_T = 1.8 \mu\text{m}$ which is close to the channel length of the fabrication devices.

It was found by the simulation that the coupling length L_T changes delicately by both V_{TG} and V_{IPG} in this structure, and the magnitude of L_T can become comparable with the channel length under suitable bias conditions. Appearance of the pair of peak and dip can thus be explained in terms of electron transfer due to coupling between two waveguides. Namely, by changing gate biases, Fermi level crosses a energy range where coupled symmetric and antisymmetric modes are formed and electrons are transferred. The reason why multiple pairs of peak and dip were observed in V_{IPG} dependence, while a single pair was observed in V_{TG} dependence, is qualitatively explained using figures 17.17(c) and (d). In the case of V_{IPG} swing, the basic curvature of the potential is changed without changing the middle barrier height, whereas the middle barrier height is changed without changing the basic potential shape in V_{TG} swing. Thus, multiple chances of electron transfer are produced in the former, while only one chance is created in the latter.

17.6 Compound semiconductor single electron devices

17.6.1 Key issues for SEDs and efforts at RCIQE

For realization of ultralarge-scale integrated circuits (ULSIs) based on single-electron transistors (SETs), the following material and device related key issues should be addressed. First, a suitable technology for the formation of high-density arrays of position- and size-controlled ultrasmall wires or dots should be established, as discussed in section 17.3. Second, a suitable device technology for the construction of SET devices on these structures should be created. These devices should operate with acceptable high voltage gains and at acceptable high temperatures. Additionally, the structure of these devices should be suited for planar integration.

Recent efforts on SET device developments have emphasized high-temperature operation of SETs, resulting in room-temperature or near room-temperature operations of metal-based SETs [20] and Si-based SETs [21–24]. However, not much attention has been paid so far to the prospect of high-density integration and basic circuit performances such as the voltage gain of the basic inverter. In particular, the voltage gain is a fundamentally important characteristic for any circuit applications, particularly those based on Boolean architecture. However, as summarized in [table 17.3](#), almost all the SETs reported previously [22, 28, 29, 38, 61–67] have exhibited voltage gains below unity, except for one metal SET with intentionally enhanced gate capacitors [68] and one Si multi-tunnel junction SET based on a long wire with a length of $1.5\text{ }\mu\text{m}$ [69, 70].

At RCIQE, we are very much interested in developing device technologies for GaAs-based and InP-based SEDs for room temperature operating ultralarge-scale and high-speed planar integrated circuit applications. Previous split-gate control of the AlGaAs/GaAs HEMT wafer structures is an attractive approach suitable for planar integration of SEDs. However, the operation temperature of the split-gate SEDs were usually limited below several 100 mK due to the weak lateral confinement of electrons. To increase lateral confinement, we have recently proposed control of AlGaAs/GaAs double heterostructure wires by new Schottky IPG and WPG structures already shown in [figure 17.1](#). These gates have been applied to GaAs 2DEG bars and selective MBE grown InGaAs wires in order to construct SETs having lateral structures suitable for planar integration.

A series of logic and memory SEDs were proposed and fabricated at RCIQE, using IPG and WPG gates. For example, devices based on IPG are schematically shown in [figure 17.18](#). These devices are based on the Schottky IPG QWTr shown in figure 17.18(a). The operation principle of the basic Schottky IPG QWTrs is depletion control of quasi-1D wire by lateral electric field from Schottky IPG as shown in figure 17.18(a). The depletion width changes linearly with the IPG voltage as shown previously [58, 71]. By forming short IPGs, single- and multidot SET can be formed, as shown in figures 17.18(b) and (c), respectively. Resistive load type logic gates can be formed by using these SETs as the switching devices

Table 17.3. Voltage gain of SETs.

Metal/metal oxide junction based SET		
Zimmerli <i>et al</i> (1992)	2.8	(Al/Al ₂ O ₃ junction + interdigital capacitance) gate
Matsumoto <i>et al</i> (1996)	0.3	STM oxidation of Ti film + side gate
Nakamura <i>et al</i> (1996)	0.02	Al/Al ₂ O ₃ junction + side gate
Semiconductor based SET		
Kouwenhoven <i>et al</i> (1991)	0.15	GaAs 2DEG + split gate
Nakazato <i>et al</i> (1994)	0.04	δ -doped GaAs MTJ + air gap side gate
Austing <i>et al</i> (1995)	0.07	GaAs vertical resonant tunnel structure + side gate
Ohata and Toriumi (1995)	0.002	poly-Si island + MOS back gate
Jinushi <i>et al</i> (1996)	0.55	GaAs 2DEG etched wire + Schottky in-plane gate
Fujisawa <i>et al</i> (1996)	0.40	GaAs 2DEG FIB in-plane gate + top Schottky gate
Takahashi <i>et al</i> (1996)	0.9	pattern-dependent oxidation of Si wire + top MOS gate
Okada <i>et al</i> (1997)	0.26	selectively grown InGaAs QWR + Schottky wrap gate
Hiramoto and Ishikuro (1997)	0.17	Si point contact MOSFET
Nakazato <i>et al</i> (1997)	4.8	Si QWR + air gap side gate (wire length = 4 μ m)
Jinushi <i>et al</i> (1998)	1.4	GaAs 2DEG etched wire + 6 in-plane gate
Satoh <i>et al</i> (1998)	1.3	GaAs 2 DEG etched wire + 3 Schottky wrap gate

and the QWTr as the active load. Furthermore, by taking advantage of the fact that wire can be controlled from both sides by two IPGs, planar realization of Tucker's pseudo-CMOS gates seems possible [72].

III-V based SETs using IPG and WPG gates have realized higher temperature operation at 30–50 K [29,30]. Conductance oscillation and Coulomb blockade characteristics were observed for a Schottky IPG SET formed on a GaAs-based etched 2DEG bar [28] and an InP-based WPG SET formed on a selective MBE grown InGaAs quantum wire [38], respectively. Appearance of grouped peaks in the latter strongly indicated the existence of shell filling in this lateral device. The conductance oscillation characteristics observed in a three-dot WPG SET formed on a GaAs-based etched 2DEG bar suggested an artificial molecule formation due to dot coupling in this device [30].

In the next section, recent results on voltage gain of our SETs and on new single electron memory devices are described.

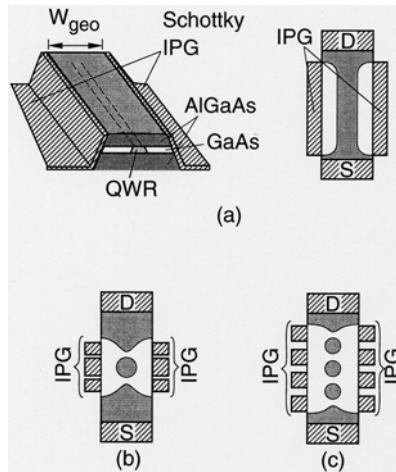


Figure 17.18. SEDs based on Schottky IPG QWTr. (a) QWTr, (b) single-dot SET and (c) multidot SET.

17.6.2 Voltage gain in GaAs-based lateral SETs having novel Schottky gates

17.6.2.1 Voltage gain of SETs

According to Likharev [3, 73], the voltage gain, G , of a SET is given by,

$$G \equiv \left. \frac{\partial V_{DS}}{\partial V_G} \right|_{I_{DS}}. \quad (17.4)$$

where G varies with the current level and temperature as shown by Likharev [73], but at a low current level near the Coulomb blockade threshold and at a low temperature, it is approximately given by

$$G = \frac{C_G}{C_D}. \quad (17.5)$$

In contrast to the conventional FET inverter circuit where the voltage gain is strongly dependent on the load impedance, the voltage gain of the resistive load SET inverter circuit or that of the pseudo-CMOS SET inverter [72] is much less sensitive to the load impedance, and the gain obtained by equation (17.5) is a good measure of the voltage gain in the actual circuit applications.

17.6.2.2 Theoretical study on gain of GaAs-based SETs having Schottky WPGs

In order to gain further insights on the voltage gain, the gain characteristics of novel, lateral GaAs-based SETs having Schottky WPGs were investigated theoretically. The structures of analysed SETs having two- and three-WPGs

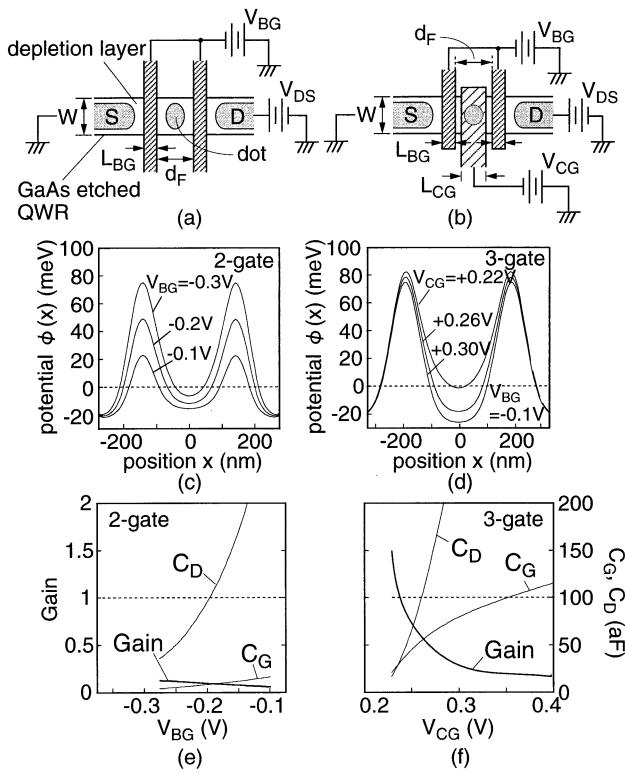


Figure 17.19. (a), (b) Structures, (c), (d) calculated potential profiles along electron transport direction, and (e), (f) calculated C_G , C_D and voltage gains for two- and three-gate devices, respectively.

are schematically shown in figures 17.19(a) and (b), respectively, with voltage supplies. In figure 17.19(a), the tunnelling barriers are controlled by controlling the shape and extension of the depletion layer using two-WPG barrier gates (BG). The dot potential is also controlled by the two-WPGs. This device is hereafter referred to as a two-gate device. On the other hand, the device in figure 17.19(b) possesses an additional centre gate for the realization of independent control of tunnelling barriers and dot potential by two outer WPGs and by the centre WPG (CG), respectively. This device is hereafter referred to as a three-gate device.

In the two-gate device, the same bias voltage, V_{BG} , is applied to the two-WPGs. On the other hand, two independent bias voltages of V_{BG} and V_{CG} are applied to the outer two-WPGs and the centre WPG in the three-gate device, respectively, for independent control of tunnelling barriers and dot charge.

The potential distributions in both type of devices were calculated for various device sizes and various bias conditions to clarify formation of a dot

Table 17.4. Device parameters, and theoretical and experimental voltage gains of WPG SETs in figures 17.19 and 17.21.

Type	W (nm)	d_f (nm)	L_{BG} (nm)	L_{CG} (nm)	Potential control	Maximum gain	
						Theory ^a	Experiment
2-gate	580	260	60	—	Barrier gate	0.15	0.2
3-gate	430	320	65	190	Barrier gate	—	0.3
					Centre gate	1.5	1.3

^a Evaluated at $n = 1$.

and tunnelling barrier as well as the control capability of the dot potential. The calculation was carried out self-consistently by solving the three-dimensional Poisson's equation using the successive over relaxation (SOR) method with quick convergence of the solution.

Figures 17.19(c) and (d) show examples of calculated potential profiles at the device centre along the electron transport direction for two- and three-gate devices, respectively. The width of the etched wire, W , gate lengths, L_{BG} and L_{CG} , separation between the two gates in the two-gate device and that between the two outer WPGs in the three-gate device, d_f , assumed in the calculations in figure 17.19 are summarized in table 17.4. As shown here, double barrier potential profiles are realized along the wire in both the structures. In the two-gate device, the tunnelling barrier and the dot potential are simultaneously changed with the barrier control gate bias. On the other hand, in the three-gate device, only the dot potential was changed by the centre gate bias, V_{CG} , keeping the shape of the potential barriers almost unchanged.

In order to see how well the dot potential, ϕ_d , and the number of electrons, n , are controlled by the gate voltage V_G in these devices, the dot potential modulation rate, $\partial\phi_d/\partial V_G$, and $n(V_G)$ were calculated, using the simulation program. According to the simulation results, the dot potential as well as the electron number change much more sensitively with the gate bias in the three-gate device than in the two-gate device. This is due to tighter charge control of the dot in the three-gate device because of its FET-like structure. From the potential simulation, the values of C_D and C_G were determined, and the gain was calculated using equation (17.5). The results are shown in figures 17.19(e) and (f) for the two-gate device and the three-gate device, respectively. In both devices, the gain increases towards pinch-off. However, only the gain of the three-gate device exceeds unity, while that of the two-gate device remains well below unity. This is due to tighter control of charge in the three-gate device.

17.6.2.3 Experimental study on voltage gain of GaAs Schottky WPG SETs

The device fabrication process was as follows. First, $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ double-hetero-quantum structure was grown by a standard MBE. The structure consisted of an undoped GaAs buffer layer (500 nm), an undoped AlGaAs bottom barrier layer (50 nm), a GaAs channel layer (20 nm), a spacer layer (10 nm) and an undoped AlGaAs top barrier layer (50 nm). Between the barrier layer and spacer layer, a Si δ -doping layer was inserted. Then, a GaAs wire with source and drain pad regions was formed by EB lithography and wet chemical etching. Subsequently, Au/Ge/Ni ohmic contacts were formed by conventional photolithography and the lift-off process. Finally, Cr/Au Schottky WPG electrodes were formed by EB lithography and the lift-off process.

Examples of observed conductance oscillations for the fabricated two- and three-gate devices are shown in figure 17.20. Both types of devices showed clear conductance oscillation peaks. These peaks are visible up to 30–40 K. The $I_D - V_G$ characteristics shown in figure 17.20(b) also has confirmed that independent control of potential in the dot and a tunnel barrier is achieved in the three-gate device as expected from the simulation.

The voltage gains of the fabricated devices were obtained by constructing the so-called Coulomb-diamond plots from the output characteristics ($I_D - V_{DS}$ characteristics). In this plot, the values of V_{DS} where the Coulomb blockade is lifted, are plotted as a function of the gate voltage. In this study, these points were determined by finding the first positive and negative peaks in the second derivative of each $I_D - V_{DS}$ curve. Then, the negative slope of the plot should give the voltage gain in the limit of small drain currents and should be equal to the gain in equation (17.5) according to the orthodox theory.

Figure 17.21 shows examples of Coulomb-diamond plots taken for two- and three-gate devices with dimensions listed in table 17.4. As seen in figure 17.21(a), the maximum voltage gain ($= \Delta V_{DS} / \Delta V_{BG}$) obtained in this two-gate device was 0.2. Although this value is much larger than that achieved in conventional split-gate SETs, it is still below unity.

As for the three-gate device, the dot potential was controlled by the two barrier gates in figure 17.21(b), and it was controlled by the centre gate in figure 17.21(c), respectively. In the case of potential control by the barrier gates, voltage gain again remained below unity as shown in figure 17.21(b), similar to the case of the two-gate devices. On the other hand, a large voltage gain of 1.3 was achieved by direct control of the dot potential by the centre gate as shown in figure 17.21(c).

Maximum voltage gains obtained in this study for two- and three-gate devices are summarized and compared with the theoretical values in table 17.4. Marked agreement is observed between experiment and theory. As expected from computer simulation, the highest voltage gain was achieved in the three-gate device with the dot potential controlled by the centre gate. This is the first report on achievement of voltage gain larger than unity in III–V SETs. We believe that

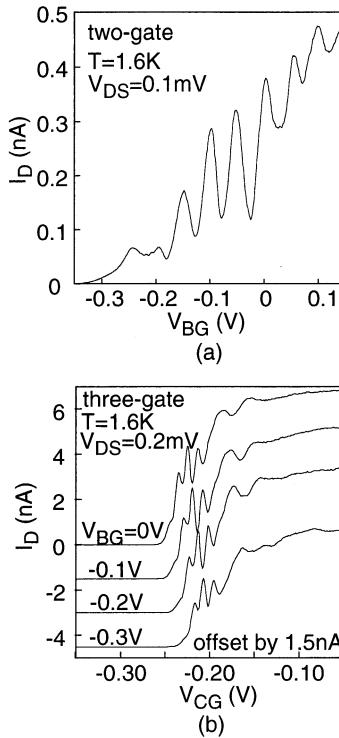


Figure 17.20. Example of conductance oscillation characteristics of the (a) two-gate device and (b) three-gate device.

this is due to efficient control of the dot potential by the centre gate as revealed by the simulation. Therefore, the three-WPG lateral SET structure having a FET-like structure seems to be promising for high-density integration and high-temperature operation of SET devices.

17.6.2.4 Gain increase by using multidot SET structure

The above method was to increase C_G in equation (17.5). An alternative method to increase voltage gain is to reduce C_D by using a multiple tunnel junction (MTJ), as pointed out by Nakazato and Ahmed [74]. To see the feasibility of such a technique in the IPG devices, an IPG multidot SET having five dots with tunnel junctions was fabricated [75]. At low temperature, the fabricated SET showed conductance oscillation and Coulomb blockade I - V characteristics. From the slope of the observed Coulomb diamond chart, maximum voltage gain of 1.4 was obtained.

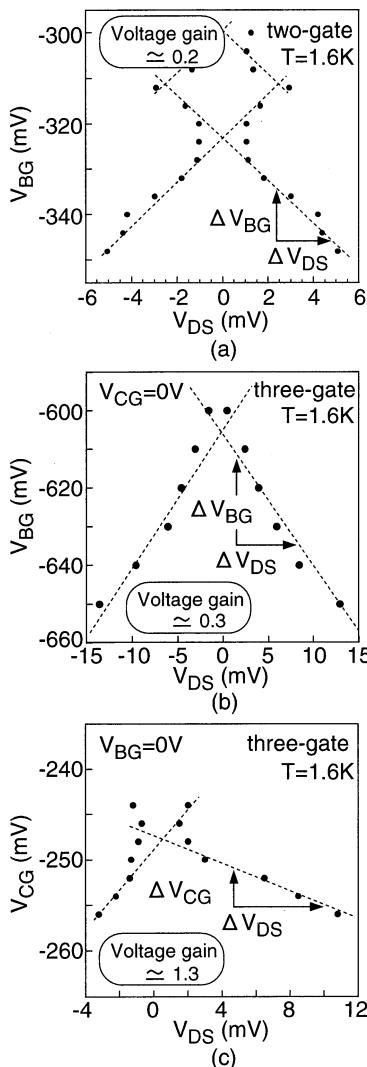


Figure 17.21. Coulomb diamond chart. (a) Two-gate device, three-gate device for the cases of potential controlled by (b) the barrier control gates and (c) the centre gate.

17.6.3 Single electron memory devices utilizing IPG quantum wire transistor and metal nanodots

17.6.3.1 Device structure and operation

The structure and the equivalent circuit of the single electron memory device proposed here are shown in figures 17.22(a) and (b), respectively. The device

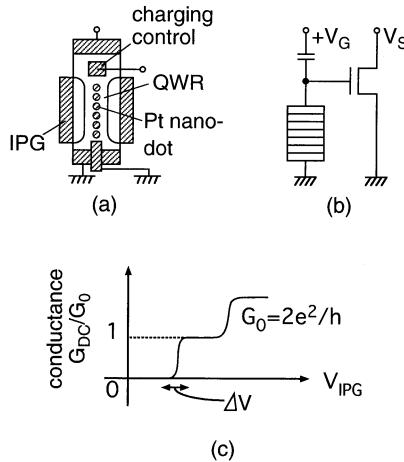


Figure 17.22. (a) Single electron memory structure, (b) equivalent circuit and (c) operation principle.

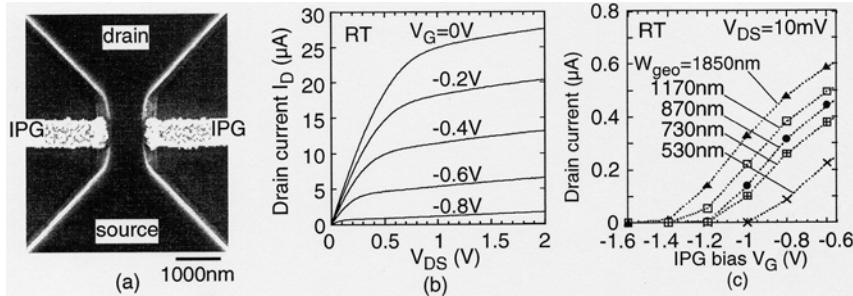


Figure 17.23. (a) SEM image of an IPG QWTr, (b) $I-V$ characteristics and (c) I_D-V_G characteristics.

consists of a QWTr memory switch and nanometre-sized metal dots formed on the top of QWTr which act as the memory node. The operation principle is to switch the QWTr between two conductance-quantized states of zero and $2e^2/h$ by the Coulomb blockade charging of the top metal dots as shown in figure 17.22(c) leading to extremely low-power memory operations.

In order to clarify the basic transport properties of Schottky IPG QWTrs, devices having geometrical wire widths of 600–1900 nm were fabricated on AlGaAs/GaAs HEMT wafers using the same procedure as described in section 17.5.2. An SEM micrograph of a fabricated IPG QWTr is shown in figure 17.23. The device showed good FET characteristics, as shown in figure 17.23(b). Detailed I_D-V_G characteristics in the linear region are shown in figure 17.23(c). Here, characteristics were systematically changed with the

width, indicating systematic change of V_{th} . Thus, fine control of the threshold voltage seems feasible in this type of QWTr.

In order to further investigate the feasibility of the memory SED shown in figure 17.22, a detailed computer simulation of potential distribution using the SOR method was carried out. The result indicated conductance switching is possible by charging metal dots with the diameter of 10–20 nm with a few electrons.

17.6.3.2 Process of nanodot Schottky contact formation

For realization of the proposed single electron memory device, dot formation is the key process. We have made an attempt to form nanometre-size Pt dots in a size- and position-controlled fashion on GaAs and InP substrates by combining our novel *in situ* electrochemical process explained in section 17.4.2 with the electron beam (EB) lithography.

Pt particles were first deposited in a self-assembled fashion on the unpatterned n-GaAs and n-InP substrates as shown in figure 17.13, in order to optimize the electrochemical condition for the formation of small Pt particles with uniform size distribution. The size distribution of such Pt particles could be represented in terms of the average in-plane diameter, the average vertical height and their standard deviations. As for the in-plane size, it was found that the average in-plane diameter of the particle and its uniformity strongly depended on the electrochemical conditions. Compared with the dc mode, the pulsed mode produced smaller, more uniform Pt particles on the III–V substrates. For example, the average diameter and a rms size deviation of Pt particles on InP formed by the dc mode were 29 nm and 10 nm, respectively. On the other hand, the pulsed mode produced smaller, more uniform Pt particles on InP with an average diameter of 26 nm and a rms deviation of 3 nm under an optimum condition.

Thus, attempts were made to form regular arrays of size- and position-controlled Pt dots by depositing Pt onto the patterned substrates having open circular windows as shown in figure 17.24(a) defined by EB lithography, using JBX-5000LS (JEOL). An electron beam with an accelerating voltage of 50 kV and a beam diameter of 50 nm was used.

Figure 17.24(b) shows the AFM image of the Pt dot array formed on n-GaAs patterned substrates with the application of one pulse. Pt dots were selectively formed only within the open windows with a period of 200 nm, thereby realizing precise dot-position control. This result indicates that the electrochemical reaction occurred only within the conductive area at the semiconductor/electrolyte interface and not through the EB resist film. Furthermore, it was also observed that Pt dots were formed at the centre of each open window without initially touching the window periphery. From the AFM observation, the diameter of a Pt dot was estimated to be about 20 nm which was much smaller than the size of the open window formed by EB lithography.

Finally, current–voltage (I – V) measurements of each nano-Schottky dot was

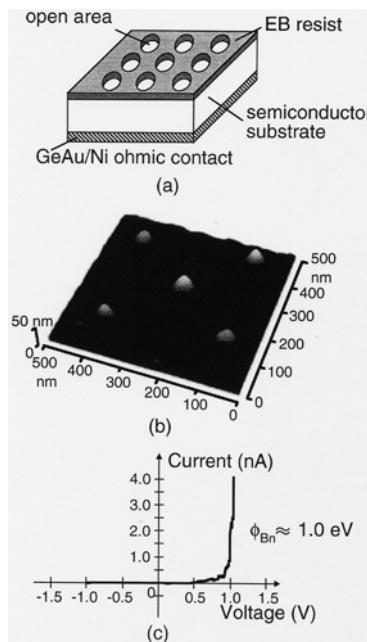


Figure 17.24. Fabrication and characterization of a Pt dot array. (a) Schematic drawing of a patterned substrate. (b) AFM image of a Pt dot array. (c) I - V characteristics of a single Pt dot/n-GaAs contact measured by an AFM system with a conductive probe.

carried out, using an atomic force microscopy (AFM) system with a conductive probe using Nanoscope II (Digital Instruments, Inc.). The typical I - V curve of Pt dot/n-GaAs contact are shown in figure 17.24(c). This conductive AFM system can provide the normal AFM observation by the optical lever method and the I - V measurements between the sample surface and back surface using a HP4156A parameter analyser. The conductive probe is fabricated from highly doped Si with a radius of curvature of about 10 nm. During the electrical measurements, the laser light was turned off in order to avoid the photovoltaic effect. The current-voltage (I - V) characteristics of a single Pt dot/n-GaAs contact clearly show rectifying behaviour. This result indicates that each Pt dot formed a well-behaved Schottky barrier with a high Schottky barrier height (SBH) (about 1.0 eV) and very low leakage currents on the n-GaAs surfaces. Thus, this electrochemical process is expected to be used for the fabrication of metal Schottky dot arrays that will be used in the proposed single electron memory device.

17.7 Conclusion

The present status of research on the quantum devices based on III-V compound semiconductors have been discussed in this paper, introducing and reviewing the recent results obtained by the author's group at the Research Center for Interface Quantum Electronics (RCIQE).

After giving a brief overview on quantum devices, recent results on quantum structure formation by MBE, formation and control of III-V nanostructure surfaces and interfaces, novel Schottky IPG quantum wave devices and novel Schottky IPG and WPG single electron devices have been presented and reviewed.

The main points that the authors would like to stress as concluding remarks are the followings:

- (1) III-V SEDs have good chances for next generation RT operating digital electronics. The combination of MBE/MOVPE selective growth and EB-lithography seems to be the most promising practical approach. Devices using IPGs and WPGs are suitable for planar integration and interface control has made remarkable progress.
- (2) Quantum wave devices are promising for high-performance complex analogue applications including quantum computation. However, RT operation may be difficult.
- (3) Key issues for the future are as follows.
 - (i) The understanding and design, based on quantum mechanics, are needed.
 - (ii) Further progress on material/fabrication technologies are needed.
 - (iii) Cooperative efforts from the basic physics level up to system architecture level are vitally important.

Acknowledgments

The work reported here is based on team efforts of many co-workers at RCIQE. The authors would like to express their sincere thanks particularly to Dr J Nakamura, Dr C Kaneshiro, Mr T Sato and Mr Y Sato for their cooperation.

The present work is supported in part by a grant-in-aid for Scientific Research on the Priority Area of 'Single Electron Devices and Their High Density Integration' from the Ministry of Education, Science, Sports and Culture (head investigator: H Hasegawa, Hokkaido University).

References

- [1] Moore G E 1979 *IEEE Spectrum* **April** 30–7
- [2] Hasegawa H 1993 *IEICE Trans. Electron.* **E76-C** 1045–55
- [3] Likharev K K 1989 *IBM J. Res. Develop.* **32** 144–58
- [4] Motohisa J and Fukui T These proceedings

- [5] Takemura R, Suhara M, Oobo T, Miyamoto Y and Furuya K 1997 *Japan. J. Appl. Phys.* **36** 1846–8
- [6] Okada H, Jinushi K, Wu N J, Hashizume T and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** 1315–19
- [7] Hashizume T, Okada H, Jinushi K and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** L635–8
- [8] Hasegawa H, Hashizume T, Okada H and Jinushi K 1995 *J. Vac. Sci. Technol. B* **13** 1744–50
- [9] Hashizume T, Okada H and Hasegawa H 1996 *Physica B* **227** 42–5
- [10] Kudoh T, Okada H, Hashizume T and Hasegawa H 1997 *Appl. Surf. Sci.* **117/118** 342–6
- [11] Nakamura J, Kudoh T, Okada H and Hasegawa H 1998 Presented at ISCS'98 (to be published in IOP conference series)
- [12] Nakazato K, Blaikie R J, Cleaver J R A and Ahmed H 1993 *Electron Lett.* **29** 384–5
- [13] Lent C S, Tougaw P D and Porod W 1993 *Appl. Phys. Lett.* **62** 714–16
- [14] Asahi N, Akazawa M and Amemiya Y 1995 *IEEE Trans. Electron Dev.* **42** 1999–2003
- [15] Mooij J E 1993 *Extended Abstracts of the 1993 Int. Conf. on Solid State Devices and Material* pp 339–40
- [16] Meirav U, Kastner M A and Wind S J 1990 *Phys. Rev. Lett.* **65** 771–4
- [17] Korotkov A N 1997 *Extended Abstracts of the 1997 Int. Conf. on Solid State Devices and Material* pp 304–5
- [18] Tarucha S, Austing D G, Honda T, van der Hage R J and Kouwenhoven L P 1996 *Phys. Rev. Lett.* **77** 3613–16
- [19] Eto M 1997 *J. Phys. Soc. Japan* **66** 2244–7
- [20] Matsumoto K, Ishii M, Segawa K, Oka Y, Vartanian B J and Hariss J S 1996 *Appl. Phys. Lett.* **68** 34–6
- [21] Ishikuro H, Fujii T, Saraya T, Hashiguchi G, Hiramoto T and Ikoma T 1996 *Appl. Phys. Lett.* **68** 3585–7
- [22] Takahashi Y, Namatsu H, Kurihara K, Iwadate K, Nagase M and Murase K 1996 *IEEE Trans. Electron Devices* **43** 1213–17
- [23] Tiwari S, Rana F, Hanafi H, Hartstein A, Crabbé E F and Chan K 1996 *Appl. Phys. Lett.* **68** 1377–9
- [24] Yano K, Ishii T, Hashimoto T, Kobayashi T, Murai F and Seki K 1994 *IEEE Trans. Electron Dev.* **41** 1628–37
- [25] Yano K, Ishii T, Sane T, Mine T, Murai F, Kure T and Seki K 1998 *IEEE Int. Solid-State Circuits Conference Digest of Technical Papers* pp 344–5, 462, 504
- [26] Yusa G and Sakaki H 1997 *Appl. Phys. Lett.* **70** 345–7
- [27] Akazawa M and Amemiya Y 1997 *Appl. Phys. Lett.* **70** 670–2
- [28] Jinushi K, Okada H, Hashizume T and Hasegawa H 1996 *Japan. J. Appl. Phys.* **35** 1132–9
- [29] Okada H, Fujikura H, Hashizume T and Hasegawa H 1997 *Solid-State Electron.* **42** 1419–23
- [30] Kasai S, Jinushi K, Tomozawa H and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 1678–85
- [31] Dutta A, Kimura M, Honda Y, Otobe M, Itoh A and Oda S 1997 *Japan. J. Appl. Phys.* **36** 4038–41
- [32] Fujikura H and Hasegawa H 1996 *J. Electron. Mater.* **25** 619–25

- [33] Fujikura H and Hasegawa H 1996 *Japan. J. Appl. Phys.* **35** 1333–9
- [34] Fujikura H, Hanada Y, Kihara M and Hasegawa H 1998 *Japan. J. Appl. Phys.* **37** 1532–9
- [35] Araki M, Hanada Y, Fujikura H and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 1763–9
- [36] Fujikura H, Araki M, Hanada Y, Kihara M and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 4092–6
- [37] Hanada Y, Ono N, Fujikura H and Hasegawa H 1998 *Solid-State Electron.* **42** 1413–17
- [38] Okada H, Fujikura H, Hashizume T and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 1672–7
- [39] Okada H, Kasai S, Fujikura H, Hasizume T and Hasegawa H 1997 *Japan. J. Appl. Phys.* **36** 4156–60
- [40] Kapon E, Shimony R, Bhat R and Hwang D M 1989 *Appl. Phys. Lett.* **55** 2715–17
- [41] Koshiba S, Noge H, Akiyama H, Inoshita T, Nakagawa Y, Shimizu A and Nagamune Y 1994 *Appl. Phys. Lett.* **64** 363–5
- [42] Kumakura K, Motohisa J, Fukui T 1997 *J. Cryst. Growth* **170** 700–4
- [43] Ono N, Fujikura H and Hasegawa H 1998 Presented at ISCS'98 (to be published in IOP conference series)
- [44] Shitara T, Tornow M, Kurtenbach A, Weiss D W, Eberl K and Klitzing K V 1995 *Appl. Phys. Lett.* **66** 2385–7
- [45] Nakamura Y, Tsuchiya M, Koshiba S, Noge H and Sakaki H 1994 *Appl. Phys. Lett.* **64** 2552–4
- [46] Fukui T, Kumakura K, Nakakoshi K and Motohisa J 1996 *Solid State Electron.* **40** 799–802
- [47] Sugaya T, Nakagawa T and Sugiyama Y 1996 *IEICE Trans. Electron.* **E79-C** 1568–71
- [48] Kodama S, Koyanagi S, Hashizume T and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** 1143–8
- [49] Hasegawa H, Kodama S, Koyanagi S and Hashizume T 1995 *Japan. J. Appl. Phys.* **34** L495–8
- [50] Fujikura H, Kodama S, Hashizume T and Hasegawa H 1996 *J. Vac. Sci. Technol. B* **14** 2888–94
- [51] Hasegawa H, Akazawa M, Matsuzaki K, Ishii H and Ohno H 1988 *Japan. J. Appl. Phys.* **27** L2265–7
- [52] Mönch W 1996 *Appl. Surf. Sci.* **92** 367–71
- [53] Hasegawa H, Sato T and Hashizume T 1997 *J. Vac. Sci. Technol. B* **15** 1227–35
- [54] Heine V 1965 *Phys. Rev. Lett. A* **138** 1689–96
- [55] Tersoff J 1984 *Phys. Rev. Lett.* **52** 465–8
- [56] Hasegawa H and Ohno H 1986 *J. Vac. Sci. Technol. B* **4** 1130–7
- [57] Fukuyama H, Kohno H and Shirasaki R 1993 *J. Phys. Soc. Japan* **62** 1109–13
- [58] Petrosyan G A and Shik A Y 1989 *Sov. Phys. Semicond.* **23** 696–7
- [59] Ogata M and Fukuyama H 1994 *Phys. Rev. Lett.* **73** 468–71
- [60] Tarucha S, Honda T and Saku T 1995 *Solid State Commun.* **94** 413–18
- [61] Nakamura Y, Cheng C D and Tsai J-S 1996 *Japan. J. Appl. Phys.* **35** L1465–7
- [62] Kouwenhoven L P, Johnson A T, van der Vaart N C, van der Enden A, Harmans C J P M and Foxon C T 1991 *Z. Phys.* **85** 381–8
- [63] Nakazato K and Blaikie R J 1994 *J. Appl. Phys.* **75** 5123–34

- [64] Austing D G, Honda T, Tokura Y and Tarucha S 1995 *Japan. J. Appl. Phys.* **34** 1320–5
- [65] Ohata A, Niiyama H, Shibata T, Nakajima K and Toriumi A 1995 *Japan. J. Appl. Phys.* **34** 4485–7
- [66] Fujisawa T and Tarucha S 1995 *Extended Abstracts of the 1995 International Conference on Solid State Electronics and Materials* pp 198–200
- [67] Hiramoto T and Ishikuro H 1997 *Appl. Phys. Lett.* **71** 3691–3
- [68] Zimmerli G, Kautz R L and Martinis J M 1992 *Appl. Phys. Lett.* **61** 2616–18
- [69] Nakazato K 1997 *Presented at Second International Research Workshop on Future Information Processing Technologies (Sapporo, Japan)*
- [70] Smith R A and Ahmed H 1997 *Appl. Phys. Lett.* **71** 3838–40
- [71] Okada H, Hashizume T and Hasegawa H 1995 *Japan. J. Appl. Phys.* **34** 6971–6
- [72] Tucker J R 1992 *J. Appl. Phys.* **72** 4399–413
- [73] Likharev K K 1987 *IEEE Trans. Magn.* **23** 1142–6
- [74] Nakazato K and Ahmed H 1995 *Appl. Phys. Lett.* **66** 3170–2
- [75] Okada H, Sato T, Jinushi K and Hasegawa H 1998 Presented at New Physics on Mesoscopic Structures '98 (to be published in *Microelectron. Eng.*)

Chapter 18

Quantum interference in corrugated conducting wire transistors

Kyoungwan Park

*Basic Research Laboratories, ETRI, Yusong PO Box 106, Taejon
305-600, Korea*

We present a review of the quantum behaviour of electrons in nanoscale semiconductor structures. Some of the nanostructures include an uneven gate conducting wire, an electrostatic Aharonov–Bohm ring, an electron grating structure, and a corrugated quantum wire. The electrical properties of the devices are attributed to the quantum interference effect.

18.1 Issues and challenges in semiconductor nanostructure

Demands for superior electronic devices and systems are increasing in order to achieve a network of high speed and capacity in present switching and computer systems. Many studies about nanoscale effects in recent years have shown that nanoscale semiconductor devices have a good potential for high speed and capacity in telecommunication and information technology. In terms of components, discrete devices consisting of nanoscale semiconductors exhibit quantum mechanical characteristics in device operations, which imply ultrahigh speed switching, ultralarge integrability, and multifunctionality. Such components will become more important as the information processing technologies begin to attract more attention, because in system applications quantum devices are expected to increase their share of functions. Semiconductor quantum devices showing functional properties that can serve the above-mentioned purposes will therefore become more attractive in the future.

Recent advances in crystal growth techniques and nanofabrication technology of semiconductor materials have enabled the production of ultrasmall structures to confine electrons in reduced dimensions, which have become the

subject of many studies in semiconductor physics. It is well known that electrons confined in a narrow potential well cause extraordinary characteristics in electrical and optical properties [1]. Experimental and numerical investigations have revealed quantum mechanical aspects of electron transport in the confining potential [2]. When the widths of the confining potentials are sufficiently small (less than a few hundred angstrom), quantum size effects in more than one dimension appear. Thereby low-dimensional physics and device studies have become possible. Semiconductor heterostructures are composed of thin layers of different bandgap semiconductors, which has been the best structure for confining the electron's movement in two-dimension [3]. It is no doubt that semiconductor heterostructures offer one of the most innovative concepts in the history of semiconductor science and have made major impacts in all aspects of electronic heterojunction device technology. High-speed electronic devices owe much to the heterojunction concept [4]. The simplest form of semiconductor heterostructure is composed of thin layers of different semiconductors, such as GaAs and AlGaAs, in an alternating manner. Electrons are confined at the heterojunction interface, where a confining potential well is formed. These two-dimensional structures of quantum wells can be further reduced to one-dimensional quantum wires and zero-dimensional quantum dots by various lateral confinement techniques [5]. The density of states is continuous for three dimensions, evolves to step-wise patterns in two dimensions (2D), and consists of discrete peaks with falling tails in one dimension (1D), while being completely discrete in zero dimensions. Some of the new transport phenomena observed in such structures include electron interference and tunnelling.

One of the most widely studied quantum behaviours is the resonant tunnelling in a potential well sandwiched by barriers [6]. In the resonant tunnelling diodes, the electron transport occurs perpendicular to the interface of a heterostructure and the current depends on the resonant energy states formed by the multiple reflection of the electron waves between the barriers. The vertical transport device, such as the resonant tunnelling diode, has an ultrahigh switching speed, because the travelling distance of the electrons between the emitter and the collector is very short (\sim a few tens of angstrom). This short distance can be easily manufactured through growth of the semiconductor substrate structures. However, the vertical transport device can be hardly integrated on a single chip for high-density operation.

On the other hand, for higher device density, faster switching speed, and lower power dissipation, silicon metal-oxide-semiconductor-field-effect-transistor (MOSFET) technology has become the prevailing technology for very large-scale integration applications today. The advances led to superior performances in computers and communication systems. Miniaturization of the MOSFET structure is a main technique to achieve very large-scale integration devices. For example, a MOSFET device of 40 nm gate length has been fabricated in research laboratories, and the results demonstrated that the nanoscale device can be operated well at room temperature [7]. Here is a partial answer to a key

question: where is the limit of the classical MOSFET scaling? This answer can be applied to nanoscale semiconductor devices. From a practical consideration, no fundamental physical effects arise in such small devices of 40 nm gate length. This is in part because the size of the active area in the device is not so small that the wave property of the electrons is easily destroyed. Therefore, further miniaturization of the device is necessary for a quantum mechanical device that has much better performances in the operation as described above.

In further scaling into the 10 nm regime key challenges should be examined with respect to both material and fabrication limitations; quantum mechanical effects are expected on such scale or smaller. They are mainly related to the impurity number fluctuations and to lithography, respectively [8]. Random fluctuations of the number of impurity atoms in the active area of a semiconductor device had been predicted to be a fundamental physics limitation to device miniaturization. The detailed microscopic impurity distribution will have non-negligible effects on device characteristics. The dependence of device characteristics on random fluctuations of the number of impurities and on the discrete microscopic random arrangement of the impurities has been simulated. The simulations of the random arrangement of the impurities predicted a spread in device properties, such as the threshold voltage and the current. The effect of the random arrangement becomes more important as the active area is scaled down further. Also, the current was found to be asymmetric in direction. Reducing device features in the lateral dimensions relies on the ability of lithographic patterning. Since photolithography has exceeded previously predicted resolution limits many times, new lithography techniques are required in order to bring the electronic device into the nanoscale regime. Although the challenges in implementing x-ray lithography techniques lie primarily in mask fabrication, x-ray lithography is a prime candidate for high-resolution patterning of critical features of 10 nm size. In the research environment, electron beam lithography has been generally used for the fabrication of nanoscale semiconductor structures. However, the key challenge for electron beam lithography lies in throughput. An alternative technique that has received recent interest is electron beam projection lithography, which offers several orders of magnitude improvement in throughput.

In such 10 nm scale devices, phase coherent transport of the electrons plays an important role in the device operations even at room temperature [9]. Phase coherent transport phenomena hold a promise for many new types of quantum devices with much improved performance over conventional devices. In particular, the lateral transport can be considered for an integrated circuit chip. In lateral quantum devices the current flows parallel to the confining potential wall. It is advantageous to employ such quantum devices in integrated circuit chips with low threshold voltage and high switching speed. However, a 10 nm scale device is not easy to fabricate with present-day capabilities.

Fabrication of lateral nanostructures mainly depends on the nanolithography technique. Nanometre scale patterning is an essential technology for these devices, as their critical feature size has to be comparable to certain lengths such

as the Fermi wavelength, the elastic scattering length, and the inelastic scattering length of electrons. One of these lengths is crucial in the design of critical device features for a certain type of experiment to be carried out [10]. In the observation of 1D transport the width of the confining potential is comparable to the Fermi wavelength. For single barrier scattering and resonant scattering experiments, for example, the active layer thickness should be in the range of the Fermi wavelength. For ballistic transport the drift length has to be much smaller than the elastic scattering length. For the investigation of quantum interference in conducting wires the length of the wire should be smaller than the inelastic scattering length.

Present nanolithography techniques cannot realize quantum interference at room temperature. The critical length for the quantum interference effect devices is the coherence length of electrons, which is a function of operating temperature. Since the inelastic scattering of the electrons with phonons destroys the phase memory of the electrons, the coherence length that is related to inelastic scattering decreases as temperature increases. Thus, quantum interference devices should at present be operated at low temperature. In this paper, we give a review of the electron interference effect and devices, and provide experimental results on the physical properties of various nanoscale devices. An uneven gate conducting wire, an electrostatic Aharonov–Bohm ring structure, an electron grating structure, and a corrugated quantum wire are all discussed.

18.2 Quantum transport; electron interference

The low-temperature electrical properties of a nanoscale semiconductor wire are nonlinear; they do not scale with length. The classical model for the resistance of the wire cannot account for the nonlinear properties at all. If a semiconductor wire is so narrow that the width and thickness are comparable to the Fermi wavelength, and the length is less than the elastic and inelastic scattering lengths, then the energy and momentum are quantized, the electron motion is coherent. Therefore, the electrical properties of the wire have to be determined quantum mechanically, while the wire resembles an electron wave guide [11]. An electron wave guide is a conducting wire that is so clean and so small that the electron waves can propagate as guided modes, that are characterized by the geometry of the conducting wire without loss of phase coherence. If no quantization of the transverse energy and momentum occurs, the electron transport is not characterized by the wave guide model. However, the quantum mechanical wave properties of the electrons should still be taken into account in a semiconductor wire of which the length is less than the elastic and inelastic scattering lengths.

Electron transport is influenced by quantum mechanical effects through both elastic and inelastic scattering. Elastic scattering, such as might occur at an impurity, changes the distribution of the electrons in the transverse quantum states, but is phase-deterministic. The electron transport associated with the phase

of the electron wave is not destroyed by an elastic scattering event. However, inelastic scattering destroys the phase memory of the electron wave. Actually, coherent transport is possible when the semiconductor wire is smaller than the inelastic scattering length.

Electrons possess wave-like properties at scales comparable or smaller than the coherence length, which is an inelastic scattering length in a particular material system, and therefore they exhibit interference much like electromagnetic waves. For the interference effect the most important parameter is the phase of the wave. The wave description based on the Schrödinger equation can become necessary in describing electron transport in submicrometre structures. In response to an applied magnetic and electric field, the magnetic vector potential and/or the electrostatic potential in the conducting path can control the phase of electrons in an electrical device. The potentials explicitly appear in the Schrödinger equation. Therefore, these potentials directly affect their physical quantities [12]. The phase change of the electron wavefunction in a magnetic vector potential and an electrostatic potential is described by the Aharonov–Bohm (AB) effect: $\Delta\Phi = e/\hbar \int \mathbf{A} \cdot d\mathbf{s}$. The magnetic vector potential contributes to the first three of the components of the four-vector (\mathbf{A}, V) in the inner product with the travelling distances, while the electrostatic potential contributes to the fourth term. This term contains the electrostatic potential V associated with electric field and time t . The effect of the electrostatic potential is to cause a phase shift in the wavefunction: $\Delta\Phi = 3/\hbar \int V dt$. The electrostatic phase shift can be interpreted as the product of the electrostatic potential experienced by the travelling electrons and the time travelled under the influence of the electrostatic potential.

There have been many studies of the AB effect. In experiments, a ring type conducting path was commonly used [13]. The electron wave is split into partial waves in two branches and then recombined in the closed loop structure. If coherence is maintained between the two paths, interference effects can be observed. The magnetic flux enclosed by the loop shifts the relative phase of the partial waves in the branches according to $\Delta\Phi$. Also, an electrostatic potential difference between two paths causes a phase difference between the partial waves following two paths. A phase shift of 2π is achieved by an enclosed magnetic flux h/e . In conducting loops, this phase shift results in conductance oscillations of the device in the magnetic field region. Only the magnetic flux surrounded by the ring affects the phase shift between two paths. The width of the branches of the ring should be much smaller than the diameter in order that the closed electron paths in the ring are uniform. Thus, the small spread in the magnetic flux makes the intensity of the conductance oscillation large. It seems that for practical applications the electrostatic AB effect, that is electrostatic phase shift, would be more suitable than the magnetic effect [14]. However, the electrostatic AB effect has not yet been clearly demonstrated in solids. It should be mentioned that the AB effect was originally proposed as an experiment to demonstrate that observable effects could be produced by electromagnetic potentials even though

there are no fields along the electron paths. In the case of the electrostatic effect, it is very difficult to keep the electrons from feeling the electric field. Actually, in all the experiments in solids for the electrostatic AB effect the electric fields penetrate the paths of the electrons. Although electric fields penetrate the electron paths, the phase of the electron wave is still controlled by the effective electrostatic potential. An electron interference effect device can be operated if a phase difference between different paths exists. In this device the general formula for the phase difference between the electron waves is useful to understand the interference behaviours: $\Delta\Phi = \Delta(\mathbf{k} \cdot \mathbf{l})$, where \mathbf{k} is the wavevector of the electron wave under the electrostatic potential, and \mathbf{l} is the path length. The above AB phase is a quantum mechanical feature of the phase of wave mechanics when the electrons are in magnetic and electrostatic potentials.

18.3 Review on lateral quantum interference devices

There have been many experimental studies reporting on the quantum interference phenomena of electrons [15–24]. As the size of electronic devices becomes comparable with the coherence length of the electron, the devices are known to reveal quantum mechanical electron interference effects which need to be explored for novel device functions. Webb

emphet al, for example, reported on the oscillatory magnetoresistances in submicrometre-diameter Au rings [15], while others reported on the interference effects of the AlGaAs/GaAs two-dimensional electron gas (2DEG) in ring structures [16], which are induced by AB interference. Since the quantum interference effects are displayed largely in the ballistic transport regime, attempts have been made to utilize these effects for high-speed quantum interference transistor devices or electron interferometers. Various forms of electrostatic quantum interference transistor structures, in which the source-drain conductance can be controlled largely via the electron interference effect by the gate voltage, have been proposed and fabricated for potential device applications [17].

In an electron interferometer, an electric potential difference between two paths causes the partial waves following two paths to recombine with different phases. Washburn *et al* investigated the electrostatic AB effect in a metal ring by employing an electric gate field in the plane of the ring [20]. With the application of the electric field, they observed a shift of the h/e oscillations in the magnetoresistance, which changed with the gate field. However, they speculated that the shift might be due to the shift of the carrier trajectories induced by the electric gate field. In the study by others, the shift of the magnetic h/e AB oscillations was measured in a AlGaAs/GaAs 2DEG-based ring structure with a metal gate placed on one of the electron paths [21]. No clear indication of the electrostatic AB effect was resolved in the transresistance of the 2DEG ring as a function of gate voltage. More recently, electron interference controlled by a gate voltage has been observed in a Young's double-slit configuration [22] and

a double-quantum-wire interferometer [23], and their interference patterns agree very well with the interference conditions of the phase difference between the two paths.

Because the electrostatic potential disturbs the carrier density, [21, 23, 24] carrier trajectories, [20, 21] the number of conduction modes [21, 24] as well as the electrostatic AB phase, the results from the experimental studies are not clear enough to extract meaningful information on the electron's behaviour in such devices. Theoretically, the existence of two different interference conditions due to the electrostatic AB effect in a one-dimensional ring has been predicted in the ballistic transport regime [25]; one is caused by the conventional interference of transmitted electrons, and the other arises due to the interference of an electron travelling completely around the ring and interfering with itself at the point of entry.

The earlier experiments in disordered normal-metal rings displayed only *one* set of h/e magnetoresistance oscillations [15] and their phase shift in electrostatic AB effect experiment [20], whose behaviours agree well with the theoretical considerations based on the multichannel scattering properties [26] or ensemble averaging in the conducting paths [27]. The number of channels (or traverse modes) at the Fermi energy is large in a metal ring, $N_T > 1000$, so that multichannel effects and interchannel scattering (or mode mixing) occurring everywhere around the ring particularly randomize the phase of the electron in the electron's multiple circulation around the ring; this is called 'localization type interference' [25]. The experimental results clearly demonstrated that the h/e magnetoresistance oscillations purely originated from the phase difference between two arms of the ring. On the other hand, in a typical quasi-one-dimensional wire such as a AlGaAs/GaAs-based quantum wire, the current is carried by only a few ($N_T < 10 \sim 20$) channels at the Fermi energy (at 17 mK) with minimal scattering. The mode mixing effect mainly occurs in the irregular part of the wire with respect to width, for example, at the junction parts for the current splitting or recombining in a ring structure, and the effect is supposed to be small. When the magnetic field penetrates the annulus of the ring, two multiple localization paths typically enclose different amounts of flux. They then no longer oscillate with the same phase in the sum over paths. So we expect that the localization type interference fringes are possibly exhibited in the transport measurement of a ballistic quasi-one-dimensional ring at low magnetic fields.

Experimental studies of the conductance properties of semiconductor ring structures gave evidence of the quantum wave guide effect in the electron propagation [16, 28]. Since then, a number of experimental results in quantum interference transistor structures have been reported; these show the electron interferences associated with the multiple transmissions [18], the multiple reflections [29], the lateral electron interference [19, 30] and the electrostatic AB effect [31]. On the basis of such results on phase coherent transport, electron diffraction phenomena as well as electron interference can be expected to appear in a ballistic semiconductor device where the path length of electrons is less than

their coherence length.

Electron diffraction by periodic arrays of quantum antidotes was investigated by a quantum mechanical calculation [32]. It was found that the electron diffraction occurs through multiple channels characterized by transverse wavevectors that differ from the wavevector of the incident electron by the reciprocal lattice vectors of the periodic arrays as one can expect from the Fraunhofer diffraction of light.

Electron waveguiding in 1D semiconductor nanostructures was identified through the observation that the conductance in quantum point contacts is quantized at discrete values of $2e^2/h$ [33]. The conductance quantization originates from the quantum confinement effect, i.e. the transverse resonant states in the quasi-1D structure. Longitudinal resonant states are established by a longitudinal quantum confinement potential, such as a double barrier potential in a resonant tunnelling diode. Also, the longitudinal resonant states can be formed by a periodically modulated potential in the 1D conducting wire. The formation of the longitudinal resonant states in an artificial 1D crystal allows not only a demonstration of an energy band structure in solid state physics, but also offers potential applications in quantum devices.

Formation of an energy-band structure in an artificial 1D crystal was investigated experimentally [34] and theoretically [35]. Kouwenhoven *et al* studied the magnetotransport properties of a channel with a periodically modulated width [34]. The 1D channel was defined by a split-gate in their experiment; one gate electrode was straight, but the other gate electrode was corrugated with fifteen notches in it. Miniband formation analogous to minibands in superlattice structures was discussed for the conductance oscillations in the voltage range of the straight gate. However, the changes of the gate voltage in the 1D conducting wire formed by a split-gate influence the energy of the transverse modes as well as the Fermi energy [11]. Therefore, effects related to the modification of the transverse modes have to be considered with regard to miniband formation.

18.4 Various characteristics of quantum interference devices

18.4.1 Uneven gate conducting wire

The amplitude of the interference term was proportional to $\cos(\Delta(\mathbf{k} \cdot \mathbf{l}))$, where \mathbf{k} is the electron wavevector, and \mathbf{l} is the travelling distance of electrons. The experimental results in a Young's double-slit configuration [22] and a double-quantum-wire interferometer [23] indicate that the electron phase is modulated by the gate voltage, which is analogous to changing the index of refraction in the Mach-Zender type optical wave guide interferometer.

In order to induce laterally different changes in the phase of the electron wavefunctions, we adopted a narrow-wide-narrow conducting path as a source-drain path in a transistor type structure, thereby producing an electron

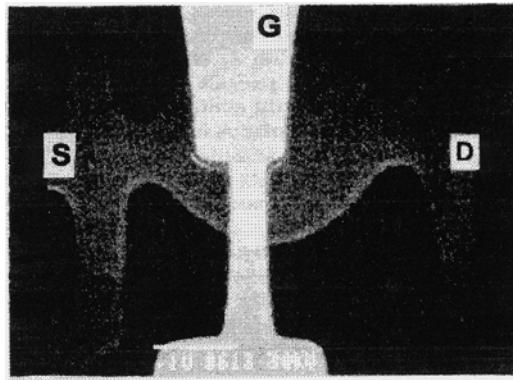


Figure 18.1. Scanning electron microscopy picture of the conducting structure with an uneven metal gate. The bright areas are the metal gate deposited on top of the AlGaAs/GaAs heterostructure.

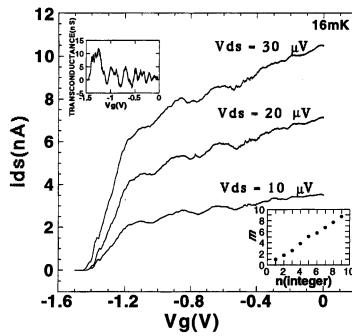


Figure 18.2. The measured source-drain current I_{ds} versus the gate voltage V_g for different source-drain voltages, V_{ds} , at $T = 16$ mK. Upper left inset: transconductance g_m versus the gate voltage. Lower right inset: constructive interference mode index m , $m = (k_f(V_g = 0) - k_f(V_g))\Delta l/2\pi$, versus integer number n .

interferometer with an uneven metal gate placed on the wide path. In this structure we are concerned with the lateral electrostatic quantum interference [30].

The conducting path of 2DEG was fabricated by means of electron beam lithography and subsequent chemical etching on a modulation-doped AlGaAs/GaAs heterostructure grown by molecular beam epitaxy. To form the gate, electron beam lithography and lift-off techniques were employed. This second step defined the $1.0/0.5 \mu\text{m}$ long uneven AuNi gate. Figure 18.1 shows the scanning electron microscopic view of the fabricated sample. In this study, source-drain current (I_{ds}) was measured as a function of the gate bias voltage (V_g) by using the lock-in technique.

The source-drain current decreases monotonically in a normal gate transistor as the reverse gate bias voltage increases, even at low temperature. In contrast to the classical behaviour, figure 18.2 shows several oscillatory features on the source-drain current of the uneven gate narrow-wide-narrow transistor. The maximum variation of the current oscillation is about 10% of the total current. In order to verify that these oscillations are truly due to the lateral quantum interference effect, we divided the conducting paths into two different regions under the uneven gate. One consists of all paths passing underneath the gate of 0.5 μm length and the other contains all paths underneath the gate of 1.0 μm length. Since the electron transport is ballistic in this structure, most of the conducting electrons move straight between the source and the drain, thereby producing a path difference of 0.5 μm between the two paths. As the sheet carrier concentration of 2DEG (n_s) varies with the gate voltage, the electron Fermi wavevector (k_f) under the gate also changes according to $k_f = (2\pi n_s)^{1/2}$ in a two-dimensional electron system. Because the narrow regions of source and drain are quasi-one-dimensional and the width of the conducting path changes adiabatically along the longitudinal direction, the magnitude of the wavevector of the conducting electrons is approximately k_f . The phase difference between the two different paths is then $(k_f V_g = 0) - k_f(V_g) \Delta l$, where Δl is the path length difference between two paths under the gate. Thus, constructive interference between the two paths occurs when $(k_f(V_g = 0) - k_f(V_g)) \Delta l = 2n\pi$, where n is an integer. In figure 18.2, the current maxima are seen to coincide with the positions calculated from the constructive interference condition. The linear approximation of the sheet carrier concentration depending on the gate voltage was used in the calculation of the Fermi wavevector. The interference mode indices, $m = (k_f(V_g = 0) - k_f(V_g)) \Delta l / 2\pi$, are calculated from the peak positions in gate voltage, and are plotted along integer numbers in the right inset of figure 18.2, where the calculated index numbers for the oscillation peaks agree well with integer numbers. Also, the transconductance (g_m) versus the gate voltage, $g_m = (\partial I_{ds} / \partial V_g)$, for V_{ds} (source-drain voltage) = 10 μV , is shown in the left inset of figure 18.2. Multiple negative transconductance oscillations are clearly seen in the regions of the negative slopes in the curve of the source-drain current versus the gate bias voltage.

18.4.2 Electrostatic Aharonov–Bohm ring transistor

In a semiconductor AB ring structure, a laterally corrugated metal gate was used in order to have a quantum interference effect. The conducting path of 2DEG has been fabricated using electron beam lithography and subsequent chemical etching on a modulation-doped GaAs/Al_xGa_{1-x}As ($x = 0.3$) heterostructure grown by molecular beam epitaxy. The average ring diameter is 1.9 μm , the width of the conducting path is 0.3 μm in the lithographic length, and the distance between the measurement probes across the ring is 3.8 μm . The heterostructures were wet-etched down to the 2DEG layer to form a well-shaped pathway. Lateral

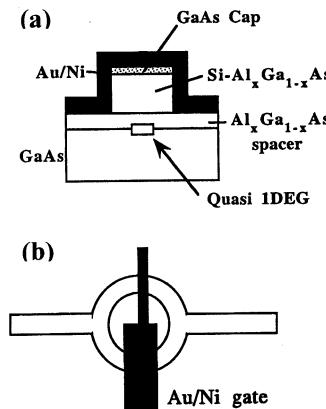


Figure 18.3. (a) Cross-sectional view of the gated conducting path. (b) Schematic diagram of the sample ring geometry.

depletion further reduces the conducting width. To form the gate, electron beam lithography and lift-off techniques were employed. The second step defined the corrugated Au/Ni gate; the metal gate length is $0.2 \mu\text{m}$ on the one conducting path of the ring, and $1 \mu\text{m}$ on the other. Figure 18.3 shows a schematic diagram of the fabricated sample.

The carrier concentration and mobility in this substrate at 1.5 K , as deduced from the measurement of Shubnikov–de Haas oscillation in a two-dimensional bar ($50 \times 150 \mu\text{m}$), were found to be $n = 3.2 \times 10^{11} \text{ cm}^{-2}$ and $\mu = 5.5 \times 10^5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively. We measured longitudinal resistance in the applied magnetic fields and gate voltages by using the lock-in technique and 10 nA driving current at 17 mK . We found periodic magnetoresistance oscillations which are in direct correspondence with the penetration of the flux $\phi_0 = h/e$ through the average area of the annulus. This result demonstrates that the mesoscopic ring was well fabricated.

Figure 18.4 shows the magnetoresistance spectrum when the gate voltage is -206.2 mV . The h/e magnetostatic AB oscillations appear in the range $B > \sim 40 \text{ G}$. However, two sets of oscillations with period h/e are clearly seen near zero magnetic field; for approximately $-40 \text{ G} < B < +40 \text{ G}$, they are shifted by π with respect to each other. The two sets of oscillation peaks are almost equal in magnitudes, so as to make the period of the resistance oscillations to appear as $h/2e$ overall. However, we rule out that these oscillations are due to the interference effect caused by the time reversal paths around the ring, because the appearance of the $h/2e$ oscillations is dependent on the gate voltage [36]. Since for the pair of electron waves in the time reversal paths the same amount of phase shift by the gate potential in the sample structure must be found, the $h/2e$ oscillations between the time reversal paths should be independent of gate

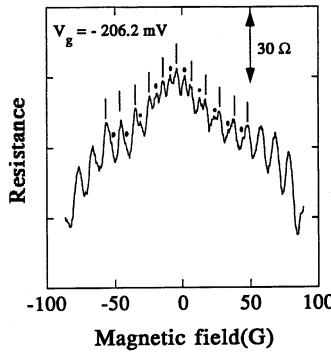


Figure 18.4. Magnetoresistance spectrum at $V_g = -206.2$ mV. Dots represent the newly formed h/e oscillations, while bars represent the background h/e oscillations.

voltage.

In this ring structure, the ballistic nature of the transport leads us to suppose that the electrons in the ring annulus can go clockwise or counterclockwise around the ring more than once. That is, we should consider the interference effect by the localization paths as well as the Mach–Zender type paths (which are composed of only transmitted-electron paths in two arms) [37, 38]. We can assume that the back scattering effect and the mode mixing effects are small on the source/drain junction side. Then the total interference intensity can be represented by the addition of the two Fabry–Perot type resonances (multiple reflection effects with clockwise circulating wave and with counterclockwise circulating wave) and their interference; the first-order term in this interference effect between two Fabry–Perot resonances is the Mach–Zender type interference. We attribute the behaviours of the new and background h/e oscillations shown in figure 18.4 to the combined effects of two Fabry–Perot resonances shifted to the opposite direction by the effective gate potential in the magnetoresistance measurements.

Figure 18.5(a) shows the consonant changes of the amplitude of the new and background h/e peaks versus the gate voltages, which are deduced from the magnetoresistance measurements. The competition and the oscillatory change of the amplitudes support the simple argument of two Fabry–Perot resonances. The measured resistances as a function of the gate voltage at zero magnetic field are depicted in figure 18.5(b). The measurements have been done in the range of low gate voltages without any mode depopulations in between. The measurements of the h/e magnetoresistance oscillations in the presence of the electrostatic potential imply that the electrostatic phase shift between the successive resistance minima in the electrostatic measurement is 2π , at least in the range of the gate voltages discussed above [31].

The electrostatic AB phase shift is induced by the time-dependent scalar potential, while no force is exerted on the electron wavefunction. When we apply

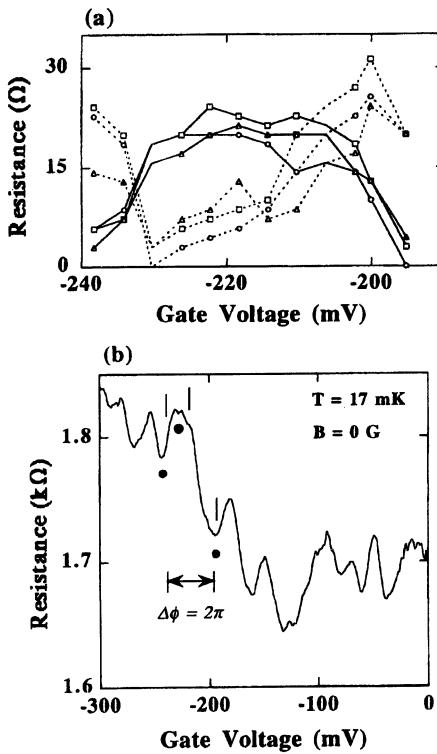


Figure 18.5. (a) The resistance changes of the amplitude of the new and background h/e peaks versus the gate voltages; the full curves are the changes when $B = -19$ G (○), -8 G (□) and 3 G (△) for the new h/e peaks and the dotted curves are the changes when $B = -13$ G (○), -3 G (□) and 7 G (△) for the background h/e peaks. The oscillatory component of the resistance is the height from the overall baseline in each magnetoresistance spectrum. (b) The resistances as a function of gate voltage at $B = 0$. Dots represent the constructive/destructive interferences caused by the electrostatic AB phase shift; $V_g = -244$ mV, -226 mV and $V_g = -197$ mV. Bars represent the gate voltages where the constructive/destructive interference effects occur in figure 18.4.

the gate voltage, the electrons experience classical electrostatic forces whose direction is perpendicular to the 2DEG layer. Thus, the small gate voltages do not affect the electron trajectory as much. The Fermi wavevector, however, changes a little, which results in a change of the phase for the quantum state of the electron [39]. The electrostatic AB phase acquired in the gate potential is given by $\Delta\phi = e/\hbar V_{\text{eff}}\langle\tau_t\rangle = \Delta(kl)$, where $\langle\tau_t\rangle$ is the harmonic mean of the travelling time through the gate potential, k is the wavevector under the gate and l is the gate length [14, 39].

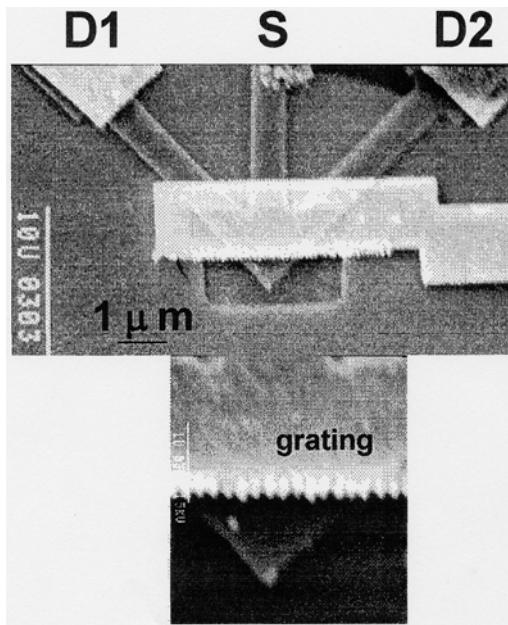


Figure 18.6. Scanning electron micrographs of the quantum diffraction transistor and the reflection grating structure.

18.4.3 Electron diffraction transistor

A single slit quantum diffraction transistor with multiple drains was proposed to have ultrafast extrinsic switching speed and a potential for multifunctionality [40]. However, there were some difficulties in fabrication of submicrometre size multidrains and in measurements of diffraction currents due to low diffraction efficiency. So, a reflection type multiple-slit aperture, which is a diffraction grating, is considered to facilitate the device fabrications and to increase the diffraction current for good device performances in quantum diffraction devices [41].

In the investigation of the electron diffraction, we fabricated a AlGaAs/GaAs-based conducting wire with a grating structure which has a source, a gate, and two drains as shown in figure 18.6. The conducting wire was bent, so that the multiple-slit type reflection grating was planted at the corner of the bent wire. We measured drain current (I_{DS}), i.e. conductance (σ) versus applied gate voltages (V_G) at various levels of source-drain voltages (V_{DS}) by using the lock-in technique at 16 mK. While the drain current was measured at D1, the drain D2 was grounded and *vice versa*. The measured values were calibrated at the constant drain-voltage bias mode by considering the effect of conductance change during the gate voltage sweep.

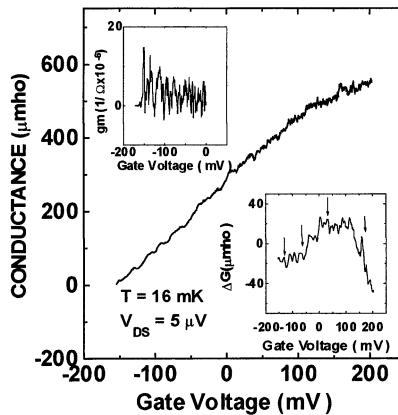


Figure 18.7. Conductance spectra versus the applied gate voltages. The insets are transconductance $g_m = (\partial I_{DS}/\partial V_G)$ when $V_{DS} = 5 \mu\text{V}$ and changes of the conductance versus the gate voltages.

Figure 18.7 shows conductance spectra as a function of the gate voltage at $T = 16 \text{ mK}$. The source-drain voltage was fixed at $5 \mu\text{V}$ to preserve the electron coherence. The pinch-off voltage ($V_{\text{pinch-off}}$) of the sample device is -155 mV . It is well known that the conductance in a conventional high electron mobility transistor which has a normal stripe gate is proportional to $V_G - V_{\text{pinch-off}}$. However, we observed oscillatory features which are superposed upon the linear conductance; these oscillations are clearly seen, for example at $V_G = -133 \text{ mV}$, with an intensity of $\sim 25\%$. Since these oscillatory behaviours are observed in the grating device and not in the conventional device, we believe that they are manifestations of electron diffraction due to the grating structure. Also, the transconductance versus the gate voltage for $V_{DS} = 5 \mu\text{V}$ is shown in the left inset. Multiple negative transconductance oscillations are clearly seen in the negative slope region of the curve of the source-drain current versus the gate voltage. The pure oscillatory conductances are derived from figure 18.7 by subtracting the linear fitting values from the measured conductances. Clear oscillations appear in the gate voltage range, which is shown in the inset. Major peaks are represented by bars, which can be attributed to the principle maxima in the diffraction spectra. Distances between the major peaks increase as the gate voltage increases from the pinch-off voltage, and three minor peaks are shown between the major peaks.

In figure 18.7, the conductance maxima represented by bars are seen to coincide with the peak positions expected from the diffraction condition. In order to verify that the observed oscillations are due to the electron diffraction effect, the Fraunhofer diffraction mode indices, $m = b/\sqrt{2}\lambda(V_G)$, where the slit separation b is 170 nm , are calculated from the peak positions in the gate voltage

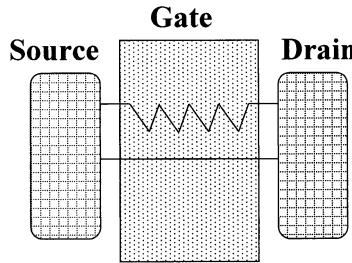


Figure 18.8. Schematic diagram of the corrugated quantum wire.

at $V_G = -133, -66, 30$, and 161 mV. The calculated index numbers for the oscillation peaks agree well with integers. We also observed three secondary maxima between the principal maxima which can be explained by Fraunhofer diffraction. We believe that the surface depletion at the side wall of the conducting channel reduces the actual number of slits in the sample devices to five.

18.4.4 One-dimensional corrugated quantum channel

In most quantum interference devices the channel regions were modified to control the electron phase [18, 19]. A corrugated quantum wire was considered to increase the yield of the interference effect. In order to investigate the above issues we fabricated a corrugated quantum wire covered with a metal gate and measured conductances as a function of the gate voltage [42]. The base material was AlGaAs/GaAs heterostructure, and fabrication procedures were similar to previous ones. A normal quantum wire structure was made for comparison. A schematic diagram of the corrugated quantum wire is shown in figure 18.8.

Figure 18.9 shows conductance spectra as a function of the gate voltage for both types of devices, $T = 16$ mK. The drain current was fixed at 1 nA to minimize the bias heating effects. The pinch-off voltages of the devices with and without corrugation are -0.21 V and -0.2 V, respectively. The conductance in the device without corrugation changes smoothly and plateau-like features are observed (the plateau-like features also exist in the device with corrugation). Since conductance steps are typically predicted in one-dimensional quantum wires, the results of figure 18.9 demonstrate that our samples are ballistic quasi-one-dimensional wires with several transverse modes. The conductance steps are originated from electrical depopulation of the conducting modes by the gate potential change. On the other hand, large oscillatory features, whose intensity is $\sim 60\%$ at -0.1 V gate voltage for example, are clearly seen in the device with the corrugation, and not in the normal quantum wire. Major peaks can be attributed to resonant maxima in the multiple transmission spectra, whose positions will be used in the calculation of the resonant transmission mode index. Distances between the major peaks are almost equal to ~ 80 mV in the negative gate voltage

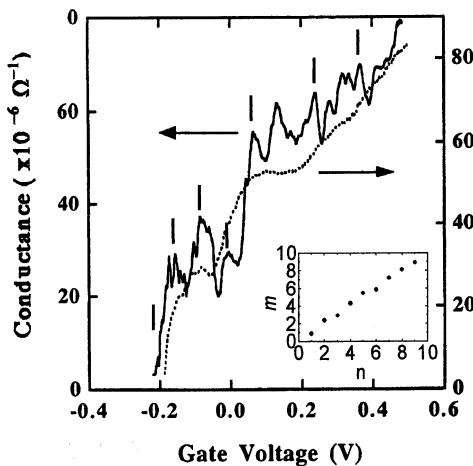


Figure 18.9. Conductance spectra versus the applied gate voltages; the line curve is for the corrugated wire, while the dotted curve is for the normal wire. The peak positions represented by bars were used in the calculation of the mode index.

range and ~ 120 mV in the positive gate voltage range, respectively. Three minor peaks also appear between the major peaks.

Since these oscillatory behaviours are observed in the corrugated quantum wire and not in the normal quantum wire, we believe that they are a manifestation of the resonance effect induced by the corrugated structure; the resonance effect is similar to the Fabry–Perot-type multiple transmission effect [30, 34].

In order to verify that the observed oscillations are due to the resonance effect, the resonance mode indices, $m = 2b/\lambda(V_g)$ (V_g is the gate voltage and b is the distance between single corrugations) are calculated from the peak positions in the gate voltage, and are plotted with integer numbers in the inset. The gate voltages alter the electron density in the device so that the Fermi wavelength is changed according to $\lambda_F = 4/n_1$ (n_1 is the one-dimensional electron density). The calculated index numbers for the resonance peaks agree with the integer number.

18.5 Summary

As the present electronics technologies are beginning to show their limitations in speed and capacity, the quantum technology is being given increased attention as an alternative way of doing telecommunications and information processing [44]. It is in this context that nanoscale semiconductor devices are being given increased attention, as they have novel electrical properties that can be utilized for such purposes [45]. Here we have focused on the low-dimensional quantum

structures and have reviewed the quantum interference effects. Quantum devices are already finding useful applications in the photonic technologies, for example, in lasers, modulators, switches and logic devices. Although much more work still needs to be done in the required electronic technologies, they are expected to play an important role in the future telecommunication and information processing technology.

Acknowledgments

The author would like to thank Professor J-C Woo (Seoul National University) for his encouragement. It is a pleasure to acknowledge Dr J Mondt (ETRI) for carefully reading the manuscript. This work has been generously supported by the Ministry of Information Communication, Korea.

References

- [1] Weisbuch C and Vinter B 1991 *Quantum Semiconductor Structures* (New York: Academic); many references can be found elsewhere
- [2] Reed M A and Kirk W P 1990 *Nanostructure Physics and Fabrication* (New York: Academic)
- Park K and Lee E-H 1998 *Physics of Semiconductor Devices* ed V Kumar and S K Agarwal (New Delhi: Narosa) p 47; many references can be found elsewhere
- [3] Alferov Z I 1997 *Compound Semiconductor 1996 (Inst. Phys. Conf. Ser. No 155)* p 1
- [4] Sharam B L and Purohit R K 1974 *Semiconductor Heterojunctions* (Oxford: Pergamon)
- [5] Craighead H G 1992 *Physics of Nanostructures* ed J H Davies and A R Long (London: Institute of Physics) p 21
- [6] Luryi S 1987 *Heterostructure Band Discontinuities: Physics and Device Application* ed F Capasso and G Margaritondo (Amsterdam: Elsevier) p 489
- [7] Ono M, Saito M, Yoshitomo T, Fiegna C, Ohguro T and Iwai H 1993 *IEDM Technical Digest* p 119
- [8] Taur Y, Buchanan D A, Chen W, Frank D J, Ismail K E, Lo S-H, Sai-Halasz G A, Viswanathan R G, Wann H C, Wind S J and Wong H-S 1997 *Proc. IEEE* **85** 486
- [9] Ferry D K 1993 *Extended Abstracts of the 1993 Int. Conf. on Solid State Devices and Materials* ed M Messe, p 312
- [10] Ando T and Fukuyama H 1992 *Transport Phenomena in Mesoscopic Systems* ed H Fukuyama and T Ando (Berlin: Springer) p 3
Washburn S 1988 *J. Res. Dev.* **32** 335 and many references therein
- [11] Timp G 1992 *Nanostructured Systems* ed M Reed (Boston, MA: Academic) p 113
- [12] See a text book of quantum physics; also Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485
- [13] Washburn S and Webb R A 1986 *Adv. Phys.* **35** 375
- [14] Datta S 1990 *Physics of Quantum Electron Devices* ed F Capasso (Berlin: Springer) p 321
- [15] Webb R A, Washburn S, Umbach C P and Laibowitz R B 1985 *Phys. Rev. Lett.* **54** 2696

- [16] Timp G, Chang A M, Cunningham J E, Chang T Y, Mankiewich P, Behringer R and Howard R E 1987 *Phys. Rev. Lett.* **58** 2814
- [17] Datta S, Melloch M R, Bandyopadhyay S and Lundstrom M S 1986 *Appl. Phys. Lett.* **48** 487
Tokura Y and Tsubaki K 1987 *Appl. Phys. Lett.* **51** 1807
- [18] Ismail K, Chu W, Antoniadis D A and Smith H I 1988 *Appl. Phys. Lett.* **52** 1071
Tsubaki K, Honda T, Saito H and Fukui T 1991 *Appl. Phys. Lett.* **58** 376
- [19] Allee D R, Chou S Y, Harris J S Jr and Pease R F W 1989 *J. Vac. Sci. Technol. B* **7** 2015
- [20] Washburn S, Schmid H, Kern D and Webb R A 1987 *Phys. Rev. Lett.* **59** 1791
- [21] de Vegvar P G N, Timp G, Mankiewich P M, Behringer R and Cunningham J 1989 *Phys. Rev. B* **40** 3491
- [22] Yacoby A, Sivan U, Umbach C P and Hong J M 1991 *Phys. Rev. Lett.* **66** 1938
- [23] Okuda M, Miyazawa S, Fujii K and Shimizu A 1993 *Phys. Rev. B* **47** 4103
- [24] Ford C J B, Fowler A B, Hong J M, Knoedler C M, Laux S E, Wainer J J and Washburn S 1990 *Surf. Sci.* **229** 307
- [25] Cahay M, Bandyopadhyay S and Grubin H L 1989 *Phys. Rev. B* **39** 12989
- [26] Buttiker M, Imry Y, Landauer R and Pinhas S 1985 *Phys. Rev. B* **31** 6207
- [27] Stone A D and Imry Y 1986 *Phys. Rev. Lett.* **56** 189
- [28] Ford C J B, Thornton T J, Newbury R, Pepper M, Ahmed H, Foxon C T, Harris J J and Roberts C 1988 *J. Phys. C: Solid State Phys.* **21** L325
- [29] Wu J C, Wybourne M N, Yindeepol W, Weisshaar A and Goodnick S M 1991 *Appl. Phys. Lett.* **59** 102
- [30] Park K, Lee S, Shin M, Lee E-H and Kwon H C 1995 *Phys. Rev. B* **51** 13805
- [31] Park K, Lee S, Shin M, Lee E-H and Kwon H C 1996 *Phys. Rev. B* **54** 1498
- [32] Leburton J-P and Lyanda-Geller Yu B 1996 *Phys. Rev. B* **54** 17716
- [33] van Wees B J, van Houten H, Beenakker C W J, Williamson J G, Kouwenhoven L P, van der Marel D and Foxon C T 1988 *Phys. Rev. Lett.* **60** 848
- [34] Kouwenhoven L P, Hekking F W J, van Wees B J, Harmans C J P M, Timmering C E and Foxon C T 1990 *Phys. Rev. Lett.* **65** 361
- [35] Leng M and Lent C S 1993 *Phys. Rev. Lett.* **71** 137
Lent C S and Leng M 1991 *Appl. Phys. Lett.* **58** 1650
- [36] Al'tschuler B L, Aronov A G and Spivak B Z 1981 *Pis. Zh. Eksp. Teor. Fiz.* **33** 101
- [37] Takai D and Ohta K 1993 *Phys. Rev. B* **48** 1537
- [38] Shin M, Park K, Lee S and Lee E-H 1995 *Japan. J. Appl. Phys.* **34** 4470
- [39] Aharonov Y 1984 *Proc. Int. Symp. on Foundations of Quantum Mechanics* ed S Kamefuchi *et al* (Tokyo: Physical Society of Japan) p 10
See, for example, Matteucci G and Pozzi G 1985 *Phys. Rev. Lett.* **54** 2469
- [40] Bandyopadhyay S, Bernstein G H and Porod W 1989 *Nanostructure Physics and Fabrication* ed M A Reed and W P Kirk (Boston, MA: Academic) p 183
- [41] Park K, Lee S, Shin M, Yuk J S, Lee E-H and Kwon H C 1997 *Appl. Phys. Lett.* **71** 3555
- [42] Park K, Lee S, Shin M, Yuk J S, Lee E-H and Kwon H C 1998 *Phys. Rev. B* **58** 3557
- [43] Park K, Lee S, Shin M, Lee E-H and Kwon H C 1995 *Solid State Commun.* **95** 717
- [44] Lee E-H 1993 *Physics of Semiconductor Device* ed K Lal (New Delhi: Narosa) p 100
- [45] Lee E-H and Park K 1997 *Compound Semiconductor 1996 (Inst. Phys. Conf. Ser. No. 155)* p 33

Chapter 19

Non-Markovian optical gain of strained-layer quantum-well lasers with many-body effects

Doyeol Ahn

*Institute of Quantum Information Processing and Systems,
University of Seoul, 90 Jeonnong, Tongdaemoon-ku, Seoul
130-743, Republic of Korea*

A non-Markovian model for the optical gain of strained-layer quantum-well lasers is studied taking into account the valence-band mixing, strain effects, many-body effects and the non-Markovian relaxation using the time-convolutionless reduced-density operator formalism for an arbitrary driven system coupled to a stochastic reservoir. Many-body effects are taken into account within the time-dependent Hartree–Fock approximation and the valence-band structure is calculated from the (6×6) Luttinger–Kohn Hamiltonian. The optical gain with Coulomb (or excitonic) enhancement is derived by integrating the equation of motion for the interband polarization. It is predicted that the Coulomb enhancement of gain is pronounced in the cases of compressive and unstrained quantum wells while it is negligible in the case of tensile strained quantum wells.

19.1 Introduction

Recently, there has been growing interest in the optical gain of semiconductor lasers with non-Markovian relaxation [1–7] partly because of the dissatisfaction with the conventional model for the optical gain in predicting the correct gain spectrum and the thermodynamic relation which dictates the transparency point in the gain spectra to coincide with the difference of the quasi-Fermi energies for the conduction and valence bonds. In addition, many-body effects such as the bandgap renormalization, enhancement of optical gain due to the attractive

electron–hole interaction (Coulomb or excitonic enhancement), and the plasma screening are also considered important in the description of the optical properties of semiconductors [8–12]. Most of the previous work on many-body effects on the optical gain assumed Markovian (or Lorentzian) line shape functions to describe the intraband relaxation and dephasing.

In this paper, a non-Markovian model for the optical gain of strained-layer quantum-well lasers is studied by extending the time-convolutionless reduced-density operator formalism developed in previous work [4, 13, 14] to include both non-parabolic valence-band structure and the effects of built-in strain potential due to the lattice mismatch between the quantum well and barrier regions. Then, we apply the model to calculate the optical gain spectra of strained-layer $\text{In}_x\text{Ga}_{1-x}\text{As-InP}$ quantum-well lasers. The valence-band structure is calculated within the (6×6) Luttinger–Kohn model which takes into account the coupling of the spin–orbit split-off band with the heavy- and light-hole bands [17–19]. The plasma screening of the Coulomb interaction is taken into account using an effective Hamiltonian in the time-dependent Hartree–Fock approximation [16].

19.2 The model

We consider a driven system of electron–hole pairs in the band-edge semiconductors surrounded by the stochastic reservoir. The microscopic behaviour of the entire system is governed by the stochastic quantum Liouville equation for the density operator $\rho_T(t)$ of the total system [6]

$$\frac{d\rho_T(t)}{dt} = -i[H_T(t), \rho_T(t)] = -iL_T(t)\rho_T(t), \quad (19.1)$$

where $L_T(t) = L_o(t) + L_i(t) + L_{\text{ext}}(t)$, is the Liouville super operator in one-to-one correspondence with the Hamiltonian $H_T(t) = H_o(t) + H_i(t) + H_{\text{ext}}(t)$ where $H_o(t)$ is the Hamiltonian of the system, $H_{\text{ext}}(t)$ the interaction of the system with the external driving field, and $H_i(t)$ the Hamiltonian for the interaction of the system with its stochastic reservoir. The physical quantities one observes are accompanied by fluctuations because the system is coupled to a heat bath or reservoir and its evolution has a stochastic element which is absent in the case of the Hamiltonian evolution of a closed system. In this work, our approach would include a phenomenology and we assume that the stationary part of the interaction can be put into an effective Hamiltonian $H_o(t)$ in the time-dependent Hartree–Fock approximation since we do not presently have the first-principle theory of how to separate the interaction into the stationary and stochastic parts. This line of thought links the well-known method of the stochastic Liouville equation to the stochastic theory of the spectral line shape. The information of the system is then contained in the reduced density operator obtained by eliminating the dynamical variables of the reservoir from the total density operator $\rho_T(t)$ using the projection operator technique (I). In a previous work [13], the equation of motion for the

reduced density operator $\rho(t)$ which is defined by $\rho(t) = \langle \underline{P} \rho_T(t) \rangle_i$, is derived and the generalized Bloch equations for the matrix elements of the reduced density operator are obtained employing the two-band model for semiconductors. Here $\langle \cdot \cdot \cdot \rangle_i$ denotes the average over the stochastic process $H_i(t)$. Many-body effects are included by taking into account the Coulomb interaction in the time-dependent Hartree–Fock approximation. In II, the optical gain was derived from the non-diagonal interband matrix element $p_k^*(t)$ that describes the interband pair amplitude induced by the optical field using a Padé approximation [10].

In this paper, we first show that the exact expression for the vertex function of the interband polarization can be obtained by simple mathematical manipulations. We start from the Laplace transformed equation of motion for the interband matrix element $\underline{p}_k^*(t)$. If the angular frequency of the optical field is ω , then the interband pair amplitude can be rewritten as

$$p_k^*(t) = e^{i\omega t} \underline{p}_k^*(t), \quad (19.2)$$

in the rotating wave approximation for electron–hole pairs in the band-edge semiconductors driven by the coherent light field. Then the linearized equation of motion for $\underline{p}_k^*(t)$ becomes [4]

$$\begin{aligned} \frac{\partial}{\partial t} p_k^*(t) &= i\Delta_k \underline{p}_k^*(t) + i \left[\mu^*(k) E_p^* + \sum_{k'} V_s(k - k') \underline{p}_{k'}^*(t) \right] [n_{ck}^o - n_{vk}^o] \\ &\quad - \int_0^t d\tau \{ \langle\langle v k | [H_i(t)(\underline{U}_o(\tau) H_i(t - \tau))] | v k \rangle\rangle_i \\ &\quad + \langle\langle c k | [(\underline{U}_o(\tau) H_i(t - \tau)) H_i(t)] | c k \rangle\rangle_i \} \underline{p}_k^*(t) \\ &\quad + i \int_0^t d\tau \int_0^\tau ds \exp\{-i[E_v^{\text{sc}}(k) - E_c^{\text{sc}}(k)]s\} \\ &\quad \times \{ \langle\langle v k | [H_i(t)(\underline{U}_o(\tau) H_i(t - \tau))] | v k \rangle\rangle_i \\ &\quad + \langle\langle c k | [(\underline{U}_o(\tau) H_i(t - \tau)) H_i(t)] | c k \rangle\rangle_i \} \mu^*(k) E_p * t - s \\ &\quad \times [n_{ck}^o - n_{vk}^o], \end{aligned} \quad (19.3)$$

where $\Delta_k = E_c^{\text{sc}}(k) - E_v^{\text{sc}}(k) - \omega$; n_{ck}^o and n_{vk}^o are the quasi-equilibrium distribution of electrons in the conduction band and the valence band, respectively; $V_s(k)$ is the screened Coulomb potential; $\underline{U}_o(t) = T \exp\{-i \int_0^t ds L_o(s)\}$ is the unperturbed evolution operator of the system with the time-ordering operator T ; $\mu(k)$ is the dipole moment, $E_c^{\text{sc}}(k)$ and $E_v^{\text{sc}}(k)$ are renormalized energies of electrons in the conduction and the valence band, respectively, and $E_p(t)$ is the optical field. We also employ the two-band model for the semiconductor and introduce two short-handed notations $|ck\rangle$ and $|vk\rangle$ such that

$$|ck\rangle = |c, \mathbf{k}\rangle \quad \text{and} \quad |vk\rangle = |v, \mathbf{k}\rangle,$$

where c and v denote conduction and valence bands, respectively, and \mathbf{k} is the electron wavevector. In equation (19.3), we assume that the electron–hole pairs

were initially in quasi-equilibrium. Otherwise, one needs to solve the coupled equations of motion for $n_{ck}(t)$, $n_{vk}(t)$ and $p_k^*(t)$ where $n_{ck}(t)$ and $n_{vk}(t)$ are non-equilibrium distributions for electrons in the conduction band and in the valence band, respectively. This would be too complicated and is out of the scope of this paper. We also assume that the interband pair amplitude follows the temporal variation of the polarization and the field amplitude adiabatically.

One can obtain $\underline{p}_k^*(t)$ in a closed form from equation (19.3) using the integrating factor:

$$\begin{aligned} \underline{p}_k^*(t) = & i \int_0^t d\tau \exp \left[- \int_\tau^t dt' (-i\Delta_k + g_1(t')) \right] \\ & \times \left[\mu^*(k) E_p^* \{1 + g_2(\tau, \Delta_k)\} + \sum_{k'} V_s(k - k') \underline{p}_{k'}^*(\tau) \right] [n_{ck}^o - n_{vk}^o], \end{aligned} \quad (19.4)$$

with

$$\begin{aligned} g_1(t) = & \int_0^t d\tau \{ \langle\langle vk | [H_i(t)(\underline{U}_o(\tau)H_i(t-\tau))] | vk \rangle\rangle_i \\ & + \langle\langle ck | [(\underline{U}_o(\tau)H_i(t-\tau))H_i(t)] | ck \rangle\rangle_i \} \end{aligned} \quad (19.5)$$

and

$$\begin{aligned} g_2(t, \Delta_k) = & \int_0^t d\tau \int_0^\tau ds \exp\{i\Delta_k s\} \{ \langle\langle vk | [H_i(t)(\underline{U}_o(\tau)H_i(t-\tau))] | vk \rangle\rangle_i \\ & + \langle\langle ck | [(\underline{U}_o(\tau)H_i(t-\tau))H_i(t)] | ck \rangle\rangle_i \}. \end{aligned} \quad (19.6)$$

In equations (19.4)–(19.6), $g_1(t)$ is the optical dephasing which is the temporal decay of the interband polarization. Recent studies on optical dephasing using the kinetic equations for the non-equilibrium Green's functions showed that the dephasing processes are time dependent and non-Markovian dephasing processes are expected to be a source for the non-exponential decay. It is convenient mathematically to manipulate equation (19.4) by taking the Laplace transformation.

We define the following Laplace transforms [4]:

$$\Psi_k(s) = \mathcal{L}\{\underline{p}_k^*(t)\}, \quad (19.7a)$$

$$\Xi(s, \Delta_k) = \mathcal{L}\left\{ \int_0^t dt' (i\Delta_k - g_1(t')) \right\}, \quad (19.7b)$$

and

$$G_2(s, \Delta_k) = \mathcal{L}\{g_2(t, \Delta_k)\} \quad (19.7c)$$

where $\mathcal{L}\{f(t)\}$ denotes the Laplace transformation of $f(t)$.

Then from equations (19.4) and (19.7), we obtain

$$\Psi_k(s) = \Psi_k^{(o)}(s) Q_k(s), \quad (19.8)$$

where

$$\Psi_k^{(o)}(s) = i\Xi(s, \Delta_k) \mu^*(k) E_p^* \left[\frac{1}{s} + G_2(s, \Delta_k) \right] [n_{ck}^o - n_{vk}^o], \quad (19.9)$$

and

$$Q_k(s) = 1 + \frac{s}{\mu^*(k) E_p^*(1 + sG_2(s, \Delta_k))} \sum_{k'} V_s(k - k') \Psi_{k'}^{(o)}(s) Q_{k'}(s). \quad (19.10)$$

In a previous work [4], we calculated the vertex function $Q_k(s)$ in the Padé approximation. In the following, it is shown that the vertex function can be obtained exactly by a simple manipulation. First, we make a summation of both sides of equation (19.10) over k :

$$\begin{aligned} \sum_k Q_k(s) &= \sum_{k'} Q_{k'}(s) \\ &= \sum_k 1 + \sum_{kk'} \frac{s}{\mu^*(k) E_p^*(1 + sG_2(s, \Delta_k))} V_s(k - k') \Psi_{k'}^{(o)}(s) Q_{k'}(s) \\ &= \sum_k 1 + \sum_{kk'} \frac{s}{\mu^*(k') E_p^*(1 + sG_2(s, \Delta_{k'}))} V_s(k' - k) \Psi_k^{(o)}(s) Q_k(s). \end{aligned} \quad (19.11)$$

Or

$$\sum_k \left[1 - \sum_{k'} \frac{s}{\mu^*(k') E_p^*(1 + sG_2(s, \Delta_{k'}))} V_s(k' - k) \Psi_k^{(o)}(s) \right] Q_k(s) = \sum_k 1. \quad (19.12)$$

In equation (19.12), the range of k is restricted to the first Brilloin zone. Then, it is obvious that

$$\begin{aligned} Q_k(s) &= \frac{1}{\left[1 - \sum_{k'} \frac{s}{\mu^*(k') E_p^*(1 + sG_2(s, \Delta_{k'}))} V_s(k' - k) \Psi_k^{(o)}(s) \right]} \\ &= \frac{1}{1 - q_k(s)}, \end{aligned} \quad (19.13)$$

where

$$\begin{aligned} q_k(s) &= \sum_{k'} \frac{s}{\mu^*(k') E_p^*(1 + sG_2(s, \Delta_{k'}))} V_s(k' - k) \Psi_k^{(o)}(s) \\ &= s \Psi_k^{(o)}(s) \sum_{k'} \frac{V_s(k' - k)}{\mu^*(k') E_p^*(1 + sG_2(s, \Delta_{k'}))}. \end{aligned} \quad (19.14)$$

Note that the expression for $q_k(s)$ is exact and is different from $q_{lk}(s)$ given in equation (16) of [4] which is the first-order term in the Padé approximation. The vertex function derived in this work which is based on the time-convolutionless reduced-density operator formalism (I and II) is equivalent to the one derived from the solution of the Bethe–Salpeter equation obtained from the many-body Green's function approach [8, 9] because the Green's function can be interpreted as the matrix elements of the reduced-density operator.

The steady-state interband pair amplitude is determined by taking the limit procedure of the Laplace transformation:

$$\begin{aligned} \underline{p}_k^*(0) &= \lim_{s \rightarrow 0} s \Psi_k(s) \\ &= \lim_{s \rightarrow 0} \frac{s \Psi_k^{(o)}(s)}{1 - q_k(s)} \\ &= i \frac{\Xi(0, \Delta_k)}{1 - q_k(0)} \mu^*(k) E_p^*[1 + g_2(\Delta_k)] [n_{ck}^o - n_{vk}^o], \end{aligned} \quad (19.15)$$

with

$$\begin{aligned} q_k(0) &= \lim_{s \rightarrow 0} s \Psi_k^{(o)}(s) \sum_{k'} \frac{V_s(k' - k)}{\mu^*(k') E_p^*(1 + s G_2(s, \Delta_{k'}))} \\ &= i \Xi(0, \Delta_k) \mu^*(k) E_p^*[1 + g_2(\Delta_k)] [n_{ck}^o - n_{vk}^o] \\ &\quad \times \sum_{k'} \frac{V_s(k' - k)}{\mu^*(k') E_p^*(1 + s G_2(s, \Delta_{k'}))} \\ &= i \Xi(0, \Delta_k) \mu^*(k) [n_{ck}^o - n_{vk}^o] \sum_{k'} \frac{V_s(k' - k)}{\mu^*(k')}. \end{aligned} \quad (19.16)$$

The interband polarization P and the susceptibility χ can be expressed through the dipole operator and the interband pair amplitude as

$$P = \frac{1}{V} \text{Tr}\{\mu(k) \underline{p}_k^*(0)\} \quad (19.17)$$

and

$$\varepsilon_0 \chi(\omega) = \frac{1}{V} \text{Tr} \left\{ i \frac{\Xi(0, \Delta_k)}{1 - q_k(0)} |\mu(k)|^2 [1 + g_2(\Delta_k)] [n_{ck}^o - n_{vk}^o] \right\}. \quad (19.18)$$

The linear gain of two-dimensional electron–hole pairs in a quantum well is given by

$$\begin{aligned} g(\omega) &= \frac{\omega \mu_c}{n_r} \frac{1}{V} \sum_{\sigma \eta lm} \sum_{\mathbf{k}_{\parallel}} |\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})|^2 \\ &\quad \times \frac{(f_c^l - f_{h\sigma}^m)[1 + \text{Re } g_2(\Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}))]}{(1 - \text{Re } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}))^2 + (\text{Im } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}))^2} \end{aligned}$$

$$\begin{aligned} & \times \{\text{Re } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})) (1 - \text{Re } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})) \\ & - \text{Im } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})) \text{Im } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})\}, \end{aligned} \quad (19.19)$$

where

$$\begin{aligned} \text{Re } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}) = & - \sum_{\mathbf{k}'_{\parallel}} \frac{V_{lmm'l}^{\text{eh}\sigma}(|\mathbf{k}'_{\parallel} - \mathbf{k}_{\parallel}|)}{|\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}'_{\parallel})|} |\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})| (f_c^l(\mathbf{k}_{\parallel}) - f_{\text{ho}}^m(\mathbf{k}_{\parallel})) \\ & \times \text{Im } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})), \end{aligned} \quad (19.20)$$

and

$$\begin{aligned} \text{Im } q_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}) = & \sum_{\mathbf{k}'_{\parallel}} \frac{V_{lmm'l}^{\text{eh}\sigma}(|\mathbf{k}'_{\parallel} - \mathbf{k}_{\parallel}|)}{|\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}'_{\parallel})|} |\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})| (f_c^l(\mathbf{k}_{\parallel}) - f_{\text{ho}}^m(\mathbf{k}_{\parallel})) \\ & \times \text{Re } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})), \end{aligned} \quad (19.21)$$

where $\Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}) = E_l^c(\mathbf{k}_{\parallel}) - E_m^{\text{ho}\sigma}(\mathbf{k}_{\parallel}) + E_G - \hbar\omega$; f_c^l and f_{ho}^m are the Fermi functions for the l th subband in the conduction band and the m th subband in the valence band of H^σ , respectively; $E_l^c(\mathbf{k}_{\parallel})$ and $E_m^{\text{ho}\sigma}(\mathbf{k}_{\parallel})$ are the l th subband energies in the conduction band and m th subband energy in the valence band of H^σ at \mathbf{k}_{\parallel} , respectively; E_G is the bandgap energy; ω is the angular frequency of photons; V is the volume; and $\mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})$ is the dipole matrix element between the l th conduction subband with a spin state η and the m th valence subband of the (3×3) block Hamiltonian H^σ .

We also have

$$\text{Re } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})) = \sqrt{\frac{\tau_c \pi}{2\gamma_{cv}(\mathbf{k}_{\parallel})}} \exp\left(-\frac{\tau_c \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel})^2}{2\gamma_{cv}(\mathbf{k}_{\parallel})}\right), \quad (19.22)$$

and $\text{Im } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_{\parallel}))$ is calculated numerically from the Kramers–Kronig relations. In this work, we assume that the correlation time τ_c and the interband relaxation $\gamma_{cv}(\mathbf{k}_{\parallel})$ are constants. Equation (19.19) is the generalization from the two-band case to the multiband case by simply adding all two-band contributions and in doing so neglecting the coupling of different polarizations. In order to include all the coupling terms, one has to rely on solving the equation of motion for the interband polarization numerically, which requires an enormous amount of computation time.

The interband screened-Coulomb matrix element $V_{lmm'l}^{\text{eh}\sigma}$ is approximated by [16]

$$\begin{aligned} V_{lmm'l}^{\text{eh}\sigma}(|\mathbf{k}'_{\parallel} - \mathbf{k}_{\parallel}|) & \sum_v \int dz_1 dz_2 \phi_l^*(z_1) g_m^{(v)*}(\mathbf{k}'_{\parallel}, z_2) \\ & \times \frac{e^2}{2\varepsilon_s(|\mathbf{k}'_{\parallel} - \mathbf{k}_{\parallel}| + \kappa)A} \exp(-|\mathbf{k}'_{\parallel} - \mathbf{k}_{\parallel}| |z_1 - z_2|) g_m^{(v)}(\mathbf{k}'_{\parallel}, z_2) \phi_l(z_1), \end{aligned} \quad (19.23)$$

where κ is the two-dimensional inverse screening length [20, 21]. Renormalization of the single-particle energies is approximated by the rigid band shift at $\mathbf{k}_\parallel = 0$ or the bandgap renormalization. Substituting equation (19.5) into equations (19.2) and (19.3), we obtain

$$\begin{aligned} \text{Re } q_{lm}^{\eta\sigma}(\mathbf{k}_\parallel) = & -|\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_\parallel)| \text{Im } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_\parallel)(f_c^l(\mathbf{k}_\parallel) - f_{h\sigma}^m(\mathbf{k}_\parallel)) \\ & \times \int \frac{d\mathbf{k}'_\parallel}{4\pi^2} I(\mathbf{k}_\parallel, \mathbf{k}'_\parallel), \end{aligned} \quad (19.24a)$$

and

$$\begin{aligned} \text{Im } q_{lm}^{\eta\sigma}(\mathbf{k}_\parallel) = & |\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}_\parallel)| \text{Re } \Xi_{lm}^{\eta\sigma}(0, \Delta_{lm}^{\eta\sigma}(\mathbf{k}_\parallel)(f_c^l(\mathbf{k}_\parallel) - f_{h\sigma}^m(\mathbf{k}_\parallel)) \\ & \times \int \frac{d\mathbf{k}'_\parallel}{4\pi^2} I(\mathbf{k}_\parallel, \mathbf{k}'_\parallel), \end{aligned} \quad (19.24b)$$

with

$$\begin{aligned} I(\mathbf{k}_\parallel, \mathbf{k}'_\parallel) = & \sum_v \int dz_1 dz_2 \phi_l^*(z_1) g_m^{(v)*}(\mathbf{k}_\parallel, z_2) \frac{e^2}{2\varepsilon_s(|\mathbf{k}'_\parallel - \mathbf{k}_\parallel| + \kappa)A} \\ & \times \exp(-|\mathbf{k}'_\parallel - \mathbf{k}_\parallel||z_1 - z_2|) g_m^{(v)}(\mathbf{k}_\parallel, z_2) \phi_l(z_1) \\ & \times \frac{1}{|\hat{\varepsilon} \cdot \mathbf{M}_{lm}^{\eta\sigma}(\mathbf{k}'_\parallel)|}. \end{aligned} \quad (19.24c)$$

Here $g_m^{(v)}$ and ϕ_l are the envelope functions for the valence and conduction bands, respectively.

19.3 Numerical results and discussion

To calculate the band structure and the electron (or hole) wavefunctions of a quantum well, the $\mathbf{k} \cdot \mathbf{p}$ method is employed. For the valence band, we use the (6×6) Pikus–Bir Hamiltonian [17–19] the elements of which are derived based on the $\mathbf{k} \cdot \mathbf{p}$ method, taking into account the spin–orbit (SO) split-off band coupling effects on band structure. As a numerical example, we have chosen a 60 \AA $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{InP}$ quantum well because it is widely used in high-speed optical-fibre communication systems. Moreover, it has a unique property such that an $\text{In}_x\text{Ga}_{1-x}\text{As}$ quantum well grown on InP substrates can be under a biaxial compressive ($x > 0.53$), no strain ($x = 0.53$), or a biaxial tensile strain ($x < 0.53$).

In figures 19.1(a)–(c), we plot the $\text{Re } q_{11}$ (full curve) and the $\text{Im } q_{11}$ (broken curve) which are defined by equations (19.2) and (19.3), respectively, for (a) $x = 0.65$ (compressive), (b) $x = 0.53$ (no strain) and (c) $x = 0.4$ (tensile), versus the in-plane wavevector (in units of $2\pi/a_0$). The carrier density of $3 \times 10^{18} \text{ cm}^{-3}$, intraband relaxation times of 0.1 ps and the correlation time of 20 fs are assumed

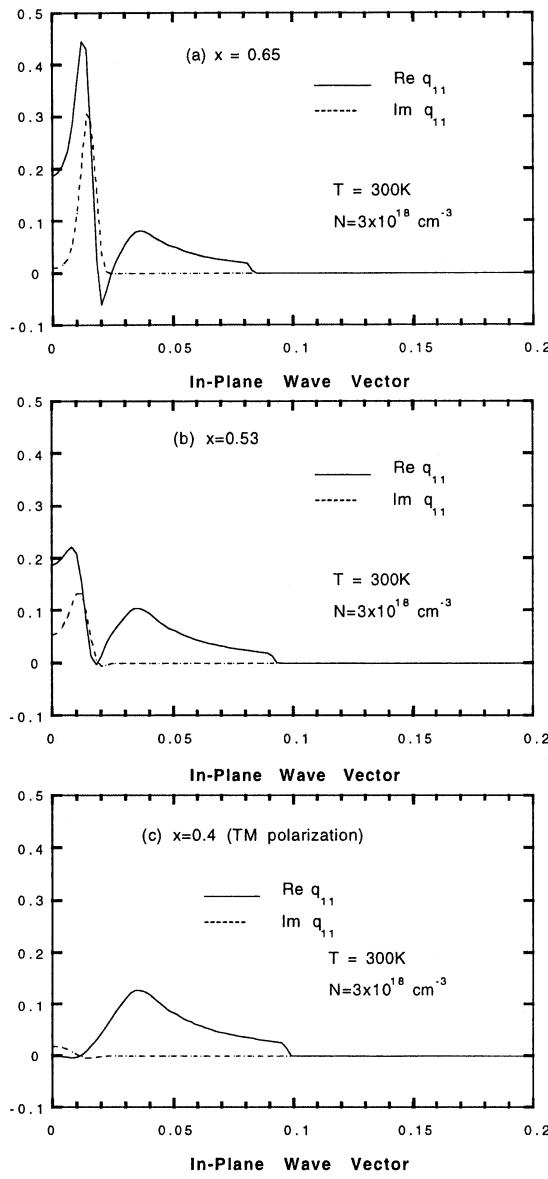


Figure 19.1. Plot of the $\text{Re } q_{11}$ (full curve) and the $\text{Im } q_{11}$ (broken curve) for (a) $x = 0.65$ (compressive), (b) $x = 0.53$ (no strain) and (c) $x = 0.4$ (tensile), versus the in-plane wavevector (in units of $2\pi/a_0$). The carrier density of $3 \times 10^{18} \text{ cm}^{-3}$, intraband relaxation time of 0.1 ps and the correlation time of 20 fs are assumed throughout the calculation.

throughout the calculation. In all three cases, $\text{Re } q_{11}$ is larger than $\text{Im } q_{11}$ in magnitude when compared as functions of the in-plane wavevector. However, they both vanish when the in-plane wavevector is far from the zone centre. In other words the Coulomb enhancement of gain is only appreciable near the zone centre. This makes sense because slower electron–hole pairs can feel the attractive Coulomb force more strongly than faster electron–hole pairs. If we ignore $\text{Im } q_{11}$, then the Coulomb enhancement factor is inversely proportional to $1 - \text{Re } q_{11}$ [15, 16]. From this figure, we expect that the Coulomb enhancement of optical gain would be appreciable for both compressive strained and unstrained quantum wells and be negligible for the tensile strained quantum well. It is predicted that the enhancement of an optical gain would be most pronounced in the case of compressive strained quantum wells. In the case of a tensile strained quantum well, we consider the Coulomb enhancement for the TM polarization because the TM mode gain is dominant over the TE case for tensile strained quantum wells.

In figures 19.2(a)–(c), non-Markovian optical gain spectra with (full curves) and without (broken curves) Coulomb enhancement are plotted for (a) $x = 0.65$ (compressive), (b) $x = 0.53$ (no strain) and (c) $x = 0.4$ (tensile), versus photon energy for carrier densities of $3 \times 10^{18} \text{ cm}^{-3}$ and $5 \times 10^{18} \text{ cm}^{-3}$. Bandgap renormalizations are taken into account in all cases. From these figures, we expect that the many-body effects are becoming more important in cases of compressive strained quantum wells and large carrier densities in the well. In previous papers [6, 15, 16], the author showed that the optical gain model with the non-Markovian lineshape function removes two anomalies which occur when the Lorentzian lineshape functions are used in the gain calculation. These anomalies are the unnatural absorption region below the bandgap energy and mismatch of the transparency point in the gain spectra with the Fermi level separation, the latter suggests that the carriers and the photons are not in thermal (or quasi-) equilibrium.

In the absence of spectral broadening the optical gain spectra are related to the spontaneous emission spectra from the detailed balance between absorption and emission of photons [22, 23]. One can easily see from this relation that there is a transparency point in the gain spectra which coincide with the Fermi (or quasi-Fermi) level separation that suggests the carriers and the photons are in thermal (or quasi-) equilibrium. In figures 19.2(a)–(c), it can be seen that the transparency points for a given carrier density converge regardless of the inclusion of the Coulomb enhancement effects, which results in the consistent picture with the thermodynamic relations. The reason why the use of non-Markovian lineshape function shows better agreement upon a transparency point with the quasi-Fermi level separation is the faster convergence of the Gaussian-like nature of the non-Markovian lineshape function as compared with the Lorentzian lineshape function in the spectral integration for the gain.

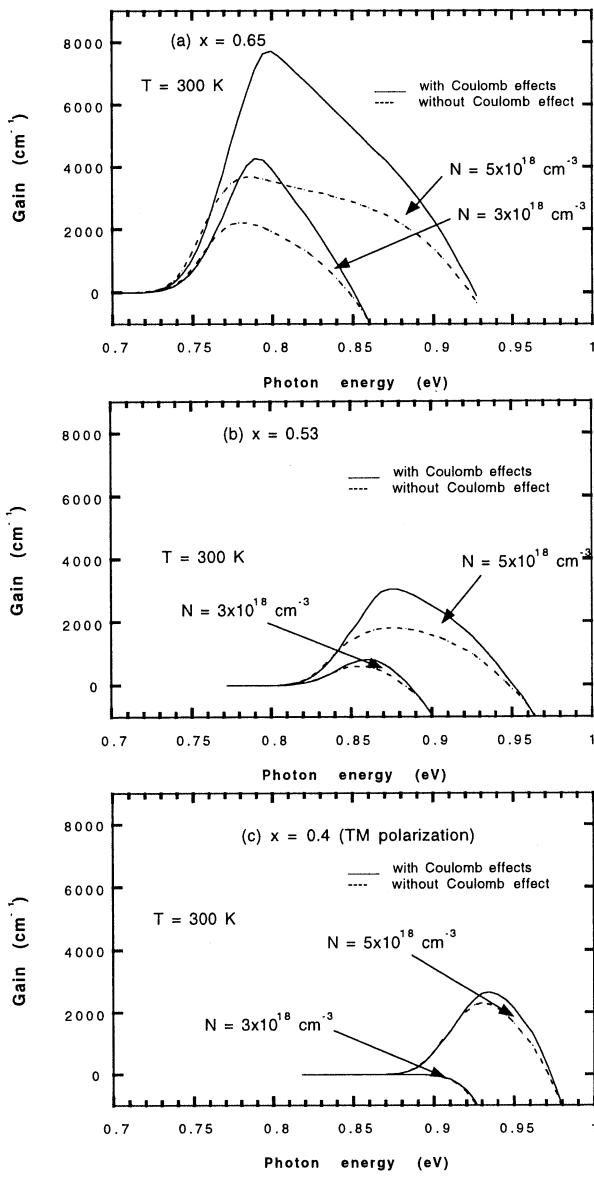


Figure 19.2. Non-Markovian optical gain spectra with (full curves) and without (broken curves) Coulomb enhancement are plotted for (a) $x = 0.65$ (compressive), (b) $x = 0.53$ (no strain) and (c) $x = 0.4$ (tensile), versus photon energy for carrier densities of $3 \times 10^{18} \text{ cm}^{-3}$ and $5 \times 10^{18} \text{ cm}^{-3}$.

19.4 Conclusions

In summary, a comprehensive model for the description of optical gain spectra of strained-layer quantum-well lasers is developed taking into account the valence-band mixing, strain effects, many-body effects and the non-Markovian relaxation. Recently developed time-convolutionless quantum-kinetic equations for electron-hole pairs near the band-edge are used to derive the optical gain and the line shape function of a quantum-well laser. Many-body effects are taken into account within the time-dependent Hartree-Fock approximation and the valence-band structure with built-in strain potential is calculated from the (6×6) Pikus-Bir Hamiltonian. The optical gain with Coulomb (or excitonic) enhancement is derived by integrating the equation of motion for the interband polarization. As a numerical example, an $\text{In}_x\text{Ga}_{1-x}\text{As-InP}$ quantum well is chosen for its wide application in optical communication systems. It is predicted that the Coulomb enhancement of gain is pronounced in the cases of compressive and unstrained quantum wells while it is negligible in the case of a tensile strained quantum well.

Acknowledgment

This work was supported by the Creative Research Initiatives at the Ministry of Science and Technology.

References

- [1] Yamanishi M and Lee Y 1987 Phase dampings of optical dipole moments and gain spectra in semiconductor lasers *IEEE J. Quantum Electron.* **QE-23** 367–70
- [2] Ohtoshi T and Yamanishi M 1991 Optical line shape functions in quantum-well and quantum-wire structures *IEEE J. Quantum Electron.* **27** 46–53
- [3] Tomita A and Suzuki A 1991 Optical line shape functions in quantum-well and quantum-wire structures *IEEE J. Quantum Electron.* **27** 1630–41
- [4] Ahn D 1995 The theory of non-Markovian gain in semiconductor lasers *IEEE J. Selected Topics Quantum Electron.* **1** 301–7
- [5] Tomita A 1996 Many-body effects, space-charge potentials, and valence-band mixing on the optical gain in quantum-well structures *Phys. Rev. B* **54** 5609–19
- [6] Ahn D 1997 Theory of non-Markovian optical gain in quantum-well lasers *Prog. Quantum Electron.* **21** 249–87
Ahn D 1998 *IEEE J. Quantum Electron.* **34** 344–52
- [7] Enders P M 1997 Enhancement and spectral shift of optical gain in semiconductors from non-Markovian intraband relaxation *IEEE J. Quantum Electron.* **33** 580–8
- [8] Haug H and Schmitt-Rink S 1984 Electron theory of the optical properties of laser-excited semiconductors *Prog. Quantum Electron.* **9** 3–100
- [9] Schmitt-Rink S, Ell C and Haug H 1986 Many-body effects in the absorption, gain, and luminescence spectra of semiconductor quantum-well structures *Phys. Rev. B* **33** 1183–9

- [10] Haug H and Koch S W 1989 Semiconductor laser theory with many-body effects *Phys. Rev. A* **39** 1887–98
- [11] Pereira M F Jr, Koch S W and Chow W W 1991 Many-body effects in the gain spectra of strained quantum wells *Appl. Phys. Lett.* **59** 2941–3
- [12] Chow W W and Koch S W 1995 Many-body Coulomb effects in room-temperature II–VI quantum well semiconductor lasers *Appl. Phys. Lett.* **66** 3004–6
- [13] Ahn D 1994 Time-convolutionless reduced-density-operator theory of an arbitrary driven system coupled to a stochastic reservoir: Quantum kinetic equations for semiconductors *Phys. Rev. B* **50** 8310–18, hereafter called I
- [14] Ahn D 1995 Time-convolutionless reduced-density-operator theory of an arbitrary driven system coupled to a stochastic reservoir. II. Optical gain and line-shape function of a driven semiconductor *Phys. Rev. B* **51** 2159–66, hereafter called II
- [15] Ahn D 1996 Optical gain of a quantum-well laser with non-Markovian relaxation and many-body effects *IEEE J. Quantum Electron.* **32** 960–5
- [16] Ahn D 1996 Non-Markovian gain of a quantum-well laser with many-body effects *Appl. Phys. Lett.* **69** 2498–500
- [17] Chao C Y P and Chuang S L 1992 Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells *Phys. Rev. B* **46** 4110–22
- [18] Chang S S and Chuang S L 1995 Modelling of strained quantum-well lasers with spin-orbit coupling *IEEE J. Selected Topics Quantum Electron.* **1** 218–29
- [19] Ahn D, Yoon S J, Chuang S L and Chang C S 1995 Theory of optical gain in strained-layer quantum wells within the 6×6 Luttinger–Kohn model *J. Appl. Phys.* **78** 2489–97
- [20] Haug H and Koch S W 1993 *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (Singapore: World Scientific)
- [21] Chow W W, Koch S W and Sargent M III 1994 *Semiconductor-Laser Physics* (Berlin: Springer)
- [22] Rees P and Blood P 1995 Derivation of gain spectra of laser diodes from spontaneous emission measurements *IEEE J. Quantum Electron.* **31** 1047–50
- [23] Henry C H, Logan R A and Merritt F R 1980 Measurement of gain and absorption spectra in AlGaAs buried heterostructure lasers *J. Appl. Phys.* **51** 3042–50

Chapter 20

Generation of non-classical lights from semiconductor light emitters

Masamichi Yamanishi, Hiroyuki Sumitomo, Masahide Kobayashi and Yutaka Kadoya

Department of Physical Electronics, Faculty of Engineering, and Hiroshima-Branch of Research Projects for Core Research for Evolutional Science and Technology (CREST), JST, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima 739-8527, Japan

Results on two types of experiments related to sub-Poissonian photon state are demonstrated; (a) the current-dependence of the squeezing bandwidth of a light-emitting-diode (LED) driven by a constant-current source at room temperature and (b) the generation of sub-Poissonian photons from a LED without recourse to high-impedance circuits at room temperature. In the interpretation of both squeezing characteristics, the significance of backward-pump process is emphasized.

20.1 Introduction

The performance of light communication systems and the sensitivity of various measurements has already been limited by the standard quantum limit (SQL) determined by Heisenberg's minimum uncertainty principle. Hence, non-classical lights, such as photon-number squeezed states and quadrature-phase squeezed states are required to circumvent such difficulty associated with the SQL. A photon-number squeezed state, having photon-number uncertainty, Δn smaller than that of coherent lights (the square root of the average number, $\sqrt{n_0}$ of Poissonian distribution) is termed a sub-Poissonian photon state. Among a variety of schemes for non-classical light generation, the scheme based on semiconductor

light emitters has been attracting much attention, because of the simplicity of the experimental arrangement, the low energy consumption, and the possibility of a large degree of noise suppression. In fact, sub-Poissonian photon-state has been successfully generated using semiconductor laser diodes [1, 2] or light-emitting diodes (LEDs) [3–8] driven by high-impedance constant-current sources, and series coupled LEDs [9, 10] driven by constant-voltage sources. In cases of practical application, such as device-uses in few photon direct-detection optical communication systems, the LED- (spontaneous emission-) mode operation is of crucial importance since LEDs are expected to have higher quantum efficiencies at very low current regime, and as a result, weak sub-Poissonian fluxes could be generated with LEDs even at pump rates far below the threshold for the onset of lasing [11].

In this paper, we will review our own experimental results on photon-number squeezing both in a constant-current and constant-voltage driven heterojunction LED. In the interpretation of obtained squeezing characteristics, the significance of backward-pump process will be emphasized. In the case of the constant-current operation (section 20.2), the experimental results of squeezing bandwidth at room temperature can be explained in a unified manner over a whole range of injection current in terms of our new model including the effect of the microscopic backward-pump process. On the other hand, in the constant-voltage operation (section 20.3), we demonstrate our experimental results, manifesting substantial noise suppression of output photon fluxes below the SQL level from the single LED surprisingly without recourse to high-impedance circuit at room temperature. The observed squeezing is also interpreted by the backward-pump model.

20.2 Squeezing in photon-number fluctuations under constant-current operation

20.2.1 Basic physics for constant-current operation

The aim of this subsection is to review theoretical models proposed so far for the physical interpretation of sub-Poissonian photon generation in LEDs driven by high-impedance constant-current sources. The current source-noise described by Johnson–Nyquist noise in a resistor, R_s , can be settled to be much weaker than the full shot-noise of a current flow, i.e. $4k_B T / R_s \ll 2eI_{\text{LED}}$ if R_s is chosen to be sufficiently high, where k_B , T , e and I_{LED} are the Boltzmann constant, the device temperature, the elementary charge and the dc pump-current of an LED, respectively. However, ‘quiet electron flows’ created by high-impedance constant-current sources do not necessarily guarantee the generation of ‘quiet photon flows’ [1, 12]. [Figure 20.1](#) shows the energy-band diagram of a hetero p–n junction LED. The conduction electrons are supplied to a p-doped active region from an n-doped wide-bandgap region across the potential barrier by thermionic emission, diffusion or tunnelling processes. Such an electron transport process is

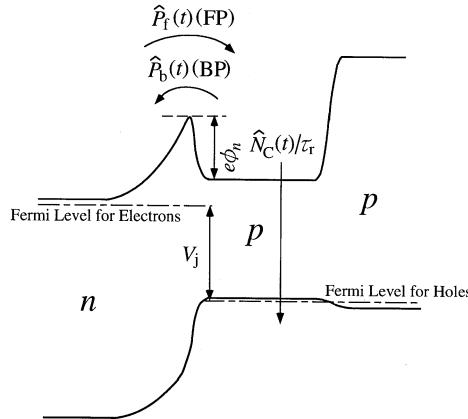


Figure 20.1. Typical energy-band diagram under the forward-bias condition of a commercial p-n heterojunction light-emitting-diode (LED).

a completely random point (stochastic) one in microscopic level as the transport events are caused by the collision of electrons with phonons. Hence, the extra noise would be added to the carrier injection process. Fortunately, however, the macroscopic pump rate is regulated even in the conventional macro-size LEDs or LDs through a negative feedback mechanism owing to the junction voltage fluctuations induced by the resultant effects of many pump events, if the time-window (measurement time) Δt is much wider than the thermionic emission time $\tau_{te} = k_B T C_{dep}/(e I_{LED})$, where C_{dep} is the junction capacitance [1, 4, 13]. In other words, the regulation bandwidth for the pump process is given by $1/(2\pi\tau_{te})$. This regulation mechanism is sometimes termed ‘macroscopic Coulomb blockade’ or ‘collective Coulomb blockade’ [4, 13, 14]. It is interesting to note that the inequality to represent the source-noise suppression, $4k_B T/R_s \ll 2eI_{LED}$, is equivalent to the condition, namely $R_s \gg$ differential resistance of LED, $r_d \cong k_B T/(e I_{LED})$, under which the junction voltage can fluctuate freely and the pump-current is kept constant despite existence of the junction voltage fluctuations. In addition to the normal injection of electrons (FP) considered above, the reverse current flow, i.e. the backward transport of conduction electrons from the p-type active to the n-type wide-bandgap regions (BP) may take place at microscopic level. The (net) average current is given by the difference between the FP and BP rates. In the noise analysis, each process has to be treated separately. The way to treat the BP process has not yet been well established and is a key issue of the present work.

Another random point process is involved in the radiative recombination process of injected carriers. Spontaneous decay of the excited electrons has a shot-noise character because vacuum-field fluctuations (VFFs) trigger spontaneous emissions. However, the free fluctuations of the electron population N_C in

the active region induced by the random emission events brings us an internal negative feedback for the subsequent emission events. Namely, the fluctuations of the photon emission rate associated with the population fluctuations cancel the shot-noise originally involved in the photon emission process, as far as the measurement time window is wider than the recombination lifetime of the carriers, τ_r , the time required for an appreciable change in the electron population. Thus, the bandwidth for the regulation of emission process due to the fluctuations of electron population is given by $1/(2\pi\tau_r)$.

It is naturally anticipated that the squeezing bandwidth of photon-number fluctuations is characterized by both the characteristic time constants of the pump and recombination dynamics. In fact, applying the macroscopic diffusion model (FP rate \cong BP rate) to a homojunction LED, Kim and Yamamoto [14] have obtained an expression for the squeezing bandwidth (3-dB rolloff bandwidth) f_c as

$$f_c = \frac{1}{2\pi(\tau_{te} + \tau_r)} = \frac{1}{2\pi \left(\frac{k_B T C_{dep}}{e I_{LED}} + \tau_r \right)}. \quad (20.1)$$

Some experimental results have been fitted rather well by equation (20.1) [4, 7]. However, it is questionable whether the macroscopic diffusion model is applicable generally to heterojunction LEDs, since the BP rate is to strongly depend on bandgap discontinuities, temperatures and injection current densities in the heterostructure devices. Indeed, our experimental results cannot be interpreted in terms of the theoretical curve estimated with equation (20.1) as will be shown later. On the other hand, Imamoğlu and Yamamoto [13] focused on another extreme case where carrier injection into the active region takes place only through the thermionic (forward) emission process in a heterojunction LED. Under the assumption that the injected carriers recombine instantaneously, the squeezing bandwidth has been given by $f_c = e I_{LED} / (2\pi k_B T C_{dep})$. However, such an assumption may not again be valid for our heterojunction LED where the BP process may take place more or less particularly at room temperature.

20.2.2 Experimental arrangement and results (constant-current operation)

A schematic diagram of the experimental arrangement for room temperature measurements is shown in figure 20.2. A high-speed AlGaAs LED (Hitachi HE8812SG), whose emitting surface is shaped hemispherically to give rise to a high external quantum efficiency [15], has a double heterojunction (DH) structure, n-Al_{0.2}Ga_{0.8}As/0.5 μm thick p-Al_{0.01}Ga_{0.99}As/p-Al_{0.2}Ga_{0.8}As and an emission spectrum, $\lambda \sim 860$ nm, peaking in the high-sensitivity wavelength range of a high quantum efficiency silicon p-i-n photodiode (PD) (Hamamatsu S6040). The LED was connected to a voltage source (a dry cell battery) through a metal-film resistor $R_s = 2$ KΩ, much higher than the differential junction resistance r_{dj} of the LED ($R_s \gg r_{dj}$), i.e. the LED was driven by a high-impedance constant-current source. For the measurement of sub-Poissonian photon states, the LED

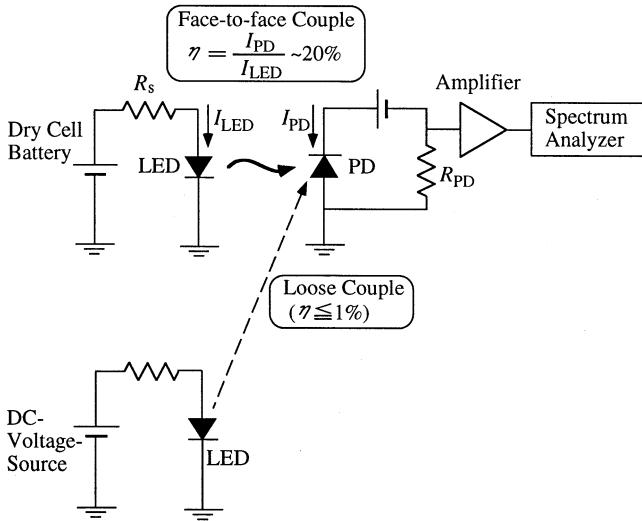


Figure 20.2. Schematic diagram of the experimental arrangement under constant-current operation. For the measurement of the sub-Poissonian photon state, the light-emitting-diode (LED) was coupled tightly to the photodiode (PD). Conversely, in order to calibrate the full shot-noise (FSN) level, another LED was coupled weakly to the PD.

was placed just in front of the PD to yield a high current-to-current conversion efficiency (photodetector current, I_{PD} /LED-pump-current, I_{LED}), $\eta \sim 20\%$ at room temperature [8]. The output current flow in the PD was converted to a voltage signal through a load resistor ($R_{PD} = 330 \Omega$). Then, the noise signal was amplified by an ultralow noise preamplifier (NF SA-220F5) and was finally fed to a spectrum analyser (HP HP4396A). The high-frequency characteristic of the detection system was primarily limited by that of the RC low-pass circuit formed by the capacitance of the PD, the input capacitance of the preamplifier and the load resistor.

In order to calibrate experimentally the full shot-noise (FSN) level (reference level), the PD was illuminated by another LED (HE8812SG) located far away from the PD to reduce the coupling efficiency, $\eta \leq 1\%$ [8]. The shot-noise level can be precisely determined by this technique within a very small measurement error of $\sim 0.2\%$. The whole experimental system was set inside a shield room to isolate it from external noise sources. Normalized noise power spectral density of the output photon flux, i.e. the Fano factor was determined experimentally as the ratio between the sub-Poissonian noise power and the shot-noise power.

Figure 20.3 shows the spectral Fano factor of detected photons, $F_d(f)$ for a LED-pump-current of 13.18 mA and a quantum efficiency of $\eta = 18.16\%$, demonstrating a substantial noise suppression below the FSN level over a

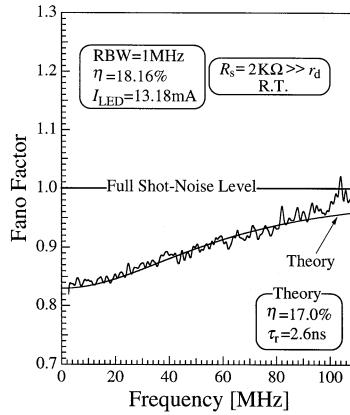


Figure 20.3. Spectral Fano factor of output photon fluxes over a wide frequency range obtained with a high-speed light-emitting-diode (LED) (Hitachi HE8812SG) driven by a constant-current (13.18 mA) source at room temperature. The data were obtained with a resolution bandwidth of 1 MHz, a video bandwidth of 3 KHz and a sweep time of 5.12 s. For a comparison with the experimental result, the theoretical curve estimated with equation (20.2) is also indicated. The values of overall quantum efficiency and recombination lifetime are used, in the estimation, to be $\eta = 0.17$ and $\tau_r = 2.6$ ns, respectively.

wide frequency range from near dc to 100 MHz [8]. In this relatively high current regime, thermionic response time of the p–n junction is much shorter than recombination lifetime in the active layer, $\tau_r = 2.6$ ns $\gg \tau_{te} = k_B T C_{dep}/(e I_{LED}) = 0.14$ ns with $C_{dep} = 70$ pF, and, in turn, the squeezing bandwidth is primarily limited by the recombination dynamics. By fitting of the experimental result with the theoretical curve for such a case [16]

$$F_d(f) = 1 - \eta + \eta \frac{(2\pi f \tau_r)^2}{1 + (2\pi f \tau_r)^2}, \quad (20.2)$$

we obtain $\eta = 0.17$, which is reasonably close to the experimental value obtained with dc measurement ($\eta = 0.1816$) and $\tau_r = 2.6$ ns, also close to the widely accepted value, $\tau_r \sim 3$ ns [17], of electron lifetime in a highly p-doped ($\sim 10^{18}$ cm $^{-3}$) GaAs which is the active layer material (though more strictly, Al_{0.01}Ga_{0.99}As) of the present LED. The short recombination lifetime of electrons enabled us to observe the wide-band squeezing of photon-number fluctuations. The bandwidth of the noise suppression, f_c , defined as the frequency at which the degree of noise reduction is half the maximum suppression at the low-frequency limit, is observed to be ~ 60 MHz in this specific case, which is the widest ever reported with LEDs. This bandwidth for the noise suppression is very close to the 3 dB bandwidth of optical output for external current modulation of the same

type LED as our's, reported in [15].

In the next subsection, the pump-current-dependence of squeezing bandwidth will be demonstrated to investigate the squeezing mechanism of photon-number fluctuations in the heterojunction LED.

20.2.3 Pump-current-dependence of squeezing bandwidth

As a result of the careful measurements carried out with the high-performance LED, we obtained systematic experimental results on current-dependence of the squeezing bandwidth, which may allow us qualitative discussion on physical mechanisms which limit the squeezing bandwidths at various injection current levels. The measured squeezing bandwidths (squares) at room temperature are plotted as a function of pump-current I_{LED} (or corresponding current density) in figure 20.4. The full curve shows the theoretical prediction estimated with equation (20.1). In the estimation, the junction capacitance, $C_{\text{dep}} = 70 \text{ pF}$ and the reasonable recombination lifetime, $\tau_r = 2.6 \text{ ns}$ were chosen independently so as to result in the best fittings between theory and experiments in the low- and high- I_{LED} limits, respectively. The postulated value, $C_{\text{dep}} = 70 \text{ pF}$ was reasonably close to the measured diode capacitance 50 pF at an applied forward voltage of 1.3 V which is slightly lower than the bias voltages used in the noise measurements, $V = 1.35\text{--}1.5 \text{ V}$. Nevertheless, the experimental values obviously deviate from the theoretical ones in the mid I_{LED} range, $I_{\text{LED}} = 0.6\text{--}10 \text{ mA}$ indicating that equation (20.1) cannot be used to explain the experimental results over the whole range of I_{LED} . In other words, the macroscopic diffusion model is not applicable to the present heterojunction LED unlike a homojunction LED [14]. In the next section, we will propose a new model which is going to interpret our experimental results over the entire current range.

20.2.4 Framework of backward-pump model

Let us go back to the energy-band diagram of a heterojunction LED under a forward-bias condition where the junction voltage is biased to be V_j as shown in figure 20.1. The LED (HE8812SG) used in the present work would be viewed electrically as a single p-n junction device since the active and p-type wide-bandgap regions of the LED are highly p-doped [15], $N_a > 1 \times 10^{18} \text{ cm}^{-3}$, much higher than the doping concentration in the n-type wide-bandgap region, $N_d \sim 1 \times 10^{17} \text{ cm}^{-3}$. In other words, the voltage drops across the p-type active region and the p-p heterojunction between the active and p-type wide-bandgap regions are negligibly low compared with the voltage drop across the p-n junction. The device performance of such a LED may be characterized by electron transport from the n-type wide-bandgap to p-type active regions and by recombination dynamics of electrons in the active region. Another important point to be considered is backward electron transport which is termed backward-pump (BP) process, from the active to n-type wide-bandgap regions. In general,

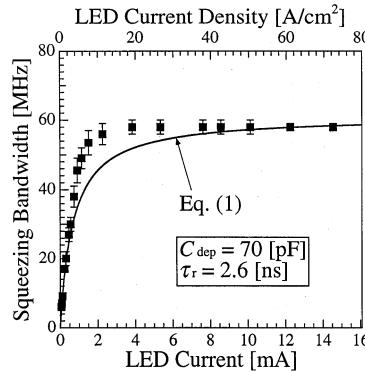


Figure 20.4. Measured pump-current(-density)-dependence of the squeezing bandwidth of the high-speed light-emitting-diode (LED) driven by constant-current sources at room temperature. The experimental results are compared with the theoretical curve estimated with equation (20.1), using $C_{\text{dep}} = 70 \text{ pF}$ and $\tau_r = 2.6 \text{ ns}$.

the excess energy of an electron thermally injected from the n-type region into the p-type active region is relaxed quickly by its interaction with phonons, other electrons and holes. However, for a high pump-current, some injected electrons may no longer be completely relaxed down to around the conduction band edge because of the hot carrier effect [18] and/or band filling. As a result, electrons with energies higher than the n-side barrier height, $e\phi_n$ of the heterojunction can go back to the n-type layer and thus the BP process may take place as shown in figure 20.1. Thus, the LED should be viewed as a quantum mechanical system including such BP process.

We start by putting the operators, written in Heisenberg picture, representing the electron FP rate, $\hat{P}_f(t)$, the electron BP rate, $\hat{P}_b(t)$, the emission (electron recombination) rate, $\hat{N}_C(t)/\tau_r$, in the active region with the electron population, $\hat{N}_C(t)$ and the photon flux, $\hat{\Phi}(t)$ detected by a PD with an efficiency η , as the sums of their steady state (average) and small fluctuation terms. The corresponding Langevin noise source operators, $\hat{\Gamma}_{pf}(t)$, $\hat{\Gamma}_{pb}(t)$ and $\hat{\Gamma}_{sp}(t)$ are added to the FP, BP and emission rates operators, respectively, and $\hat{F}(t)$ is involved in the photon flux operator, to represent a complete set of time-evolution of these quantities,

$$\hat{P}_f(t) + \hat{\Gamma}_{pf}(t) = P_{f0} + \Delta \hat{P}_f(t) + \hat{\Gamma}_{pf}(t), \quad (20.3)$$

$$\hat{P}_b(t) + \hat{\Gamma}_{pb}(t) = P_{b0} + \Delta \hat{P}_b(t) + \hat{\Gamma}_{pb}(t), \quad (20.4)$$

$$\frac{\hat{N}_C(t)}{\tau_r} + \hat{\Gamma}_{sp}(t) = \frac{N_{C0}}{\tau_r} + \frac{\Delta \hat{N}_C(t)}{\tau_r} + \hat{\Gamma}_{sp}(t), \quad (20.5)$$

$$\hat{\Phi}(t) = \Phi_0 + \Delta \hat{\Phi}(t) = \eta \frac{N_{C0}}{\tau_r} + \eta \frac{\Delta \hat{N}_C(t)}{\tau_r} + \hat{F}(t). \quad (20.6)$$

We define a parameter α_0 ($0 \leq \alpha_0 < 1$) as the ratio of dc BP to FP rates,

$$\alpha_0 = P_{b0}/P_{f0} = I_{b0}/I_{f0}, \quad (20.7)$$

where I_{f0} and I_{b0} are the average microscopic forward- and backward-currents, respectively. The net average current is given by $I_{\text{LED}} (\equiv I_0) = I_{f0} - I_{b0} = e(P_{f0} - P_{b0})$, and, hence, the dc values (average values) of the FP and BP rates are expressed as

$$P_{f0} = \frac{P_0}{1 - \alpha_0}, \quad (20.8)$$

$$P_{b0} = \frac{\alpha_0}{1 - \alpha_0} P_0 = \frac{\alpha_0}{1 - \alpha_0} \frac{N_{C0}}{\tau_r}. \quad (20.9)$$

The small-amplitude fluctuation $\Delta \hat{P}_b(t)$ of the BP rate can be naturally viewed as being proportional to the fluctuation of the electron population $\Delta \hat{N}_C(t)$ in the active region and being independent of the junction voltage fluctuations, $\Delta \hat{V}_j(t)$. Hence, we put

$$\Delta \hat{P}_b(t) = \frac{\alpha_d}{1 - \alpha_d} \frac{\Delta \hat{N}_C(t)}{\tau_r}, \quad (20.10)$$

where α_d ($0 \leq \alpha_d < 1$) would be a function of N_{C0} . Similarly, for a small shift of the dc bias point, one obtains, by using the same parameter α_d ,

$$\Delta P_{b0} = \frac{\alpha_d}{1 - \alpha_d} \frac{\Delta N_{C0}}{\tau_r}. \quad (20.11)$$

The parameter, α_d may be understood as a differential ratio of the BP to FP rates in the dc limit,

$$\alpha_d = \frac{dP_{b0}}{dP_{f0}} = \frac{dI_{b0}}{dI_{f0}} \quad (20.12)$$

because of the continuity of particle flow,

$$\Delta P_{f0} - \Delta P_{b0} = \frac{\Delta N_{C0}}{\tau_r}. \quad (20.13)$$

Energy relaxation times, τ_e , for injected electrons with excess energies are short, < 100 ps enough to accomplish quickly the stationary Fermi–Dirac distribution of an electron temperature, T_e [18]. Hence, the parameters, α_0 and α_d may be written as functions of the electron population N_{C0} , electron temperature T_e and n-side barrier height $e\phi_n$ of the active region shown in figure 20.1. Such short energy relaxation times may justify the independence of the differential ratio, α_d on the frequency over the entire range, $f \lesssim 100$ MHz in the present work, allowing us the uses of the same parameter, α_d in equations (20.10) and (20.11). Note, however, that the differential ratio α_d is, in general, not necessarily equal to the dc ratio α_0 . We can determine, independently of the noise measurement, the dependence of α_0 and α_d on pump-current I_{LED} , using measured current-versus-voltage (I – V) characteristic and differential resistance of the LED used in the noise measurement [19].

20.2.5 Theoretical bandwidth for noise suppression in comparison with experimental results

The explicit expression for the spectral Fano factor of photon-number detected by a PD in the case of high-impedance noise suppression $R_s \gg r_{dj}$ can be obtained by solving quantum mechanical Langevin equations, which is given by [19]

$$\begin{aligned} F_d(f) = 1 - \eta & \left[1 + \frac{2(1 - \alpha_d)}{1 - \alpha_0} (\alpha_d - \alpha_0) (2\pi f \tau_{te})^2 \right] \\ & \times [1 + 2\alpha_d (2\pi f \tau_r) (2\pi f \tau_{te}) + (1 - \alpha_d)^2 (2\pi f \tau_r)^2 (2\pi f \tau_{te})^2 \\ & + (2\pi f \tau_r)^2 + (2\pi f \tau_{te})^2]^{-1}, \end{aligned} \quad (20.14)$$

where f is the frequency, and $\tau_{te} = C_{dep}r_{dj} = (C_{dep}nk_B T/eI_{LED}) \cdot (1 - \alpha_0(V_{j0})/(1 - \alpha_d(V_{j0})))$ is the junction-response-time in the condition that the BP processes exist. The ideality factor n is slightly larger than unity, $1 \leq n \lesssim 1.2$, in the LEDs used in the present experiments. The deviation of n from unity may not only be due to recombination processes in the depletion layer, but due to carrier transport through graded-gap heterojunctions [19].

Maximum squeezing depth, which is obtained in the condition that $f \rightarrow 0$, keeps constant, $F_d = 1 - \eta$ without relating to the existence of the BP processes. Equation (20.14) leads to the squeezing bandwidth (3-dB rolloff bandwidth) f_c ,

$$\begin{aligned} (2\pi f_c)^2 = & \frac{1}{2(1 - \alpha_d)^2 \tau_r^2 \tau_{te}^2} \\ & \times \left[- \left\{ \tau_r^2 + \tau_{te}^2 + 2\alpha_d \tau_r \tau_{te} - \frac{4(1 - \alpha_d)}{1 - \alpha_0} (\alpha_d - \alpha_0) \tau_{te}^2 \right\} \right. \\ & + \left(\left\{ \tau_r^2 + \tau_{te}^2 + 2\alpha_d \tau_r \tau_{te} - \frac{4(1 - \alpha_d)}{1 - \alpha_0} (\alpha_d - \alpha_0) \tau_{te}^2 \right\}^2 \right. \\ & \left. \left. + 4(1 - \alpha_d)^2 \tau_r^2 \tau_{te}^2 \right)^{1/2} \right]. \end{aligned} \quad (20.15)$$

In the case of $\alpha_0 = \alpha_d = 0$, equation (20.14) reduces to a form described by the product of two Lorentzians,

$$F_d(f) = 1 - \eta \frac{1}{\{1 + (2\pi f \tau_{te})^2\}\{1 + (2\pi f \tau_r)^2\}}, \quad (20.16)$$

and equation (20.15) reduces to

$$(2\pi f_c)^2 = \frac{1}{2\tau_r^2 \tau_{te}^2} \left\{ -(\tau_r^2 + \tau_{te}^2) + \sqrt{\tau_r^4 + \tau_{te}^4 + 6\tau_r^2 \tau_{te}^2} \right\}. \quad (20.17)$$

On the other hand, in another extreme case of $\alpha_0 = \alpha_d \rightarrow 1$, equation (20.14) reduces to a simpler form described by a single Lorentzian,

$$F_d(f) = 1 - \eta \frac{1}{1 + \{2\pi f(\tau_{te} + \tau_r)\}^2}, \quad (20.18)$$

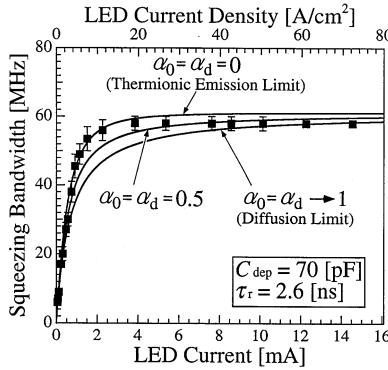


Figure 20.5. Pump-current(-density)-dependence of the observed squeezing bandwidth compared with the theoretical curves including the effects of the backward-pump (BP) process. The theoretical curves for three different values of $\alpha_0 = \alpha_d = 0, 0.5$ and 1 estimated with equation (20.15) by assuming $\tau_{te} = C_{dep}k_B T/(eI_{LED})$ are indicated. The values of junction capacitance and recombination lifetime used for the estimation are the same as in figure 20.4.

and equation (20.15) leads to

$$f_c = \frac{1}{2\pi(\tau_{te} + \tau_r)}, \quad (20.19)$$

which is actually equation (20.1). In the former case, $\alpha_0 = \alpha_d = 0$, the FP and recombination processes in a diode can be viewed as being two cascade processes since no BP process exists at all. This case may be termed the thermionic emission limit. In contrast, in the latter case, $\alpha_0 = \alpha_d \rightarrow 1$, both the pump and recombination processes are completely linked so that they are inseparable from each other. This is the so-called macroscopic diffusion limit, discussed in [14].

Let us presume that the BP rate, P_b is a linear function of the electron population, N_C , i.e. $\alpha_d = \alpha_0 = \text{constant}$. In figure 20.5, the experimental result of the squeezing bandwidth, f_c is compared with three theoretical curves estimated with equation (20.15) by putting $\tau_{te} = C_{dep}k_B T/(eI_{LED})$ and by using the different constant values, 0, 0.5 and 1, of α_0 and α_d and the same values of C_{dep} and τ_r as in figure 20.4. With the increasing pump-current I_{LED} , the experimental f_c begins to deviate from the theoretical curve labelled $\alpha_0 = \alpha_d = 0$ (equation (20.17)) at $I_{LED} = 4$ mA, and eventually approaches the curve labelled $\alpha_0 = \alpha_d \rightarrow 1$ (equation (20.19)), suggesting that both α_0 and α_d are no longer constant but increasing functions of I_{LED} . In other words, the pump situation of our LED moves from the thermionic emission toward diffusion limits as I_{LED} increases.

Here, we briefly discuss possible current-dependence of the parameters, α_0

and α_d in a heterojunction diode, in comparison with a homojunction diode. In a homojunction diode, injected electrons may easily go back causing $\alpha_0 = \alpha_d \rightarrow 1$, as discussed in [14]. On the other hand, in a heterojunction diode as in the present diode, the BP process for injected electrons may be less significant than in the homojunction diode since the n-side potential barrier $e\phi_n$ formed by bandgap discontinuity of the heterojunction may block the backward transport of injected electrons. However, the BP rate may be superlinearly increased by bandfilling and/or the hot carrier effect as I_{LED} increases, as already pointed out in the early part of this subsection. The superlinear increase in the BP rate yields increasing α_0 and α_d with I_{LED} , which will be, indeed, demonstrated elsewhere [19].

To summarize, we have investigated experimentally and theoretically the bandwidth for the squeezing of photon-number fluctuations in constant-current-driven heterojunction LEDs operating at room temperature, focusing on the driving-current-dependence of the bandwidth. The experiments using high-performance (high-speed, highly efficient) AlGaAs LEDs and carried out with careful measurements have led to the observation of apparent squeezing up to 100 MHz and allowed us a precise quantitative comparison with the theory.

In low- and high-injection current regimes, the squeezing bandwidths have been explained as limited by the response time of the p–n junction and the recombination lifetime in the active region, respectively, as already pointed out in previous work. However, it has been found in the present work that, in the intermediate current region, previous theory including the FP and BP processes in diffusion limit, originally developed for homojunction LEDs, fails to interpret the experimental results. We have proposed a theoretical model, generally applicable to heterojunction LEDs, where the dc and ac ratios of the BP to FP rates are included as parameters α_0 and α_d . It has been revealed that changes in the pump situation are included in a quite natural manner in the expression for the squeezing bandwidth, and as a consequence, that the transition of the situation from thermionic emission to diffusion limits associated with the changes of α_0 and α_d appears as the change of the bandwidth. The α_0 - and α_d -values can be determined by the measurements of current-versus-voltage characteristics and the differential resistance of the LED. The uses of these obtained α_0 - and α_d -values lead to an excellent interpretation of the experimental results of injection-current-dependence of the squeezing bandwidth over the whole range of the injection current used in the noise measurements. This will be demonstrated in a separate article [19].

20.3 Squeezing in photon-number fluctuations due to the backward pump process

20.3.1 Basic physics for constant-voltage operation

Sub-Poissonian photon states have been generated with constant-current-driven LED's, as demonstrated in the last section. The fundamental mechanisms

responsible for the generation of sub-Poissonian photon states with such semiconductor devices have relied on Coulomb repulsive force between electrons. The regulation of the electron stream by the Coulombic effect was converted into a photon stream. The most typical way to generate sub-Poissonian photon state using the Coulombic effect may be macroscopic Coulomb blockade where the charge dipoles at the p-n junction arisen by the pump processes suppress the subsequent pump events. However, if the junction is driven directly by a voltage source, the charge dipoles (system fluctuation) dissipate instantaneously to the voltage source (reservoir). Therefore, the resistor inserted between the junction and voltage source is essentially required to prevent the quick dissipation (constant-current operation as long as the regulation mechanism relies on the Coulombic effect). On the other hand, our new scheme for generation of sub-Poissonian photon states which will be proposed in this section is based on the backward pump process of electrons. The fundamental issues, which characterize the backward pump (BP) process are rises of temperature and the quasi-Fermi level of an electron system in the active region accompanied with an increase of the injection current. These phenomena have no relation with the Coulombic effect, so that if the BP process brings about regulation of pump events, generation of sub-Poissonian photon states is possible even in the constant-voltage operation where the junction is driven directly by a voltage source. The generation of sub-Poissonian photon states under the constant-voltage operation is of high interest from the application point of view because it is very hard to realize the generation of such states under the constant-current operation in extremely low current (≤ 16 nA), high-frequency regime operations [11]. In such regimes, the extremely high resistance ($R_s \geq 1$ M Ω) required for the constant-current operation would be unfortunately shunted by the vacuum impedance or a stray capacitance.

In the following subsections (subsections 20.3.2 and 20.3.3), we will demonstrate our experimental results, manifesting substantial noise suppression of the output photon fluxes below standard quantum limit (SQL) level from a single LED surprisingly without recourse to high-impedance circuit at room temperature. The observed squeezing will be interpreted in terms of our new model including the effects of microscopic backward pump process.

20.3.2 Experimental arrangements and results (constant-voltage operation)

A schematic of the experimental arrangement is shown in [figure 20.6](#). The arrangement is almost the same as the constant-current operation case except for one point that the test LED (HE8812SG) was intentionally shunted by the large capacitance ($C = 0.5$ μ F) to keep the LED bias voltage constant. The lead wires of the LED and the capacitance were tightly soldered to metal plates to reduce the wiring inductance as much as possible. We measured photon number noise of the output photon flux both at room and low temperatures.

Normalized noise power spectra, i.e. Fano factors obtained with three

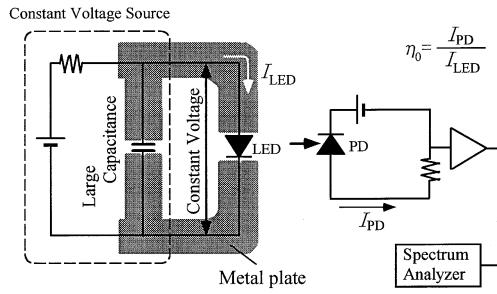


Figure 20.6. Experimental arrangement for noise measurements of the LED under constant-voltage operation. The terminals of the LED were intentionally shunted by the large capacitance to keep the LED voltage constant.

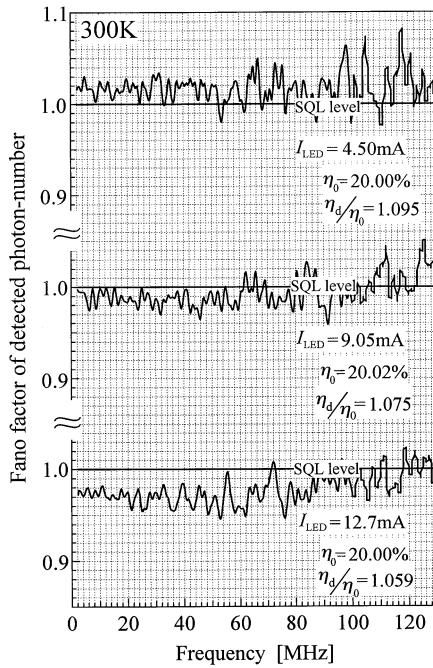


Figure 20.7. Measured spectra of Fano factor in constant-voltage operation at room temperature. The data were obtained by the spectrum analyser of the setting equal to the case in [figure 20.3](#).

different drive currents at room temperature are shown in figure 20.7. Despite the constant voltage source operation, the LED was found to generate sub-Poissonian photon fluxes in the high-current regime ($I_{\text{LED}} \geq 7 \text{ mA}$). The Fano factor of

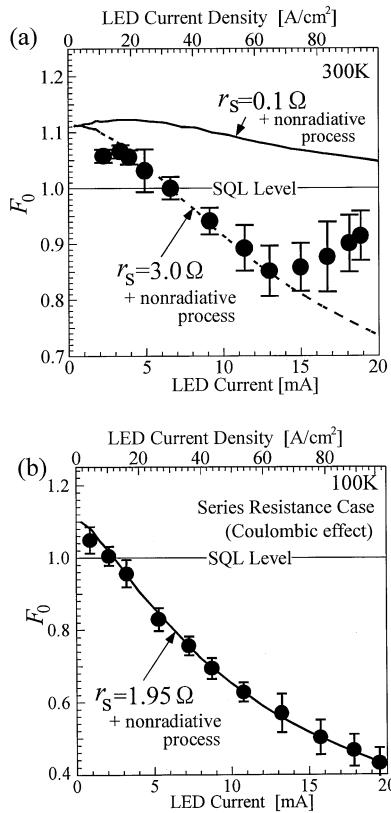


Figure 20.8. (a) The measured pump-current(-density)-dependence of the Fano factor, F_0 for the overall output photons in the low-frequency limit at room temperature (full circles). The full curve shows the theoretical values using the Coulomb blockade model estimated with $r_s = 0.1 \Omega$. Unrealistic values of $r_s = 3.0 \Omega$ have to be assumed to fit the experimental result (broken curve) with the theoretical one (4–13 mA). (b) Measured pump-current(-density)-dependence of the F_0 in the low-frequency limit at low temperature, 100 K. The theoretical curve (full curve) is estimated with a measured series resistance $r_s = 1.95 \Omega$ and the condition, $\alpha_0 = \alpha_d = 0$ at low temperature.

detected photon number, F_d , may be related to that of overall emitted photon-number, F_0 , i.e. $F_d = 1 - \eta_r + \eta_r F_0$ where η_r is the detection efficiency for all output photon emitted from LED [20]. The room- and low-temperature data of the Fano factor F_0 obtained from the experimental values of F_d in the low-frequency limit (5–10 MHz) are plotted in figures 20.8(a) and (b), respectively.

In order to discuss the mechanism of the squeezing, we estimated the squeezing degree at the low-frequency limit on the basis of the usual Coulomb

blockade model ($\alpha_0 = \alpha_d = 0$), given by [20];

$$F_0 = 1 - \left\{ \frac{\eta_r}{\eta_d} \left(1 + \frac{r_s}{r_{dj0}} \right)^2 \left(2 - \frac{\eta_d}{\eta_0} \right) - \frac{\eta_r}{\eta_0} - 2 \frac{\eta_r}{\eta_0} \cdot \frac{r_s}{r_{dj0}} \right\} \\ \times \left\{ \frac{\eta_r^2}{\eta_d^2} \left(1 + \frac{r_s}{r_{dj0}} \right)^2 \right\}^{-1} \quad (20.20)$$

where, $r_{dj0} = nk_B T/eI_{LED}$ is the differential junction resistance of the LED, η_0 the overall quantum efficiency and η_d the differential quantum efficiency, and the effect of the non-radiative process is considered through three kinds of efficiencies, η_r , η_0 and η_d .

First, let us focus on the low-temperature case. We obtained the theoretical result shown by the full curve in figure 20.8(b) computed with equation (20.20) using measured internal series resistance involved in the LED, $r_s = 1.95 \Omega$ and measured efficiencies. The theoretical result is in very good agreement with the experimental one, so that one can convincingly conclude that the observed squeezing comes out from the Coulomb blockade effect. In other words the squeezing can be viewed as arising from noise suppression by the series resistance r_s .

Next, let us consider the room-temperature case. In general, the semiconductor resistance is substantially lowered with temperature rise. The series resistance, r_s mainly originating from the p-doped layer of the present LED [19]. The temperature dependence of hole concentration and mobility in p-type $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with various Al mole fraction x and doping level have been extensively investigated in [21]. According to their experimental result [21], the series resistance, r_s , of the present LED at room temperature is evaluated to be lower than 0.2Ω , which is much lower than r_{dj} in our experimental condition ($I_{LED} \leq 20 \text{ mA}$). That is, the p-n junction in the LED can be viewed as being driven directly by a truly constant-voltage source. The full curve in figure 20.8(a) is the theoretical one estimated with equation (20.20) using $r_s = 0.1 \Omega$, which, however, badly deviate from the experimental results. An unrealistic value of the series resistance, $r_s = 3.0 \Omega$ must be assumed to fit the theoretical result with the experimental one (4–13 mA) as shown by the broken curve in figure 20.8(a). Nevertheless, the theoretical result still cannot explain the high-current behaviour of F_0 . It should be emphasized that the increasing tendency of F_0 in the high-current regime can never be interpreted by the usual Coulomb blockade model. Thus, we have to look for another mechanism which can work even in a constant-voltage operation. We have found that the mechanism is based on the BP process as will be manifested in the next subsection.

20.3.3 New squeezing mechanism based on the backward pump process

The theoretical F_0 at the low-frequency limit in the case of the constant-voltage operation on the basis of the BP model can be obtained by solving quantum

mechanical Langevin equations, which is given by [20];

$$\begin{aligned}
 F_0 = & 1 - \left\{ 2(1 - \alpha_d) \left[\left\{ \frac{\eta_r}{\eta_d} (1 - \alpha_d) \left(1 + \frac{r_s}{r_{dj,eff}(1 - \alpha_{d,eff})} \right) + \alpha_d \right\} \right. \right. \\
 & \times \left(1 + \frac{r_s}{r_{dj,eff}(1 - \alpha_{d,eff})} \right) \left. \right] \\
 & + \frac{\eta_r}{\eta_0} \left[(1 - \alpha_d)^2 \frac{1 + \alpha_{0,eff}}{1 - \alpha_{0,eff}} + \left(1 + \frac{r_s}{r_{dj,eff}(1 - \alpha_{d,eff})} \right)^2 (1 - \alpha_d)^2 \right. \\
 & \left. \left. + 2 \frac{\eta_r}{\eta_0} \frac{r_s / r_{dj,eff}(1 - \alpha_d)^2}{(1 - \alpha_{d,eff})(1 - \alpha_{0,eff})} \right\} \right. \\
 & \times \left\{ \frac{\eta_r}{\eta_d} (1 - \alpha_d) \left(1 + \frac{r_s}{r_{dj,eff}(1 - \alpha_{d,eff})} \right) + \alpha_d \right\}^{-2} \quad (20.21)
 \end{aligned}$$

where, $\alpha_{0,eff} = \alpha_0/[1 + (\eta_r/\eta_0 - 1)(1 - \alpha_0)]$, $\alpha_{d,eff} = \alpha_d/[1 + (\eta_r/\eta_d - 1)(1 - \alpha_d)]$ and $r_{dj,eff} = (nk_B T/e I_{LED}) \cdot (1 - \alpha_{0,eff})/(1 - \alpha_{d,eff})$ is the differential junction resistance of the LED on condition that the BP processes exist. In the deduction of equation (20.21), we also took account of the non-radiative process in addition to the BP process. If no BP processes exist (for instance, in the low-temperature case), equation (20.21) reduces to equation (20.20) with $\alpha_0 = \alpha_d \rightarrow 0$.

In the absence of non-radiative processes, three kinds of efficiencies become the same, $\eta_0 = \eta_d = \eta_r$, and, as a result $\alpha_{0,eff} = \alpha_0$ and $\alpha_{d,eff} = \alpha_d$, and eventually, F_0 can be simply expressed by,

$$F_0 = 1 - \left\{ (r_s/r_{dj})^2 + 2 \left(\frac{\alpha_d - \alpha_0}{1 - \alpha_0} \right) [(1 - \alpha_d) + r_s/r_{dj}] \right\} / (1 + r_s/r_{dj})^2. \quad (20.22)$$

In the room-temperature case, $r_s \ll r_{dj}$ for our experimental conditions as we have already pointed out. Hence, the term r_s/r_{dj} in equation (20.22) may be disregarded, resulting in

$$F_0 = 1 - 2 \frac{\alpha_d - \alpha_0}{1 - \alpha_0} (1 - \alpha_d). \quad (20.23)$$

Thus, equation (20.23) indicates that nonlinearity involved in the BP process, i.e. $\alpha_d > \alpha_0$ brings us squeezing ($F_0 < 1$), even if there is no series resistance at all. In fact, α_d is possibly larger than α_0 , which will be demonstrated by the measurement of the current-versus-voltage (I - V) characteristic and differential resistance of the LED used in the noise measurement [19]. In the case of the constant-voltage operation, maximum squeezing depth (represented by F_0) is related to the existence of the BP process as mentioned above unlike the constant-current operation case. A surprising point is that the noise reduction below the shot-noise level can occur despite the existence of two stochastic random processes (FP and BP processes). This is because of the correlation between

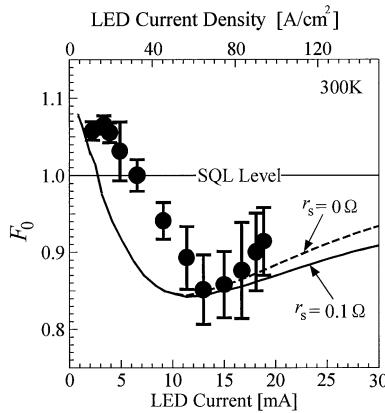


Figure 20.9. Theoretical result of the pump-current(-density)-dependence of F_0 in the low-frequency limit at room temperature on the basis of the BP model using $r_s = 0.1 \Omega$ (full curve) and $r_s = 0 \Omega$ (broken curve). The experimental results were well reproduced by the theoretical ones.

two kinds of pump rates. The nonlinear correlation between FP and BP processes (resulting from the relation $\alpha_d > \alpha_0$) is at the heart of the pump regulation. This will be demonstrated in detail in a separate paper [20].

The experimental results of F_0 can be compared with our theory, equation (20.21) by using measured η and α . Figure 20.9 shows the experimental result of the pump-current-dependence of F_0 at room temperature in comparison with the theoretical values estimated by equation (20.21), using $r_s = 0.1 \Omega$ and $r_s = 0 \Omega$. The experimental results were well reproduced by the theoretical ones. It is possible to interpret the increasing tendency of F_0 in the high-current region as a result of the decrease in the difference of α_0 and α_d with increasing current because both α_0 and α_d approach unity in the high-current region. In the low-current regime, output photon fluxes show super-Poissonian character ($F_0 > 1$) because of the non-radiative process [22, 23]. It can be concluded that the behaviour of F_0 is determined by non-radiative processes in the low-current regime, while it is done by the BP process in the high-current regime.

The mechanism for photon-number squeezing due to the BP process at room temperature is completely different from those in previous theory, in which sub-Poissonian statistics for external current is presupposed with recourse to the high-impedance circuit.

20.4 Conclusions

We conclude that the BP process plays essentially important roles in photon-number squeezing in heterojunction LEDs. In the case of the constant-current

operation, the pump-current-dependence of the squeezing bandwidth manifests the transition of carrier transport mechanism across the p–n junction, from the thermionic emission to diffusion limits. On the other hand, in the constant-voltage operation, the BP process brings about the new scheme to generate a sub-Poissonian light, without the need for a high-impedance constant-current source. The physical mechanisms behind the squeezing and the current-dependence of BP parameters, α_0 and α_d will be extensively discussed elsewhere [24]. The present experiments carried out with commercially available LEDs under standard operating condition have indicated the general importance of the BP process in heterojunction LEDs.

References

- [1] Yamamoto Y, Machida S and Nilsson O 1986 *Phys. Rev. A* **34** 4025
Yamamoto Y and Machida S 1987 *Phys. Rev.* **35** 5114
- [2] Machida S and Yamamoto Y 1988 *Phys. Rev. Lett.* **60** 792
Richardson W H, Machida S and Yamamoto Y 1991 *Phys. Rev. Lett.* **66** 2867
- [3] Tapster P R, Rarity J G and Satchell J S 1987 *Europhys. Lett.* **4** 293
- [4] Kim J, Kan H and Yamamoto Y 1995 *Phys. Rev. B* **52** 2008
- [5] Hirano T and Kuga T 1995 *IEEE J. Quantum Electron.* **QE-31** 2236
- [6] Shinozaki G, Hirano T, Kuga T and Yamanishi M 1997 *Japan. J. Appl. Phys.* **36** 6350
- [7] Abe J, Shinozaki G, Hirano T, Kuga T and Yamanishi M 1997 *J. Opt. Soc. Am. B* **14** 1295
- [8] Kobayashi M, Kohno M, Kadoya Y, Yamanishi M, Abe J and Hirano T 1998 *Appl. Phys. Lett.* **72** 284
- [9] Edwards P J and Pollard G H 1992 *Phys. Rev. Lett.* **69** 1757
- [10] Sumitomo H, Kadoya Y and Yamanishi M 1999 *Opt. Lett.* **24** 40
- [11] Yamanishi M, Watanabe K, Jikutani N and Ueda M 1996 *Phys. Rev. Lett.* **76** 3432
- [12] Yamamoto Y and Haus H A 1992 *Phys. Rev. A* **45** 6596
- [13] Imamoğlu A and Yamamoto Y 1993 *Phys. Rev. Lett.* **70** 3327
- [14] Kim J and Yamamoto Y 1997 *Phys. Rev. B* **55** 9949
- [15] Kurata K, Ono Y, Ito K, Mori M and Sano H 1981 *IEEE Trans. Electron Dev.* **ED-28** 374
- [16] Yamanishi M and Lee Y 1993 *Phys. Rev. A* **48** 2534
- [17] For instance, Casey H C Jr and Panish M B 1978 *Heterostructure Lasers* part A (New York: Academic) ch 3, p 161
- [18] For instance, Shah J 1978 *Solid-State Electron.* **21** 43
- [19] Kobayashi M, Yamanishi M, Sumitomo H and Kadoya Y *Phys. Rev. B* submitted
- [20] Sumitomo H, Yamanishi M and Kadoya Y unpublished
- [21] Zukotynski S, Sumski S, Panish M B and Casey H C Jr 1979 *J. Appl. Phys.* **50** 5795
- [22] Fujisaki H and Shimizu A 1997 *J. Phys. Soc. Japan* **66** 34
- [23] Fujisaki H and Shimizu A 1998 *Phys. Rev. A* **57** 3074
- [24] Sumitomo H, Yamanishi M and Kadoya Y unpublished