PHYSICA E

LOW-DIMENSIONAL SYSTEMS & NANOSTRUCTURES

Proceedings of the Tenth
International Winterschool on New
Developments in Solid State Physics

"New Frontiers in Low-Dimensional
Physics"

held in Mauterndorf, Austria
23–27 February 1998

Guest Editors:

G. Bauer
H. Heinrich
F. Kuchar

19990106 023

NORTH-HOLLAND

http://www.elsevier.nl/locate/physe

PHYSICA E

An interdisciplinary journal of research on low-dimensional systems and nanostructures

Aims and scope

Physica E (Low-dimensional systems and nanostructures) contains papers and review articles on the fundamental and applied aspects of physics in low-dimensional systems, including semiconductor heterostructures, mesoscopic systems, quantum wells and superlattices, two-dimensional electron systems, and quantum wires and dots. Both theoretical and experimental contributions are invited. Topics suitable for publication in this journal include optical and transport properties, many-body effects, integer and fractional quantum Hall effects, single electron effects and devices, and novel phenomena.

Abstracted/indexed in:

Current Contents: Physical, Chemical and Earth Sciences; Aluminium Industry Abstracts, EI Compendex Plus, Engineered Materials Abstracts, Engineering Index, INSPEC, Metals Abstracts, Physics Briefs

Subscription information

Physica E (ISSN 1386-9477) is published bimonthly. For 1998, volumes 2–3 are schedule for publication. Subscription prices are available upon request from the Publisher. A combined subscription with *Physica A, Physica B, Physica C* and *Physica D* is available at a reduced rate.

Subscriptions are accepted on a prepaid basis only and are entered on a calender year basis. Issues are sent by surface mail except to the following countries where air delivery via SAL is ensured: Argentina, Australia, Brazil, Canada, China, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, ,Pakistan, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries air mail rates are available upon request. Please address all enquiries regarding orders or subscriptions to:

Elsevier Science B.V.
Customer Support Department
P.O. Box 211, 1000 AE Amsterdam
The Netherlands
Telephone: + 31-20-485 3757; Fax: + 31-20-485 3432

Claims for issues not received should be made within six months of our publication (mailing) date.

North-Holland, an imprint of Elsevier Science   **Printed in The Netherlands**

PHYSICA E

PHYSICA is published in five sections:

PHYSICA A (*Statistical mechanics and its applications*) publishes research in the field of statistical mechanics and its applications. Statistical mechanics sets out to explain the behaviour of macroscopic systems by studying the statistical properties of their microscopic constituents. Applications of the techniques of statistical mechanics are widespread, and including applications to physical systems such as solids, liquids and gases; applications to chemical and biological systems (colloids, interfaces, complex fluids, polymers and bio-polymers, cell physics); and other interdisciplinary applications to biological, economical and sociological systems.

PHYSICA B (*Condensed matter*) contains papers and review articles in the realm of physics of condensed matter. Both experimental and theoretical contributions are invited, although theoretical papers should preferably be related to experimental results. (For example "A theory on nuclear spin relaxation in metals" would be more suitable for Physica B, while a theoretical discussion on "Screening effects in the electron gas" is more fitting for Physica A.)

PHYSICA C (*Superconductivity*) serves as a rapid channel for publications on superconductivity and related subjects. This includes theoretical papers on the fundamental issues raised by high-T_C superconductivity, reports on measurements of a wide variety of physical properties of high-T_C superconductors, on new materials and new preparation techniques, on thin-film and device-oriented work and on theoretical results pertinent to such experiments. New results in the traditional areas of superconductivity as well as on novel phenomena (e.g. heavy-electron superconductivity) will also be included.

PHYSICA D (*Nonlinear phenomena*) contains papers and review articles reporting experiments, techniques and ideas which, although they may be derived and explained in the context of a particular field, advance the understanding of nonlinear phenomena in general. Contributions of this type in the recent literature have dealt with: wave motion in physical, chemical and biological systems; chaotic motion in models relevant to turbulence; quantum and statistical mechanics governed by nonlinear field equations; instability, bifurcation, pattern formation and cooperative phenomena.

PHYSICA E (*Low-dimensional systems and nanostructures*) contains papers and review articles on the fundamental and applied aspects of physics in low-dimensional systems, including semiconductor heterostructures, mesoscopic systems, quantum wells and superlattices, two-dimensional electron systems, and quantum wires and dots. Both theoretical and experimental contributions are invited. Topics suitable for publication in this journal include optical and transport properties, many-body effects, integer and fractional quantum Hall effects, single electron effects and devices, and novel phenomena.

PHYSICA E

LOW-DIMENSIONAL SYSTEMS & NANOSTRUCTURES

Editors:

T. ANDO
T. CHAKRABORTY
B.D. McCOMBE
M.A. REED
D. WEISS

Proceedings of the Tenth International Winterschool on New Developments in Solid State Physics

"New Frontiers in Low-Dimensional Physics"

held in Mauterndorf, Austria
23–27 February 1998

Guest Editors:

G. Bauer
H. Heinrich

Johannes Keppler Universität
Liz, Austria

F. Kuchar

Montanuniversität
Leoben, Austria
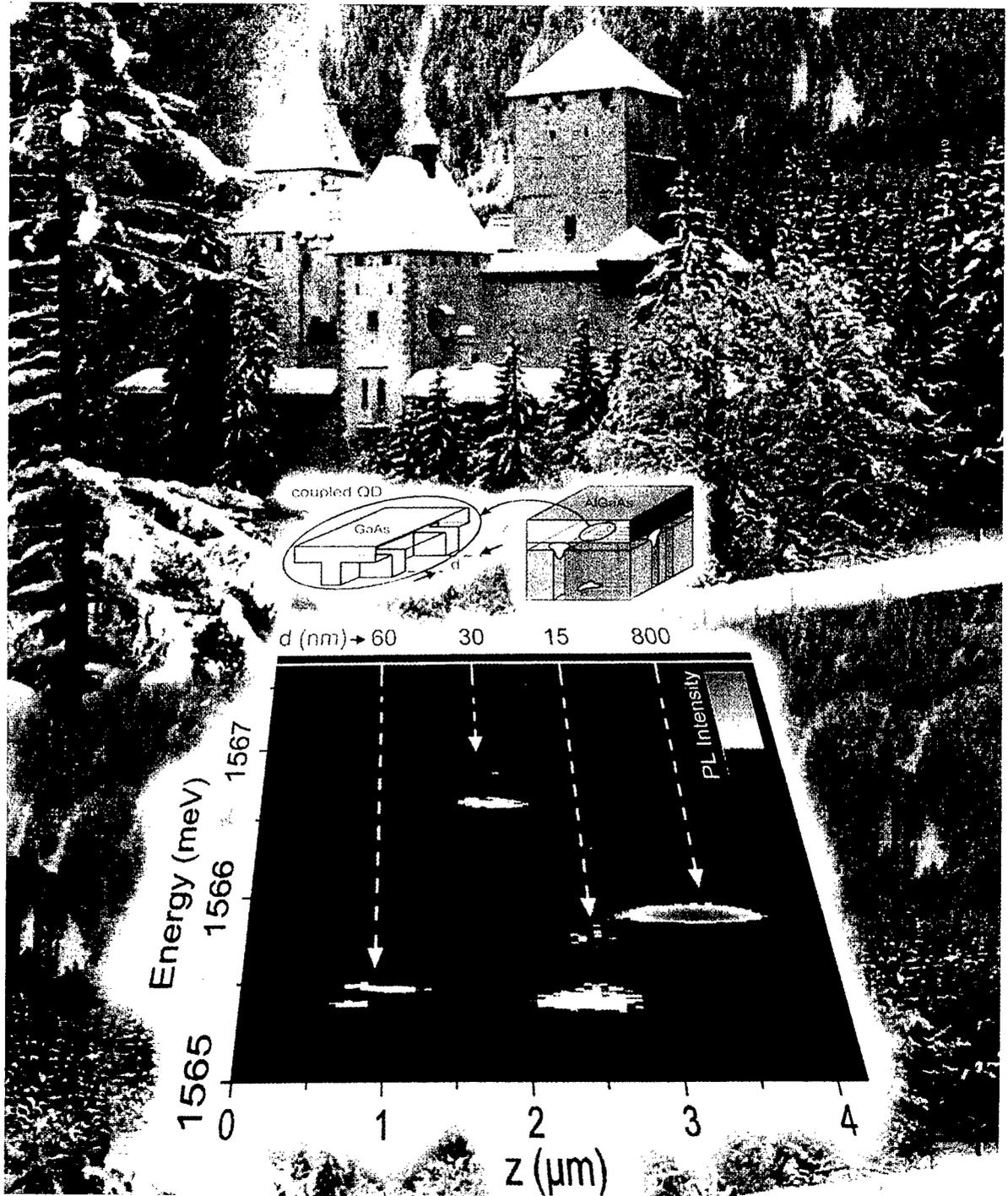
Preface

These Proceedings contain the invited talks which were presented at the 10th International Winterschool on New Developments in Solid State Physics "New Frontiers in Low-Dimensional Physics", held at the Mauterndorf Castle, Mauterndorf, Salzburg, Austria.

The major theme of this Meeting was, like that of its predecessors, most recent developments in the field of the growth, characterization, electronic, optical, acoustic and magnetic properties of low-dimensional systems, including also general aspects like the impact of nanostructure research on conventional Solid State Electronics. Topics which were presented included mechanisms of self organized growth of wire and dot structures, cleave edge overgrowth for the fabrication of atomically precise quantum dots, ground and excited states in quantum dots, spectroscopical investigations of dots, quantum dot lasers, quantum transport in dots. Corrections due to electron–electron interaction to the transport in disordered conductors were treated theoretically and the recently observed metal-to-insulator transition in two-dimensional systems was presented both with its implications for the scaling theory as well as in the light of new experimental results.

Traditional highlights in this series of conferences like the integer quantum Hall and the fractional quantum Hall effect were represented with talks on the direct imaging of the quantum Hall liquid, on higher order composite fermions, on the direct observation of the 1/3 fractional charge and recent advances in the field of single electron transistors for probing of two dimensional electron systems as well as the investigation of 1D and 2D phenomena by using bilayer systems were included.

Progress in lasers was covered by talks on blue emitting heterostructure lasers, on edge and surface emitting dot lasers, and in particular on quantum cascade lasers. A completely new concept was shown for optoelectronic devices based on a combination of surface acoustic waves and of band gap engineered semiconductor devices. Finally the new field of nanoscale magnetic systems was presented with a talk on carrier induced ferromagnetism in two-dimensional II–VI heterostructures.

We hope that these proceedings, as in the past, will provide the readers with a survey of the rapid developments in this subfield of solid state physics.

In addition to the works published here, there were also talks presented by H. Sakaki (University of Tokyo) on "10 nm scale edge and step-quantum wires and related structures: progress in their design, epitaxial synthesis and physics", S.Y. Chou (Princeton University) on "Road to room temperature silicon single electron memory and switch", R.B. Laughlin (Stanford University) "The possibility of d + id order in cuprate superconductors and its relation to the quantum Hall effect", L. Eaves (University of Nottingham) on "Magneto-resonant suppression of conduction in quasi-1D Wannier–Stark superlattices", and F. Capasso (Bell Laboratories Lucent Technologies) on "Quantum cascade lasers: fundamental performance limits and chaotic resonators". About 70 contributed papers were presented as posters, which are not printed in these proceedings.

We would like to thank many graduate students from Linz and Leoben for their invaluable help with the organization of the meeting and Elsevier, in particularly J. Clark, E. van Wezenbeek and J. van Leest for their contributions towards publication of these proceedings.

Günther Bauer,
Friedemar Kuchar
Helmut Heinrich
Guest Editors

Conference Organization

Local Organization

K. Hammerschmid
H. Kirchberger
K. Rabeder
E. Wirtl

Organizing Committee

G. Bauer
F. Kuchar
H. Heinrich

The conference organizers would also like to thank the following companies for their support:

Bruker Analytik GmbH, Germany
Cryophysics GmbH, Germany
Digital Instruments GmbH, Germany
Springer Verlag, Germany
Omicron Vakuumphysik GmbH, Germany
Oxford Instruments GmbH, Germany
Pfeiffer Vacuum Austria GmbH, Austria

Contents

Impact of nanostructure research on conventional solid-state electronics:
The giant isotope effect in hydrogen desorption and CMOS lifetime

K. Hess [a,*], L.F. Register [a], B. Tuttle [a], J. Lyding [a], I.C. Kizilyalli [b]

[a] *Beckman Institute, Coordinated Science Laboratory, Department of Electrical and Computer Engineering and Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Lucent Bell Laboratories, Orlando, FL 32819, USA*

Abstract

A theory of desorption of silicon–hydrogen/deuterium bonds and of depassivation of the silicon–oxide interface of CMOS integrate circuits is presented. First, the physics of two competing depassivation mechanisms and the hydrogen/deuterium isotope effect for each are discussed in the light of recent STM and MOS device studies. A phenomenological model of depassivation in MOS devices based on STM data then is presented that addresses the potential significance of these competing desorption mechanisms and their respective isotope effects on device reliability and lifetime. Finally, initial work to develop a more rigorous, first-principles theory of desorption is described. In the process it is demonstrated that experimental and theoretical methods of nanostructure physics have a direct relevance to conventional silicon-based integrated circuit technology. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 73.40.T; 64.70.K; 71.10

Keywords: Solid state transistors; Semiconductor–oxide interfaces; Density functional theory for condensed matter; Hydrogen passivation of defects

1. Introduction

Nanostructure research and its extension to the atomistic range is seen as a futuristic or even exotic leap beyond current integrated circuit technology. However, the basis of current technology is often rooted in processes that operate on the nanoscale

and beyond, down to molecular or atomistic dimensions. For example, quantum wells are an integral and important part of semiconductor laser diodes, and superlattice intermixing is used for optoelectronic pattern definition [1,2]. Silicon technology has similar examples and size quantization effects are currently included in device design [3]. In this paper we report an isotope effect that was found during research on nanostructure lithography [4] and that turned out to explain the atomistic dynamics of transistor aging and

---

* Corresponding author. Fax; +1 217-244-4333; e-mail: k-hess@ux1.cso.uiuc.edu.
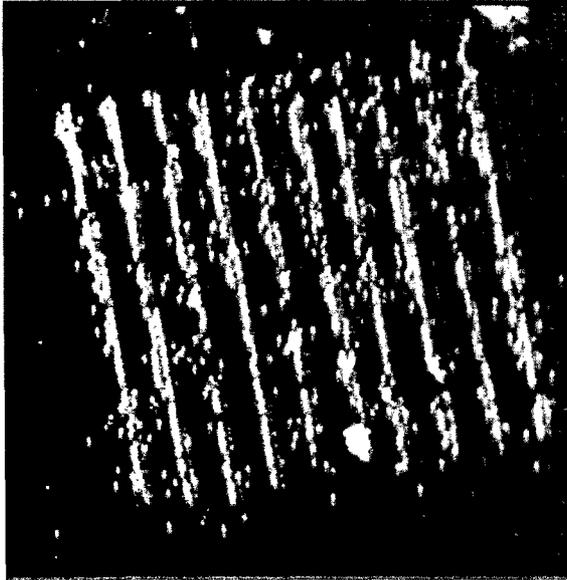
Fig. 1. Atomic scale lines written along single dimer rows by depassivating hydrogen on a silicon (1 0 0) surface.

Fig. 2. Comparison of hydrogen and deuterium desorption yields at 3 V and 11 K as a function of STM current showing a dramatic isotope effect and current dependence; both suggesting a multiple scattering vibrational heating mechanism.

more generally the hot-carrier reliability of integrated circuits [5].

2. Desorption of hydrogen and deuterium from silicon

In the past, we have performed scanning tunneling microscopy (STM) measurements that were geared toward nanoscale lithography [4]. Silicon surfaces were passivated with hydrogen (H) and then selectively depassivated by STM to form silicon nanostructure patterns that could be used for further chemical processing as illustrated in Fig. 1. In the course of these investigations, it was found that passivations with deuterium (D) were significantly more resistant to STM de-passivation. In other words, it took higher voltages or significantly higher STM current densities to remove a given deuterium atom from the surface than necessary for hydrogen. The isotope effect was of the order of a factor of 100 at high STM voltages (injected electron energies) and much higher still at lower voltages. Two mechanisms were identified to be responsible for this effect. At high voltages, excitation of the Si–H or Si–D bonding electron to an excited state is responsible for a force that repulses and desorbs

the H or D atom. This is similar to the well-known electronic surface desorption or photochemistry as described, e.g., by Menzel and Gomer [6], and has a known large isotope effect arising from the greater difficulty of accelerating D away from the surface as compared to H. The second mechanism is probably related to multiple vibrational excitation of the Si–H or Si–D system caused by multiple collisions with the STM electrons during persisting vibrations caused by previous collisions. Because the localized D vibration (at least the bending mode) is more closely matched than those of H to the vibrational energies of the bulk, dissipation to the bulk and cooling is more efficient for D and leads to reduced heating and probability of desorption by multiple collisions [7]. A typical measurement of the desorption yield for H and D in this regime is shown in Fig. 2. These measurements were taken at low temperature (11 K) and 3 V STM voltage. The strong dependence on the STM current is a signature of a process requiring multiple scattering events.

Metal oxide silicon (MOS) transistors are usually processed with hydrogen to saturate the always present dangling bonds at the silicon/silicon-dioxide (Si/SiO_2) interface. It was also known that aging of transistors due to the presence of hot carriers was hydrogen related. Therefore, in analogy to the STM experiments an experiment was performed using MOS transistors (n-channel) and processing with both H and D.
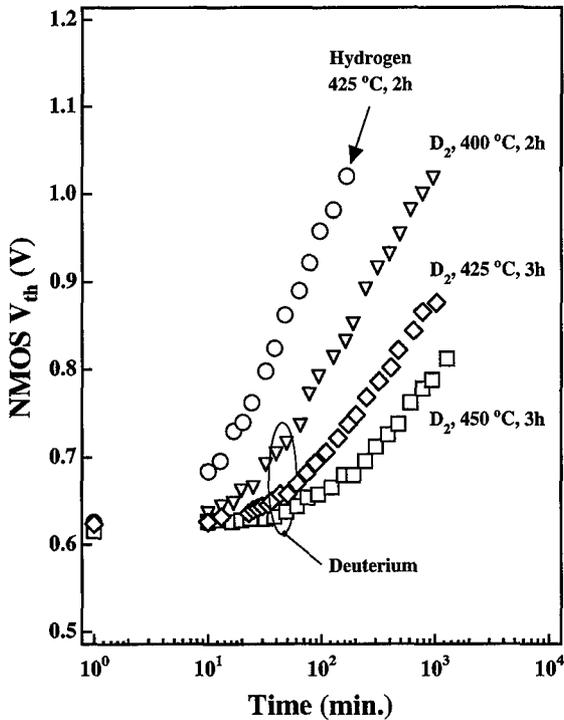
Fig. 3. Threshold voltage of a MOS transistor under stress vs. time. Note that 10% degradation often is regarded to be the safety limit and, therefore, to signify the transistor lifetime. The passivation with deuterium instead of hydrogen shows a clear advantage. The measurements were taken on a fully processed wafer after several layers of metalization.

Then hot-carrier aging experiments were performed [5]. These experiments amount to stressing devices at voltages higher than the operating voltage and monitoring the time-dependent change of device properties such as the threshold (switch on) voltage. A typical result for transistors of complementary MOS (CMOS) technology is shown in Fig. 3. The large isotope effect and its technical potential are obvious and have since been verified extensively.

To address depassivation in MOS devices requires merging a number of diverse fields including hot-carrier transport, molecular bonding and dynamics, and scattering theory as required to model the coupling between hot-carriers and the Si–H/D bond. The first area has been the subject of extensive research over the years. In this manuscript the latter two subjects are addressed. Specifically, in Section 3, a phenomenological model of depassivation based on STM

data is developed to explore the potential significance of these two desorption mechanisms and their respective isotope effects, and in Section 4 initial work to develop a more rigorous, first-principles theory of desorption is described.

3. Phenomenological model of depassivation in MOS devices

The competition between the two desorption mechanisms (single and multiple collisions) could be key to understanding hot-carrier degradation in submicrometer MOS field-effect transistors (MOSFETs). To explore the possibilities, a simple model of depassivation in MOS devices was developed.

To the extent that the two desorption mechanisms can be treated separately, the total desorption rate may be approximately subdivided as

$$D \simeq D_e + D_v, \tag{1}$$

where D_e is the desorption rate via excitation of bonding electron from the bonding to an antibonding state, and D_v is desorption via heating of the vibrational modes of the bonding state. Even if there is mixing of these processes, Eq. (1) still establishes a lower boundary on the desorption rate.

For desorption via excitation of the bonding electron, the desorption rate D_e can be written,

$$D_e \simeq \int_{E_{thres}}^{\infty} dE\, I(E)\bar{\sigma}(E)P(E), \tag{2}$$

where $I(E)$ is the carrier impact frequency on the surface per unit area per unit energy, $\bar{\sigma}_e(E)$ is scattering cross-section for excitation of the bond electron (with a weighted average over the angle of incidence), $P(E)$ is the probability that excitation of the bonds electronic state will actually lead to desorption, and E_{thres} is the threshold energy for scattering.

For desorption via excitation of the vibrational mode(s) of the bond, i.e. "bond heating", the cumulative effect of multiple carrier impacts must be addressed. A model that has proven successful in explaining the STM data [8] is the truncated harmonic

oscillator model for which the desorption rate is given by,

$$D_v \simeq \left\{ \left(\frac{E_b}{\hbar\omega} + 1 \right) [R_{em} + \exp(-\hbar\omega/k_B T_L)/\tau] \right\}$$
$$\times \left[\frac{R_{ab} + 1/\tau}{R_{em} + \exp(-\hbar\omega/k_B T_L)/\tau} \right]^{-(E_b/\hbar\omega)}, \quad (3)$$

where $\hbar\omega$ is the phonon-energy of the Si–H/D bond, E_b is the barrier height to desorption, T_L is the background lattice temperature, and τ is the phonon lifetime. R_{em} is the total phonon emission rate divided by the phonon occupation number plus one,

$$R_{em} \simeq \int_{\hbar\omega}^{\infty} dE \, I(E) \bar{\sigma}_{em}(E)[1 - f(E - \hbar\omega)], \quad (4)$$

and R_{ab} the total phonon absorption rate divided by the phonon occupation number,

$$R_{ab} \simeq \int_{0}^{\infty} dE \, I(E) \bar{\sigma}_{ab}(E)[1 - f(E + \hbar\omega)], \quad (5)$$

where $\bar{\sigma}_{em}(E)$ and $\bar{\sigma}_{ab}(E)$ are the scattering cross-sections for bond-phonon emission and absorption, respectively. The leading term in Eq. (3) (in curly brackets) is the rate of excitation from the vibrational state just below the desorption barrier, $n \simeq E_b/\hbar\omega$, to the state just above the desorption barrier from which H/D desorption quickly follows, i.e. the escape rate from the highest bound state. The following term is the approximate probability of the Si–H/D bond being in the highest bound state. While the truncated harmonic oscillator model could be considered oversimplified, particularly near the barrier top, this latter term represents a quite general physical result, a simple thermal distribution but with a bond temperature greater than T_L established by the competition between the net rate of bond-phonon emission due to carriers incident on the bond and the decay of vibrational energy from the Si–H/D bond to the surrounding crystal lattice.

Scattering cross-sections and other model parameters were obtained by comparison to STM data [9] with $R_{em} \simeq R_{ab}$, with adjustments made for differing surface overlaps of initial and final electron state wave-functions between the MOS system with electrons incident from the channel on the oxide, and the STM system with electrons incident from a vacuum. As discussed in the next section, there may be a number of pathways with differing barrier energies and

vibrational frequencies along which desorption could occur. A full accounting of these will only be possible from first-principles total energy calculations such as are now being initiated, as described in the next section. However, for the purpose of the phenomenological calculations, the bond-phonon energy was taken to be that of the bending mode with a barrier height of 1.5 eV for this mode, which provides a good fit to the current dependence of the STM data for desorption by bond heating. However, a good fit also can be obtained via the stretching mode of the Si–H/D system. (It should be noted that for all practical purposes it is only the product of the phonon-mode lifetime and the scattering cross sections for phonon emission and absorption that are of importance, and can be fit.) The threshold energy for excitation of the bonding electron was taken to be 3.7 eV [10]. For the latter process, STM data was available for essentially the entire energy range above threshold. However, for bond heating, STM data was available only for energies above 2 eV, so the energy dependence of the (STM) scattering cross sections was extrapolated to lower energies using a broad Lorentzian curve fit to the existing data, consistent with the presumption of an energy resonance for this process [8]. Note that this apparent resonance combined with an energy dependence for the interface overlap of the channel electron states leads to a selectivity for hot carriers for this mechanism as well. However, the selectivity for hot carriers still will not be as strong as that for excitation of the bonding electron; any electron above energy $\hbar\omega$ has a non-zero probability of emitting a phonon.

To determine the impact frequencies of carriers on the interface as a function of carrier density and electric field, first the normalized carrier distribution in energy as a function of field was obtained from published results of semiclassical Monte Carlo simulations [11]. This distribution was then multiplied by the impact rate per carrier on the interface as a function of energy (including a rough allowance for nonparabolicity) and the total carrier density.

Fig. 4 shows n-channel MOSFET device lifetimes for 10% depassivation of a hydrogen passivated surface for each desorption mechanism considered separately and for them combined. At high-carrier densities and lower fields, and thus lower electron energies, the desorption rate via bond heating becomes comparable to that via excitation of the bonding elec-
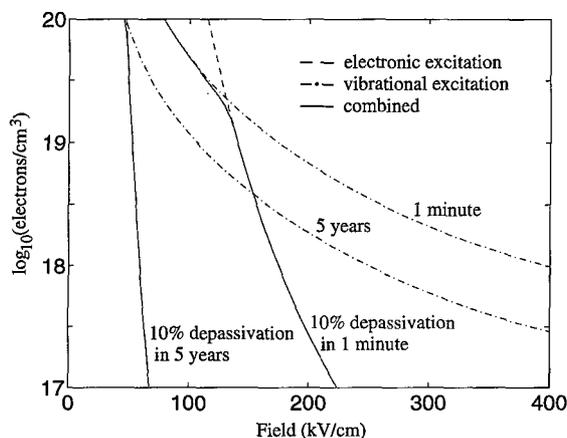
Fig. 4. N-channel MOSFET lifetimes for 10% surface depassivation as a function of interface carrier density and electric field, for depassivation due to excitation of the bonding electron only, due to excitation of the bond-phonon mode (bond heating) only, and due to both mechanisms combined.
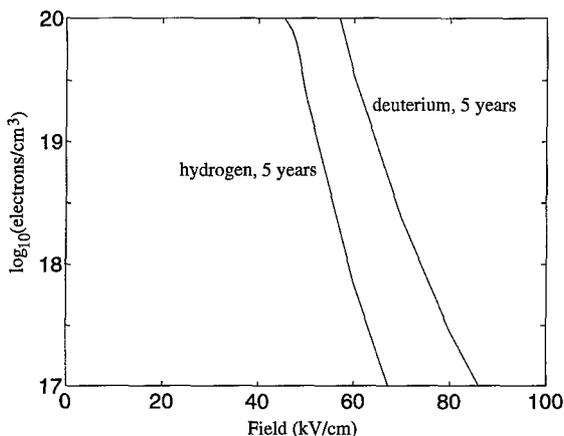
Fig. 5. MOSFET lifetimes for 10% surface depassivation as a function of interface carrier density and electric field, for hydrogen and deuterium passivation.

tron. Further, these results were obtained assuming the electrons have an opportunity to achieve a field-dependent-only distribution; however, in MOSFETs with extremely short channels (0.1 μm), characterized by significant ballistic transport and low supply voltages, the high-energy tails of the electron distribution would be reduced. Thus, in such devices bond heating would be of increased relative importance. Fig. 5 shows the isotope effect on device lifetime as a function of field for both processes combined; for bond heating the effect is larger but under the conditions

assumed in for this figure, excitation of the bonding electron dominates desorption. This latter result is qualitatively consistent with experimental measurements of the isotope effect in CMOS devices [5].

Despite the simplicity of the model, these results suggest that both desorption mechanisms are plausible sources of depassivation in future MOS devices, and that the isotope effect could allow for significantly more aggressive MOS device designs. However, more precise calculations will be required to obtain definitive answers. Hot-carrier distributions must be obtained as a function of specific device geometries and supply voltages, such as only possible with full-band Monte-Carlo methods. Still more challenging, scattering probabilities as a function of incident-carrier (electron or hole) momentum must be calculated from first-principles for both the incident-carrier-bond-electron interaction and for the incident-carrier-bond-phonon interaction. As noted, STM data is only available within limited energy ranges to date, and while it can be expected that desorption of Si–H/D bonds at a silicon–oxide interface occurs by the same fundamental mechanism as at the silicon–vacuum interface (as evidenced by a qualitatively similar isotope effect for both systems) quantitative differences between the two systems are possible. The first step toward these goals is the development of methods for detailed first-principles total energy calculations of the Si–H/D bond.

4. First-principles total energy calculations

As discussed in the previous sections, hot-carriers provide a source of energy to allow desorption of H/D from Si–H/D bonds at the Si–SiO$_2$ interface. These desorbed H/D leave behind dangling bonds (Si$_{db}$) and thereby create electronic defects. The dissociation of Si–H bonds is described by the equation

$$Si\text{–}H \longrightarrow Si_{db} + H^*, \tag{6}$$

where H*, the final configuration for the hydrogen atom, depends on the local chemistry and is currently, for the most part, unknown. Several relevant configurations have been investigated with first-principles total energy calculations [12,13]. These calculations are based on density functional theory within the local
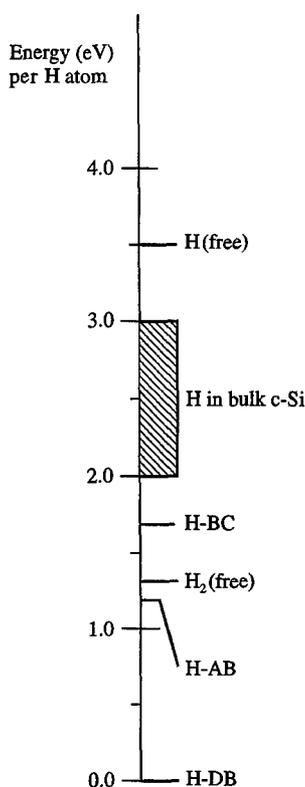
Fig. 6. Relative total energies per H atom important for Si–H/D dissociation at the Si–SiO$_2$ interface as determined from first-principles total energy calculations.

Fig. 7. Local configuration for an isolated Si–H bond in silicon along with the competing dissociation paths. The large filled circles represent silicon atoms and the small open circle represent a hydrogen atom.

density approximation. We employ ab initio pseudo-potentials, plane waves for the electronic basis functions, and periodic supercells to model the bulk environment. For details regarding the calculations see Refs. [12,13] and references therein.

The energetics for some relevant configurations are presented in Fig. 6. The zero of energy is set at the energy of H passivating an isolated dangling bond in bulk crystalline silicon. The energy of this Si–H bond should be similar to such bonds at the Si–SiO$_2$ interface and in bulk a-SiO$_2$. The candidates for H* in Eq. (6) include neutral hydrogen or H$_2$ molecules in an open SiO$_2$ interstice which can be approximated by the respective species in free space (H(free) and H$_2$(free) in Fig. 6), and hydrogen interstitials in bulk c-Si or bulk a-SiO$_2$. The range of energies given in Fig. 6 for H in bulk c-Si indicates of the uncertainty of the calculation and the variation of the H interstitial energy as

a function of the Fermi level. Unfortunately, no first-principles calculations exist for H in bulk SiO$_2$; their energies may be somewhat lower than those of H in c-Si. Note that H and D are electronically identical, and so, therefore, are there bonding energies; the isotope effects results from differing dynamic properties of the bonds due to the differing masses of the isotopes.

To quantify H/D desorption and transistor lifetime predictions necessitates an understanding of the atomistic mechanisms and associated energies for processes described by Eq. (6) for the Si(1 0 0)–SiO$_2$ interface. We have, as a first step, examined the adiabatic dissociation of an isolated Si–H bond in a crystalline silicon environment. The two dissociation paths are indicated in Fig. 7. We have determined from first-principles total energy calculations the structure, energetics and electronic structure of H to dissociate from the dangling bond (DB) site into neighboring minima, which for Path II is at the AB site.

Given a sufficiently large void above an Si–H bond, the dissociation along Path I of Fig. 7 will end in the formation of a silicon dangling bond and a neutral H in free space with the corresponding desorption energy E_d ∼3.6 eV (see Fig. 6). Of course, at the Si–SiO$_2$ interface the local chemistry above the Si–H bond may reduce E_d significantly. Upon examining the electronic structure of H along Path I, we find,

after a displacement of 0.5 Å at a relative total energy of 1.6 eV, a localized electronic state forms near the conduction band edge. Thus, before full dissociation occurs in the neutral charge state the H complex can accept free electrons, significantly reducing further barriers to dissociation.

For Path II, the hydrogen initially moves perpendicular to the bond direction, as indicated in Fig. 7, and then curves toward the BC site which is a local maximum. Finally, the hydrogen will rest at the AB site (see Fig. 7). The maximum at the BC site is rather flat. For instance, sites within 0.5 Å of the BC site have roughly the same energy which is $\sim$1.7 eV higher than H-DB. Also, the H-BC configuration has an associated localized electronic state near both band edges. The defect wave functions are localized on the H atom and the silicon dangling bond. A sketch of the H-BC defect is given in Fig. 6. Near the BC site the hydrogen complex can accept both free carriers and holes. The hydrogen at the AB site is lower in energy than at the BC site by 0.5 eV. Also, no localized electronic gap levels are associated with hydrogen at the AB site. Finally, we find that lattice relaxations are significant for Path II, lowering the final total energies by greater than 2 eV. Such large lattice effects indicate that Path II may be less favored for nonequilibrium desorption. An examination of Si–H dissociation in models of the Si–SiO$_2$ interface will be the next step.

5. Conclusions

We have shown that there exists a close correspondence of STM hydrogen and deuterium desorption measurements on silicon surfaces and the hot-carrier degradation and aging of transistors. Experimental and theoretical methods of nanostructure physics have thus a direct relevance to conventional silicon-based integrated circuit technology. Rudimentary beginnings of a first principle theory of these effects have demonstrated that transistor aging by hot carriers as well as the associated giant isotope effect can be qualitatively understood by combining hot-carrier transport, a scattering theory for the hot-carrier interaction with the Si–H/D bond, and a calculation of desorption energies by density functional theory. The wide variations of desorption energies depending on the desorption pathway, the complications due to the interface with silicon dioxide and the details of hot-carrier interactions with the Si–H bond, indicate however, that there is much work yet to be performed before it will be possible to quantitatively understand the isotope effect in transistors and to predict transistor lifetimes.

References

[1] S.L. Chuang, Physics of Optoelectronic Devices, Wiley, New York, 1995.
[2] W.D. Laidig, N. Holonyak, Jr., M.D. Camras, K. Hess, J.J. Coleman, P.D. Dapkus, J. Bardeen, Appl. Phys. Lett. 38 (1981) 776.
[3] K. Hess, Advanced Theory of Semiconductor Devices, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[4] J.W. Lyding et al., Applied Surface Science 130–132 (1998) 221.
[5] J.W. Lyding, K. Hess, I.C. Kizilyalli, Appl. Phys. Lett. 68 (1996) 2526, and I.C. Kizilyalli, J.W. Lyding, K. Hess, IEEE Electron Dev. Lett. 18 (1997) 81.
[6] D. Menzel, R. Gomer, J. Chem. Phys. 41 (1964) 331.
[7] C.G. Van de Walle, W.B. Jackson, Appl. Phys. Lett. 68 (1996) 2441.
[8] T.-C. Shen et al., Science 268 (1995) 1590.
[9] E. Foley, J. Lyding, Phys. Rev. Lett. 80 (1998) 1336.
[10] C.H. Hu et al., Trans. Electron Dev. 32 (1985) 375 and references therein.
[11] A. Abramo et al., IEEE Trans. Electron Dev. 41 (1994) 1646.
[12] B. Tuttle, unpublished.
[13] C.G. Van de Walle, Phys. Rev. B 49 (1994) 4579.

Scanning single-electron transistor microscopy: Imaging individual charges

M.J. Yoo *, T.A. Fulton, H.F. Hess, R.L. Willett, L.N. Dunkleberger, R.J. Chichester, L.N. Pfeiffer, K.W. West

Lucent Technologies Bell Laboratories, Murray Hill, NJ 07974, USA

Abstract

We describe the single-electron transistor scanning electrometer (SETSE), a novel scanned probe microscope capable of mapping static electric fields and charges with submicron (100 nm) spatial resolution and fractional electron charge sensitivity ($0.01e$). The active sensing element of the SETSE is a single-electron transistor fabricated at the end of a sharp glass tip. Images of the surface electric fields of a GaAs/AlGaAs heterostructure sample taken before and after brief exposures to light show individual photoionized charge sites as well as 100 nm length scale fluctuations in the dopant and surface charge distribution. We also describe SETSE images and measurements of depleted regions, local capacitance, band bending, and work functions at submicron length scales on the surface of this model semiconductor device. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Scanning single-electron transistor microscopy

Inspired by the development of the scanning tunneling microscope, a variety of surface scanning probes have been developed to measure and map electrical properties of materials surfaces on a microscopic scale. For example, noncontact techniques such as scanning capacitance microscopy [1], scanning Kelvin probe microscopy [2], and electric-field-sensitive atomic force microscopy (EFM) [3] have all been explored. Indeed, the last has in one instance shown the remarkable ability to detect the presence of individual charges and to obtain images of insulating surfaces in which a charged spot of one or two electrons is apparent.

We describe a novel low-temperature scanning electrometer [4] operating on a different principle, one which has one to two orders of magnitude greater charge resolution and a similar spatial resolution (100 nm) compared with the EFM. This microscope, the single-electron transistor (SET) scanning electrometer, or SETSE, uses the SET as a probe to sense the electrically induced charge on its small (100 nm) metal island held in proximity to the sample surface (Fig. 1C). It can detect ~1% of an electron charge ($0.01e$). Because all of the important geometrical parameters are known, one can assign a quantitative interpretation to the SETSE signal. Also, during

* Corresponding author. Current address: IBM Research, P.O. Box 218, Yorktown Heigths, New York, NY 10598, USA. Fax: +1 914-945-4531.

operation the SETSE, unlike the EFM, does not require the application of high electric fields (10^6 V cm^{-1}) between the tip and surface, an important consideration for many interesting but easily perturbed semiconductor systems. All of these features enable a broader class of experiments to be explored.

As an example, we studied the electric fields at the surface of a semiconductor GaAs/Al$_x$Ga$_{1-x}$As heterostructure sample. These fields arise from localized charges at and near the surface as well as from the voltage bias and work function of any underlying electrode. Statistical fluctuations in the density of surface charges are evident, although the individual charges are too closely spaced (<10 nm) to be resolved directly. However, we are able to image the more widely spaced individual charges that are produced by brief light exposure. These photoinduced charges change the surface electric field to reveal isolated, well-resolved sites where dopant atoms or other electron traps have lost or gained an electron. In addition, we acquired images (with <1 mV resolution) of surface potential variations near biased electrodes, images related to band bending near sample edges, and images of surface fields and corresponding local capacitance. We also describe some measurements of the local work function.

1. Microscope and sample description

The SET is a submicrometer-sized tunneling device whose current flow is governed by the Coulomb blockade effect [5]. It consists of a small metal island connected to metal source and drain leads by two small tunnel junctions. The current tunnels through the junctions at a rate determined by the island's electrostatic potential with respect to the source and drain. This potential is in turn controlled by the electric field that the island experiences from external sources, such as fixed charges of capacitively coupled electrodes on a nearby sample (Fig. 1). At low temperatures and proper voltage bias, the current flowing through the SET fluctuates periodically as this electric field increases. In fact, the current passes through a full period each time the electric field lines terminating on the island induce a charge of exactly one additional electron (Fig. 1A) [6]. Hence, monitoring of the current through the SET as it is scanned over the sample provides a means of

mapping the electric field emanating from the sample surface [7].

Fabrication of the SET probe tip involves the evaporation of three separate areas of a thin (10–20 nm) aluminum film onto a specially shaped glass fiber. The end of the fiber has a shallow conical taper that terminates at the tip in a flat, nearly circular area ~ 100 nm in diameter (Fig. 1B and C) [8]. A circular patch of film covering the tip constitutes the field-sensitive island. The films for source and drain leads spread out from the edges of the tip and extend up the sides of the fiber to electrical contacts. The source and drain leads are deposited first by separate evaporations from the side and rear. After an in situ exposure to oxygen that creates the oxide tunnel barriers, a final end-on evaporation deposits the island and forms the tunnel junctions. The three electrode shapes are defined by natural shadowing.

The sample described in this work has electrostatic features on a variety of length scales. It is a GaAs/Al$_x$Ga$_{1-x}$As heterostructure grown by molecular beam epitaxy (Fig. 1C). A δ-doped layer of Si atoms of density 5×10^{12} cm^{-2} is grown 22 nm below the sample surface within an Al$_x$Ga$_{1-x}$As region. Many of these Si atoms ionize and act as electron donors. Most of the electrons are trapped in states at the GaAs surface, whereas a small fraction go to the GaAs/Al$_x$Ga$_{1-x}$As interface located 60 nm below the sample surface. Here they form a metallic sheet, a two-dimensional electron gas (2DEG) [9]. To describe its electrostatic behavior, the sample may be regarded as a conductor (the 2DEG) topped by 60 nm of insulator (with a dielectric constant of 13) containing the donor and surface charge layers [10]. Potential fluctuations on short length scales are produced by these charge layers. Electric fields on a larger scale may be produced by the application of a voltage between the 2DEG and a series of 1 μm-wide, 30 nm-thick metal stripes evaporated upon the sample surface. A sufficiently negative voltage applied to one of these gates depletes the 2DEG, driving the electron gas laterally away from its edge.

The SET-tipped fiber is installed in a low-temperature scanning microscope probe stage, allowing three-dimensional positioning of the tip near the sample with sub-nanometer precision and stability. For best resolution and signal, the tip is held
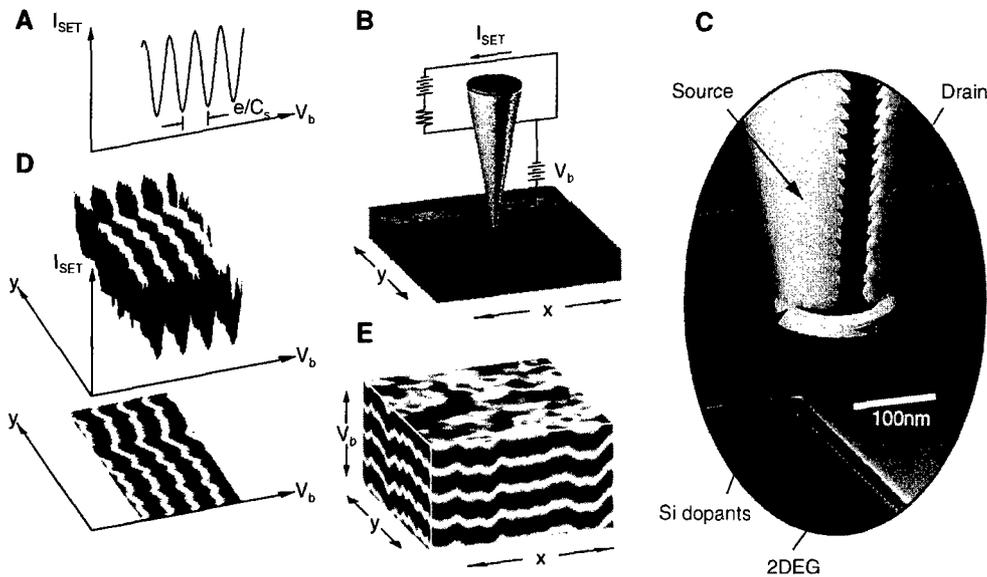
Fig. 1. (A) Typical current oscillations $I_{SET}(V_b)$ of a SET. (The zero of I_{SET} is offset; the amplitude is actually $\sim$20% of the average). (B) Schematic depiction of the SET probe tip suspended above the GaAs/Al$_x$Ga$_{1-x}$As heterostructure near a gate. (C) Magnified view of the tip and a cutaway view of the sample. (D) $I_{SET}(V_b)$ versus y for a 2 μm line scan: (top) 3D representations after (A); (bottom) 2D color representation of the same data. The wiggling of the stripes is caused by variations in the electrical charge density of the surface. (E) Color representation of a complete data set of $I_{SET}(V_b)$ versus x and y for a 2 μm by 2 μm raster scan over a nongated region. The top of the data block maps the electric field of the sample surface as detected by the SETSE. The side of the block is taken from (D).

as close to the surface as possible without contact, typically at a height of 25 nm. Electrical contacts (Fig. 1B) enable the application of a voltage V_b between the SET and the 2DEG (or other electrodes) on the sample and also provide the means of biasing the SET and monitoring its current.

2. Theory of operation

The electric field between tip and sample is monitored by recording the SET current I_{SET} as the probe is scanned without feedback in a plane at a fixed height z above the planar sample surface. Collection of a typical data set involves positioning the probe over the surface and sweeping the 2DEG bias voltage V_b by an amount sufficient to cause the induced charge on the island to vary by several electrons. This charge variation in turn causes $I_{SET}(V_b)$ to oscillate through several full cycles, one for each electron (Fig. 1A). A sequence of such curves is taken as the tip is scanned along a line in the y direction (Fig. 1D). Finally, a raster scan

in x and y results in a complete $I_{SET}(x, y, V_b)$ data set (Fig. 1D); contour lines in the xy plane indicate lines of constant electric field, with successive contour lines representing electric fields that differ by a fixed amount corresponding to an induced charge of one electron. Acquisition of such a data structure is typical of our measurements.

Because I_{SET} varies nearly sinusoidally with V_b [11], we can characterize I_{SET} at each point by a period and phase. The period (e/C_s) (Fig. 1A) is determined by the capacitance C_s between the tip and the sample electrode and varies primarily with height z. The phase, which essentially counts the number of electrons induced on the tip, contains most of the information about the spatial distribution of electric fields. It can be regarded as resulting from an effective surface potential V_s. That is, one may model the SET current as

$$I_{SET} = A \sin \frac{2\pi Q}{e} = A \sin \frac{2\pi C_s (V_b + V_s)}{e}.$$

Here Q is the total charge induced on the probe tip by the sample, expressed as a part $C_s V_b$ due to the electrode bias and a part $C_s V_s$ induced by all other sources.

Least-squares fits of $I_{SET}(V_b)$ at each tip position yield maps of the effective potential V_s and the local capacitance C_s [12].

The surface potential V_s can be understood as the sum of two contributions: the work function difference (contact potential) between the probe and the sample [13] and the potential from charged centers distributed across and below the surface. For every charge beneath the tip, only a fraction f of the total electric flux terminates on the island [14]: An individual electron charge produces a contribution fe to Q and fe/C_s to V_s. The fraction f increases with both lateral and vertical proximity of the probe to the charge, approaching a maximum of 10% for typical scan height of 25 nm. A single-electron charge on the surface can then induce $0.1e$ on the probe island (a change in V_s of ~2.5 mV), about 10 times the noise level.

The capacitance C_s depends on the height of the tip above the surface and is used to measure and set this height. A parallel-plate picture predicts that the capacitance varies as $1/z$ when the scan height is small compared with the tip diameter. At large z, a weaker $1/\log(z)$ dependence is expected [15]. The period of the oscillations in $I_{SET}(V_b)$ (Fig. 2A), given by e/C_s, decreases sharply as the tip approaches the sample surface. Because the capacitance is infinite for zero separation between the tip and 2DEG, the location of the sample surface can be determined to within 2 nm, allowing for dielectric corrections, by extrapolation of the curve of inverse capacitance versus scan height (Fig. 2B). Fits to the z dependence also give estimates for the tip diameter of order 100 nm.

The local contact potential of the sample can also be extracted from the data of Fig. 2A. An examination shows that there is a bias voltage for which the induced island charge does not vary as the tip is lowered to the sample surface. At this voltage (here ~0.6 V), the work function difference between the SET and sample has been nulled. For most SETSE measurements, we operate with a bias approximately nulling the contact potential, which minimizes any perturbation of the sample by electric fields from the tip.

3. Imaging individual charges

Maps of V_s constructed from xy scans of the same 2DEG region initially show increasingly fine random
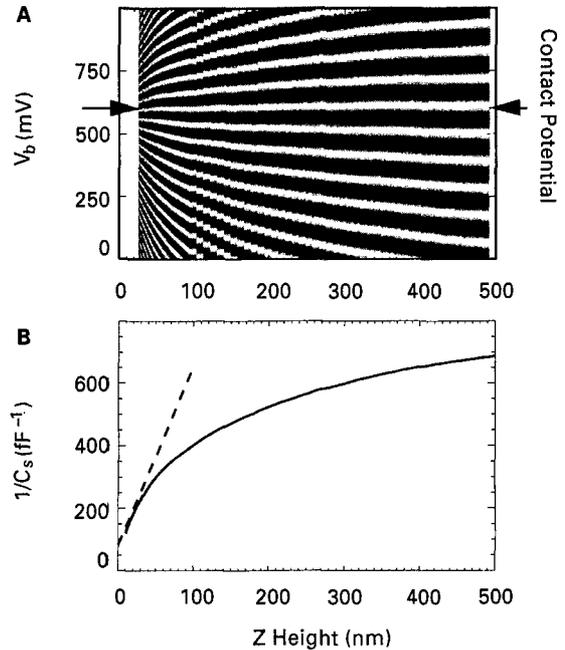
Fig. 2. Capacitance versus scan height. (A) $I_{SET}(V_b)$ at fixed location (x, y) along a sequence of heights z above the 2DEG. The convergence of the stripes is caused by proximity to the surface. Arrows indicate the contact potential voltage, from which the local work function can be obtained. (B) Inverse capacitance $1/C_s(z)$ as extracted from the period of the data in (A) (solid) and as predicted by parallel plate model (dashed).

structure with decreasing z (Fig. 3), eventually revealing 100 nm features. This resolution is limited by the tip diameter. The fluctuation amplitude also increases as the tip approaches the surface, reflecting the enhanced coupling between the SET and sample charges. Typically, there are about 300 surface and ionized dopant charges underneath the tip. We believe that the statistical fluctuations in this number across the sample are imaged in Fig. 3C; however, numerical simulations indicate that a random placing of these dopants and trapped charges would produce fluctuations in V_s nearly twice as large as those observed in Fig. 3C, which suggests a more uniform distribution. Such fluctuations in the potential play an important role in limiting the mobility of the underlying 2DEG and give a means of characterizing the disorder.

The SETSE can image individual electron charges on and within the semiconductor sample (Fig. 4). In the absence of light, maps of V_s (Fig. 3) are repeatable to within 1%. This precision allows one to im-
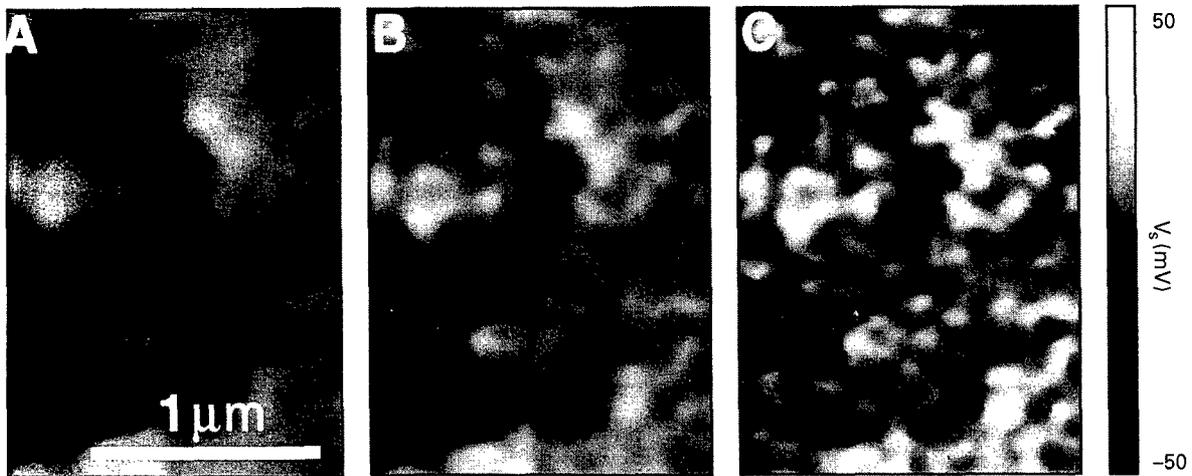
Fig. 3. Surface potential V_s in the xy plane (relative to the average value of each image) showing fluctuations from dopants and surface charges as seen at scan heights of (A) 95 nm, (B) 50 nm, and (C) 25 nm. The spatial resolution and sensitivity improve with height.
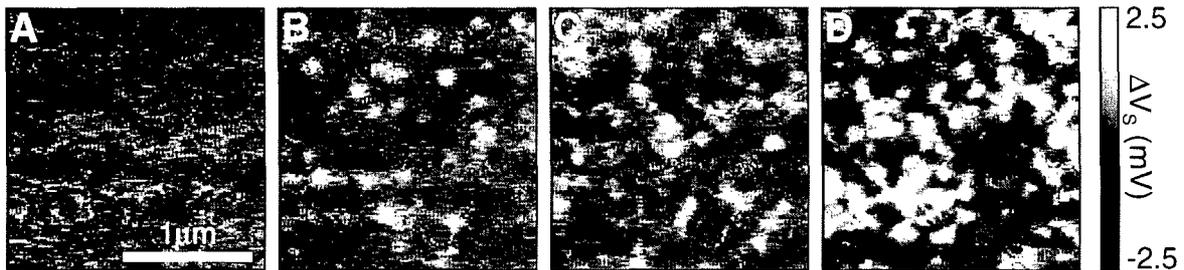
Fig. 4. Photoactivity of single dopant atoms and surface traps shown in xy maps of changes in $V_s(\Delta V_s)$, induced by exposure to (A) no light and light at (B) 850 nm, (C) 770 nm, and (D) 690 nm. Individual photo-ionized (neutralized) sites with a single-electron charge are seen as light (dark) spots; that is, light spots mark new positive charges, and dark spots, new negative charges. The sensitivity of the color scale is about an order of magnitude greater than that in Fig. 3.

age isolated photo-ionized or neutralized sites (Fig. 4) by subtracting maps made before and after brief, low-intensity illumination. The differences show up as a few small circular spots, presumably donor atoms or traps, whose apparent size ($\sim$100 nm) is limited by the spatial resolution. Sufficiently energetic photons create free carriers that can transfer electrons between various sites, primarily the silicon dopants and the surface traps, converting some neutral sites into charged ones and vice versa. The more intense signals are likely from higher (surface) sites, and the weaker signals, from lower (donor layer) sites. Simulations of the charge-coupling fraction f for a single charge of $\pm e$ buried at these levels predict changes in V_s of about 1.5–2.5 mV, as observed. It should be noted that the SETSE senses net charge differences only, so the data,

in principle, also can be explained by a single atom or charge trap with more complicated photoactive mechanism involving several localized charges with total net charge $\pm e$. The sample is markedly more photoactive for wavelengths above the band gap (760 nm) of the $Al_{0.15}Ga_{0.85}As$ layer. Here greater absorption and a larger number of photogenerated carriers within the $Al_xGa_{1-x}As$ in the region of the donors are responsible for the increased response.

4. Other applications

The SETSE has also been used to measure other electrostatic potentials on larger length scales. For example, the electric fields of the charged surface states

are responsible for band bending and can be detected externally by the SETSE. The SETSE can also map electrostatic potentials induced by applied gate voltages. Surface potential changes at such distances from the gate are sometimes observed to be hysteretic and may result from charge injection into the surface or dopant layers.

Local capacitance measurements from the SETSE provide topographical information complementary to electric field data. Because of the small well-defined island size, there is less stray capacitance background than in other capacitance-based scanned probe techniques, enabling a more quantitative measurement. Furthermore, the SETSE can be operated at frequencies approaching DC, permitting the study of systems with slow response.

5. Conclusion

Further improvement of the SETSE will involve development of SETs and probes with higher sensitivity or smaller dimensions. The resolution of the present SETSE could be increased to perhaps 10 nm [16] with smaller fibers and tips and somewhat thinner films. This design would increase the operating temperature in proportion. Even room-temperature operation, although ambitious, in conceivable, requiring the development of molecular- or even atomic-sized SET tips. There are a number of fundamental and applied problems that might benefit from this microscopic technique with its sensitivity to potential, charge, capacitance, contact potential, conductance, and dielectric composition. For instance, the properties of the electrons in the quantum Hall regime and their edge states [17] are attractive subjects for SETSE experiments. In principle, single-electron capacitance spectroscopy [18], which is a powerful technique for recording the complete electron energy structure of mesoscopic systems, could be implemented with the SETSE to add spatial mapping capability.

Acknowledgements

References

[1] C.C. Williams, J. Slinkman, W.P. Hough, H.K. Wickrama-singhe, Appl. Phys. Lett. 55 (1989) 1662.

[2] M. Nonnenmacher, M.P. Ogoyle, H.K. Wickramasinghe, Appl. Phys. Lett. 58 (1991) 2921; A.K. Henning et al., J. Appl. Phys. 77 (1985) 1888.

[3] Y. Martin, D.W. Abraham, V. Raineri, H.K. Wickramasinghe, Appl. Phys. Lett. 52 (1988) 1103.

[4] M.J. Yoo et al., Science 276 (1997) 579.

[5] T.A. Fulton, G.J. Dolan, Phys. Rev. Lett. 59 (1987) 109; L.S. Kuzmin, K.K. Likharev, JETP Lett. 45 (1987) 389; D.V. Avering, K.K. Likharev, in: B.L. Altshuler, P.A. Lee, R.A. Webb (Eds.), Mesoscopic Phenomena in Solids, North-Holland, Amsterdam, 1991, pp. 173–271; See also H. Grabert, M.H. Devoret (Eds.), Single Charge Tunneling, Plenum, New York, 1992.

[6] The net charge on the island is an integral number of electrons. Any (nonintegral) variable charge induced on the island by external sources is compensated by charges of opposite sign induced by the capacitance to other electrodes, primarily that of the tunnel junctions to the source and drain. For the SETs used here, typical numbers are voltage biases $\sim$1 mV, temperatures $<$2 K (required so that the thermal energy $kT < e^2/C$), currents $\sim$1 nA with field-induced variations of $\sim$50%, total island capacitances of $\sim$0.1 fF, and sample-to-island capacitance of $\sim$0.001 to 0.01 fF. Scanning electron micrographs show typical island diameters of 100 nm.

[7] The field measurements referred to in the text are more accurately measurements of the electric flux terminated on the tip island. One must also allow for distortion of the field from the presence of the metal electrodes of the tip, particularly in proximity to a conductor, where the surface fields are altered to a greater degree.

[8] Such substrates were fabricated from bare 80 μm-diameter single-mode optical fiber by a process described by E. Betzig, J.K. Trautman, T.D. Harris, J.S. Weiner, R.L. Kostelak, Science 251 (1991) 1468.

[9] The 2DEG has an area density of 1×10^{11} electrons per square centimeter.

[10] Surface charges from sources other than the ionized silicon donors and their electrons pinned at the surface, such as surface debris or surface ions [K. Domansky, Y. Leng, C.C. Williams, J. Janata, D. Petelenz, Appl. Phys. Lett. 63 (1993) 1513], might also contribute. However, it is not necessary to invoke such additional sources to explain the data.

[11] The sinusoidal dependence approximately describes the observed behavior at the comparatively high temperatures and magnetic fields that were used.

[12] In this expression, V_s is determined only modulo e/C_s, and the values are chosen so that V_s varies smoothly with position within a map.

[13] Because variations in the work function can be modeled as a charge dipole layer at the conducting surface, there is ambiguity in distinguishing between unresolved charge layers close to the electrode and variations in work functions.

[14] The flux terminates on the tip and the 2DEG in rough proportion to their distances from a centered charge weighted by the dielectric constants with the 2DEG receiving most of the flux.

[15] C.F. Eyring, S.S. McKeaown, R.A. Millikan, Phys. Rev. 31 (1928) 900.

[16] SETs having sharpened tips showed good response at 4.2 K in preliminary tests and might display resolutions of order 50 nm.

[17] D.B. Chlovskii, K.A. Matveev, B.I. Shklovskii, Phys. Rev. B 47 (1993) 12605.

[18] R.C. Ashoori, Nature 379 (1996) 413.

Paired additions of electrons into quantum dots: How and why does localization destroy Coulomb blockade to produce pairing?

R.C. Ashoori [a,*], N.B. Zhitenev [a], L.N. Pfeiffer [b], K.W. West [b]

[a] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA*

Abstract

Several years ago, we discovered a profound violation of Coulomb blockade in quantum dots. Not only did electron additions deviate from near perfect periodicity with electrostatic potential on the dot, the additions actually occurred in pairs and sometimes even as bunches. This behavior only occurs in quantum dots that with diameters larger than 0.4 μm. Moreover, it only arises when density of electrons within the dot is low enough such that the mean spacing between electrons is larger than the Bohr radius ($r_s \leqslant 1$). We have completed a systematic study of dots of different sizes. For the largest dots, pairs and bunches occur randomly within the addition spectrum. In intermediate sized dots the pairing occurs periodically with each 4th and 5th electron added to the dot as a pair. In the periodic case, we have strong evidence that one of the two electrons in a pair is localized at the edge of the dot while the other four additions in the sequence are two states mainly localized to the interior. Such a bifurcation of the electron droplet into inner and outer shells may arise from negative compressibility within the dot or other interaction effects. Apparently, *any* type of localization is sufficient to create the pairing and bunching effect. In large dots with randomly localized electrons, the pairing appears at random, while in intermediate sized dots with a core-edge localization, the pairing occurs periodically. While we have succeeded in understanding the conditions necessary for pairing, the physics behind this extraordinary effect remains a mystery. © 1998 Published by Elsevier Science B.V. All rights reserved.

Keywords: Quantum dots; Paired electrons; Localization; Coulomb blockade

1. Introduction

In mesoscopic physics, the electron–electron repulsion created by the Coulomb interaction translates into simple and fairly obvious result. More energy is required to add each successive electron to a puddle of electrons. If electrons spread their charges throughout the puddle in the same fashion as classical charge residing on a metallic puddle of the same shape, then the Coulomb interaction dictates that electron additions occur every time the potential of the puddle is changed by an amount e/C. Here, e is the electron charge, and C is the capacitance of the puddle to the environment.

Despite the fact that the electron density in small quantum dots or "artificial atoms" containing only a

* Corresponding author. Fax: +1 617 258 6883; e-mail: ashoori@ mit.edu.

few electrons deviates substantially from that in a metallic disk of similar size, Coulomb blockade remains a useful and nearly valid approximation to the addition spectra of few electron quantum dots [1–7]. In GaAs quantum dots, this behavior arises principally because the charging energy e^2/C is typically about an order of magnitude larger than single-particle quantum level spacings within the dot. Electrostatics effects largely dominate the spectra.

Given the relatively very large magnitude of the charging energy compared to all other energies in the problem made, our first observations of the charging of large dots (lithographic diameter $\approx 1\,\mu$m) containing few electrons seemed particularly difficult to fathom. While some of the additions appeared random (as might be expected considering that the first electrons are likely added to random minima of an unscreened disorder potential), many exact or nearly exact pairs were present in the spectra [8,9]. With the application of magnetic field, the electron additions comprising pairs would continue to "stick" together until finally splitting at high fields.

The pairing behavior was apparent in large dots, but not in small ones (those with lithographic diameters of 0.3 μm or less). To understand this phenomenon better, we carried out a systematic study of dots of differing sizes. For intermediate sized dots, we found a stunning surprise. The pairing behavior occurs *periodically* with every 4th and 5th electron entering the dots as a pair.

Our single electron capacitance spectroscopy (SECS) experiments are unique in permitting study of the single electron addition spectra of large dots containing few electrons. This is the regime in which pairs are observed, and this is presumably the reason the pairing has not been observed with other experimental methods. Other transport experiments have permitted the study of large dots containing many electrons [3,10] and small dots containing few electrons [7,11]. SECS allows the study of any size dot over an extremely wide range of electron occupancies.

The picture emerging from our studies is the following: Any type of localization is sufficient to create the pairing and bunching effect. In large dots with randomly localized electrons, the pairing appears at random, while in intermediate sized dots with a core-edge localization, the pairing occurs periodically. We have succeeded in understanding the conditions nec-
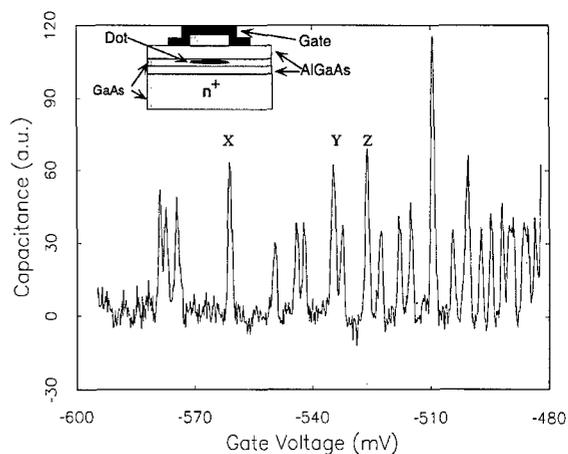
Fig. 1. In-phase signal as a function of gate bias using a 0.40 mV excitation at a magnetic field of 0.87 T. This data is part of the set used to generate Fig. 2b. Inset: Schematic representation of the dot.

essary for pairing, but the physical mechanism behind this extraordinary effect remains a mystery.

2. Samples

A schematic representation of our samples is shown in the inset to Fig. 1a. Several different wafers were used in the studies described here, and similar results were obtained for each one. Here we refer to them as wafers A and B. They are similar to the ones described in Refs. [6,8,9]. The AlGaAs/GaAs wafers contain the following layers (from the bottom to the top): 3000 Å n+ GaAs, 400 Å (150 Å for wafer B) GaAs spacer layer, 136 Å AlGaAs/GaAs superlattice tunnel barrier (85 Å AlGaAs tunnel barrier for wafer B), 175 Å GaAs quantum well (150 Å for wafer B), 500 Å AlGaAs blocking barrier, 300 Å GaAs cap layer. A mesa with deep ohmic contacts down to n+ GaAs is initially defined. Then a circular Cr/Au gate electrode is fabricated on the top of the mesa. 8 dots from sample A and 3 dots from sample B were studied with gate diameters ranging from 1.6 to 0.2 μm. Plasma etching produces a short pillar (300 Å tall) using the gate electrode as a mask. Electrons remain in the quantum well only in the region below the pillar [8]. A larger overlapping metal electrode then provides electrical connection to the gate. Our single electron capacitance

spectroscopy (SECS) measurements are carried out using an on-chip bridge circuit described in Ref. [8].

3. Large dots

Fig. 1 displays the electron addition spectrum at zero magnetic field for a dot of 1000 nm lithographic diameter produced on wafer B. The gate voltage scale can be directly converted to an energy scale $\Delta E = \alpha \Delta V_g$ with the lever-arm $\alpha \approx 0.25$ for samples created on wafer B and $\alpha \approx 0.5$ for samples produced on wafer A. These values are determined from bulk capacitance measurements to determine the geometrical parameters in the wafer as described in Ref. [8]. The gate voltage position of the Nth capacitance peak, when multiplied by the lever arm, directly measures the chemical potential μ_N of the dot containing N electrons [2].

For gate biases below -580 mV in Fig. 1, there are no electron additions to the dot, and the three peaks seen between -580 and -570 mV are for the first three electrons in the dot. A magnetic field of 0.87 T (the effect of field is inconsequential here) is applied normal to the plane of the dot, and the temperature in these measurements is 0.3 K. Not that the spectrum does not at all display the usual periodic Coulomb blockade sequence. However, this is reasonable, because the dot is quite large, and simple calculations of the disorder potential suggest that the first electrons added to this system are localized in isolated potential minima [8,9]. Note also that the mean spacing between peaks decreases as more electrons are added to the system. In the simplest picture, this behavior arises because more potential minima are available to electrons at higher energies.

Interestingly, not all the peaks shown are precisely the same height. This is surprising because in a capacitance experiment the integrated capacitance in any one peak should simply be equal to the total charge induced on the plates of our "quantum dot capacitor" from one electron entering the dot. The solution is simple: some peaks are taller simply because they are close to other peaks; moreover, some peaks have double the height of isolated peaks. As we shall show, these peaks arise from electron pairs.

Fig. 2a and b display the magnetic field dependence of capacitance data such as that shown in Fig. 1. The
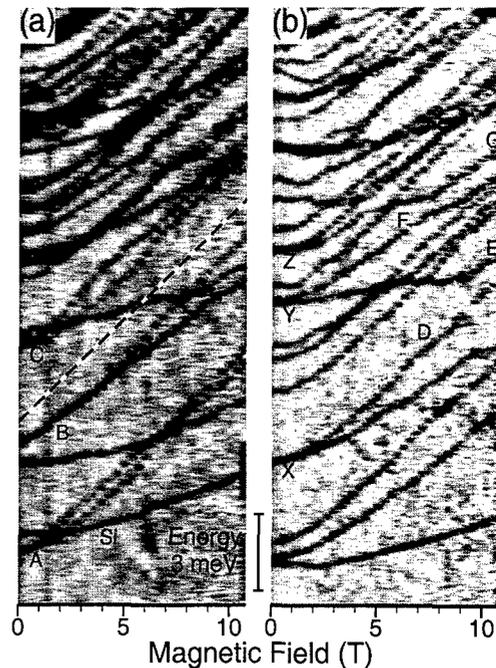
Fig. 2. (a) and (b) Gray-scale images of the sample capacitance as a function of energy and magnetic field. The vertical axis is derived from the gate bias. Curve Si in (a) fits well to the theoretical result for the field dependence of a D^0 state at the center of the quantum well. Such states arise from electrons trapped on silicon donor atoms which migrated into the well during the wafer growth. Samples from other wafers (grown in a fashion to limit silicon migration) do not display these such states, but nevertheless display pairs. The dashed line in (a) represents the diamagnetic shift of a free electron in magnetic field, $\hbar\omega_c/2$ [6].

data are now plotted in gray scale with the vertical axis representing gate bias (here converted into energy units at the position of the dot) and the horizontal axis reads the magnetic field applied normal to the plane of the quantum dot. The dark traces therefore, represent the evolution of the energies required to add successive electrons to the dot. The two data sets in Fig. 2 are from two different thermal cyclings of the same dot. Thermally cycling the sample to room temperature changes the ionization state of the dopants in the sample and thereby changes the disorder potential and the addition spectrum.

There exists an intriguing feature of the data which is apparent in both data sets shown in Fig. 2. We first refer to Fig. 2a. In several cases at zero magnetic field, two electrons enter the well at the same gate bias. This

is apparent in curves A, B, and C which split into doublets at higher fields. This behavior may reflect accidental degeneracies. However, the fact that this phenomenon is observed for *all* those low lying levels in Fig. 4a that are not bound to impurities, suggests a physical mechanism. The fact that the slopes seen in Fig. 2a all of steep slope variety indicates that the degeneracies are not associated with an impurity site.

Examining Fig. 2b, we also find some degeneracies at zero magnetic field. These are labeled X, Y, and Z. The data in Fig. 1 were taken during the same cool-down of the sample, and the X, Y, and Z peaks are the same ones as those labeled in Fig. 2b. Following the upper branch, D, of X to high fields, there is evidence for level repulsion with a state labeled E. We infer from this level repulsion that states D and E reside in the same puddle. Tracing state E back to low magnetic fields, we find that it is one of a pair of states, Y, degenerate in low field. Following the upper branch of this pair, F, to high fields, we see level repulsion with the state, G. Curve G belongs to another low field degenerate pair, Z. This behavior suggests that there exists a site which adds two electrons at a time rather than just the first two electrons. This pairing behavior cannot simply result from coincidental overlaps of peaks. We consistently observe pairs in separate sample cool-downs, and the pairs are seen to "stick" together over a wide range of fields.

We note that all of our experiments are performed with the dot effectively in equilibrium with the charge reservoir in the bottom electrode beneath it. The AC excitations used to perform the capacitance measurement are always kept smaller than kT in the quantum dot. Usually, electrons avoid entering the same potential minimum due to electron repulsion (Coulomb blockade). We observe the reverse. The transfer of one electron into the well causes a second to follow, although we cannot determine whether they occupy the same potential minimum. While this behavior seems reminiscent of an Anderson negative-U system [12], we find serious discrepancies with such models.

Two theoretical models [13,14], have been suggested to explain the origin of the exact pairs. Both models predict a dramatic suppression of the tunneling rate as soon as two electrons are joined into a pair, since both electrons must be added into the dot in a coherent fashion. Having studied a large number of exact pairs in the frequency range 50 KHz–1 MHz we have never observed a significant drop of the tunneling rate when the traces merge. This suggests that the paired electrons tunnel into the dot *independently*, although they are added to the system at precisely the same gate voltage. Remarkably, the data indicate that filling one state of a pair has no effect on the energy of the other state in the pair. Rather than behaving as a negative-U system, the pair acts instead as a "zero-U" system in which the interaction between the electrons has simply been cancelled.

4. Small and intermediate sized dots

The pairing behavior has been apparent in large dots, but smaller sized dots did not display pairs. Instead, the smallest dots display spectra consistent with calculations [6] for artificial atoms. To explore this behavior, we conducted a systematic study of different size quantum dots. We discovered several astonishing results in crossover regime between small and large size dots.

Fig. 3a displays the electron addition spectrum at zero magnetic field for a dot of 500 nm lithographic diameter produced on wafer A. For gate biases below −500 mV, the quantum dot is empty. Remarkably, some of the peaks shown are of double height indicating the tunneling of two electrons in the dot at the same gate voltage. Altogether, we can distinctly resolve the first 600 electron additions into this dot. The gate voltage scale can be directly converted to an energy scale $\Delta E - \alpha \Delta Vg$ with the lever arm α, 0.50 for wafer A.

The magnetic field evolution of a portion of the electron addition spectrum is shown in Fig. 4a. The gray scale map displays the first 150 additions, with capacitance peaks visible as black traces. Examination of the bottom of Fig. 4a shows that the first 7 electrons enter the dot at widely spaced voltages. They may enter into a single potential minimum or minima spaced closely enough that the Coulomb repulsion between the sites is sufficient to keep the peaks widely spaced. Beyond the 7th electron trace, something extraordinary occurs. Three electrons enter the dot in very rapid succession in gate voltage over the full range of magnetic fields. The next two electrons also join in a bunch (pair). For higher N, other bunches can be seen. We note that the experiment shows no hysteretic effects. The bunching
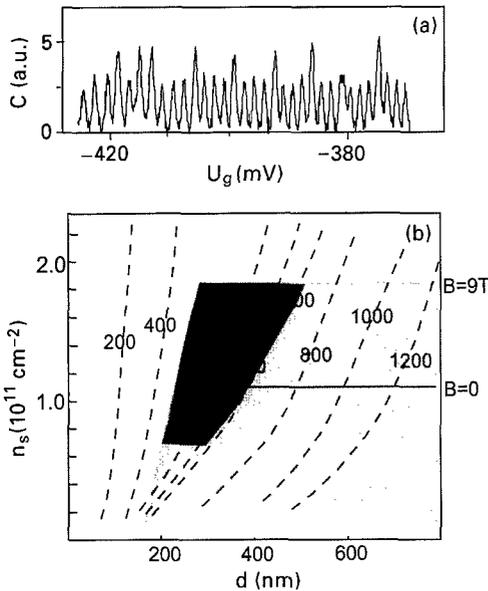
Fig. 3. (a) Quantum dot capacitance as a function of gate voltage. Each peak denotes the appearance of an electron in the dot. Double height peaks indicate the addition of two electrons. T : − 0.3 K. (b) Experimental phase diagram for $B = 9$ T. Dashed curves show the variation of lateral diameter d of electron puddles with the electron density for different quantum dots. Shaded area denotes the range of bunching; dark shading indicates the range of periodic bunches. The bunching ceases at an electron density which increases with magnetic field (see text). The density limit at $B = 0$ is demarcated by a horizontal line.

is a phenomenon that occurs with the dot in *equilibrium* with its surroundings.

After about 40 electrons are added to the dot, the bunching develops into a *periodic* pattern, with one bunch appearing for each 4–6 electrons added to the dot. As N is increased beyond about 80, the bunching ceases for zero magnetic field. Instead, the electron additions occur with nearly perfect periodicity, as is typical of CB. However, for nonzero magnetic field strengths, the bunching phenomenon returns. Bunches again occur periodically in gate voltage, and the period is about the same as that for the zero field bunches. A zoom-in of this behavior is shown in Fig. 4b. The onset of bunching shifts to larger magnetic fields with increasing N, and the bunches are no longer observable at fields up to 13 T for more than about 200 electrons in the dot.

The behavior of each electron trace can be described roughly as follows. The magnetic field at which all
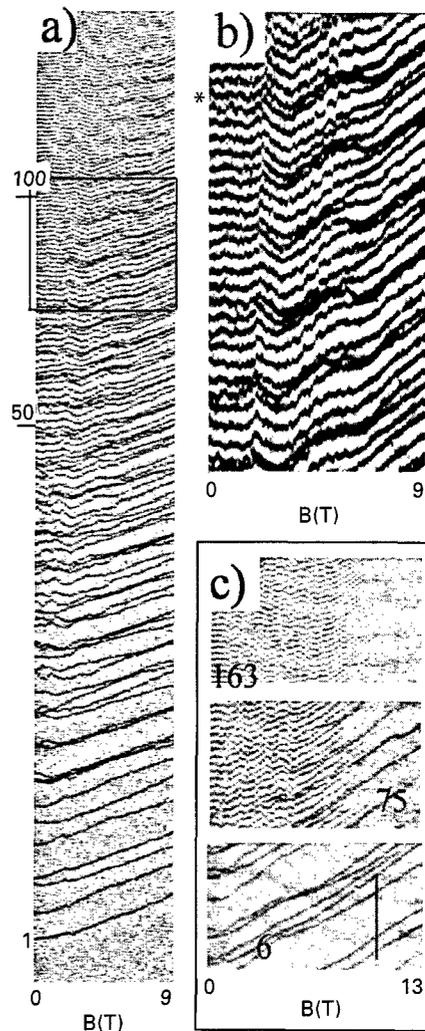
Fig. 4. Gray-scale image of the measured capacitance. Black denotes capacitance peaks. Electron occupancies are indicated as numbers. (a) Vertical axis – gate voltage ranging from 2511 (bottom) to 2328 mV (top). $T = 0.3$ K. (b) Zoom-in of spectrum surrounded by box in (a). (c) Segments of the addition spectrum measured after thermally cycling the dot to room temperature. $T = 50$ mK. Vertical bar corresponds to energy change of 5 meV (common for all images in (c)).

electrons fall into the lowest Landau level, $v = 2$, can be readily identified as a maximum in the traces at around $B = 2$ T [6]. As in two-dimensional systems the chemical potential peaks just as higher Landau levels depopulate completely. Jumps in the traces at higher magnetic fields, where both spin levels of the

lowest Landau level are filled, are usually interpreted as single electron spin-flips [15]. The flatness of the traces around $B = 6$ T demarcates total spin polarization of the dot. We refer to this range as the vicinity of $v = 1$. For higher fields, the traces rise nearly linearly with magnetic field.

A bunched pair of traces in Fig. 4b is marked with an asterisk *. These traces are fairly representative of all of the other traces which appear as electron pairs. Starting at some nonzero magnetic field the two traces are seen to stick together but then they split as the field approaches that which yields $v = 1$. Passing through $v = 1$, the lower trace of the bunched pair splits from the trace above it, only to join with the trace below it.

The bunching phenomenon is reflected in the rate at which the electrons tunnel into the dot. At zero magnetic field, the rate of electron tunneling between the n+ substrate and the quantum well is about 5 MHz. Measurements at a much lower frequency of $f = 200$ KHz are only sensitive to the tunneling resistance if the tunneling is strongly suppressed by electrons correlations within the dot [4,6]. At very low temperatures ($T < 0.1$ K) the tunneling rate drops substantially in particular regions of magnetic field and electron occupancy.

Fig. 4c shows a measurement of the addition spectrum of the same dot at base temperature $T = 30$ mK after thermal cycling up to room temperature. For low N, shown on bottom part of Fig. 4c, contrast in all electron traces is the same over the entire range of magnetic field, indicating that the electron tunneling rate is much larger than the measurement frequency. The middle segment of Fig. 4c displays the capacitance spectrum in a range of larger N (75–95 electrons in the dot). Notice here that some of the traces extinguish as magnetic field increases. As the peaks diminish in strength, the phase of the electron tunneling signal lags relative to the AC excitation [4]. This detectable decay of the tunneling rates begins in the vicinity of $v = 1$, for sufficiently large number of electrons in the dot.

The only traces observable at the highest magnetic field of $B = 13$ T in Fig. 4c extend from paired traces. Examination of the intensity and phase of these unextinguished traces shows that they typically result from only a single electron rather than two electrons tunneling. We note that the DC bias in the experiment is adjusted very slowly so that the electron occupancy in the dot changes even though peaks are not seen in the capacitance experiment. Finally, at higher N (Fig. 4c, upper part), the bunching disappears, and all traces extinguish equally.

The boundary for the onset of the bunching is remarkably similar for all dots in which bunches are observed, regardless of their size. This boundary moves to higher magnetic fields as the average electron density (note, *not N*) in the dot is increased roughly according to the linear relation

$$n_{\text{onset}} = (1.1 + 0.08 \times B\{\text{Tesla}\}) \times 10^{11}\,\text{cm}^{-2}. \quad (1)$$

Our findings are summarized on an approximate phase diagram shown in Fig. 3b. Each dashed curve represents the evolution of one measured quantum dot sample: as we increase the electron density n_s within the dot, the lateral diameter of electron pool d also grows. d is determined from the average spacing between addition peaks using a simple parallel plate capacitor approximation.

The bunching is absent *either* for very small dots at arbitrary electron density [4,6,7] *or* at sufficiently large electron density in arbitrary large dots (non-shaded area). The nearly periodic bunching (pairing) pattern is observed for dots created with lithographic diameters of about 0.5 μm (dark shaded). The details of the random bunching pattern at small electron numbers vary with thermal cycling of the sample to room temperature. In sharp contrast, the *periodic* bunching behavior remains qualitatively unchanged. The same 5 electron period is consistently detected for different thermal cycling and different samples. For larger dots the bunching still occurs, but the bunches appear to occur randomly with gate voltage rather than periodically (shaded).

5. Discussion

We believe that pairs of electrons in the quantum dot observed the large dots are a special case of the bunches in the regime of electrons strongly localized within a large (1 μm lithographic diameter) dot. In dots of similar size, we have seen more examples of bunches with the traces of two and sometimes three electrons that exactly overlap over a range of magnetic fields. In general, paired traces from dots with smaller lithographic diameters do not coincide exactly.

The phase diagram shown in Fig. 3b suggests that the bunches are intrinsically associated with electron localization within the quantum dot. Small dots likely consisting of one electron puddle do not display the bunching effect, while the effect appears in larger dots at low densities when distinct electron puddles may exist. In the case of large dots whose lateral size significantly exceeds the effective screening length, the direct Coulomb repulsion between different electron droplets may be strongly suppressed. The addition of one electron in one region may not inhibit the addition of a second electron in a remote location. Such localization is either caused by fluctuating potential or may arise intrinsically within a single dot due to interactions.

We speculate that the latter phenomenon gives rise to the periodic bunches. The reproducibility of the periodic bunching pattern in several different dots and upon different thermal cycling of the same dot cannot be ascribed to a peculiarity of the disorder potential. Taking into account the nearly cylindrical symmetry of the confining potential, the most plausible scenario for spatial segregation of electrons is the separation of an annulus of electrons at the circumference of the dot. To examine this idea, we use the results of classical modeling of the electron distribution within the dot [16,17]. Such a model is justified in the limit of infinite magnetic field. Electrons are then considered as classical point charges that form a nearly triangular lattice with significant lattice deformation only at the dot edges. The sequence of the electron entrances into a classical dot can be calculated. Levitov [17] demonstrated recently that for a classical dot containing 50–150 electrons, four electrons enter the center of the dot in succession, and the fifth enters the outer row (circumference). Our recent experiments on dots with added control gates strongly corroborate this view [18].

The idea that one of the bunched electrons appears at the edge of the electron droplet is consistent with the contrast observed in the tunneling rate. The drop of the tunneling rate for $v < 1$ illustrated in Fig. 4c can be considered as a special case of the Coulomb gap observed for larger systems [19–22]. The origin of the tunneling suppression can be understood semiclassically. The tunneling process suddenly adds one more electron into a dot. Until the system relaxes to its new ground state, the tunneling process is not finished. Hence, the effective tunnel barrier depends on the disturbance of the density distribution and its relaxation rate. The higher tunneling rate observed for one trace in each bunch can be explained if that electron is introduced into the edge of the dot. That electron has fewer and more distant neighbors compared with an electron introduced into the bulk of the dot.

The above model may explain the periodicity seen in our data and its increasing prominence at large magnetic field. However, it still does not explain the *pairing*. Studies of double dot systems (albeit in which each dot already contains many electrons) show that spatial segregation does not directly lead to paired electron additions [23]. In fact, the residual Coulomb interaction between dots inhibits pairing even when the two dots are tuned to individually add electrons at the same gate bias. However, significant differences exist between these dots and ours. For example, in the range for pairing, our dots contain a much lower electron density.

What can compete with the usually dominant Coulomb addition energy to disturb the addition spectrum so profoundly? Hartree–Fock calculations demonstrate [24] that exchange can mediate a local attraction between electrons, tending to keep the system compact. The $v = 1$ state is believed to be fully spin-polarized, and exchange maintains $v = 1$ as the lowest energy state of the system over a range of magnetic fields [24,25]. The switching of the bunches at $v = 1$ (Fig. 4b) appears to effectively broaden this range for some of the traces suggesting the involvement of the exchange interaction in bunch formation.

References

[1] R.H. Silsbee, R.C. Ashoori, Phys. Rev. Lett. 64 (16) (1990) 1991.

[2] H. van Houten, C.W.J. Beenakker, Phys. Rev. Lett. 63 (17) (1989) 1893.

[3] P.L. McEuen, E.B. Foxman, U. Meirav et al., Phys. Rev. Lett. 66 (1991) 1926.

[4] R.C. Ashoori, H.L. Stormer, J.S. Weiner et al., Phys. Rev. Lett. 71 (1993) 613.

[5] M.A. Kastner, Physics Today 46 (1) (1993) 24.

[6] R.C. Ashoori, Nature 379 (1996) 413.

[7] S. Tarucha, D.G. Austing, T. Honda et al., Phys. Rev. Lett. 77 (17) (1996) 3613.

[8] R.C. Ashoori, H.L. Stormer, J.S. Weiner et al., Phys. Rev. Lett. 68 (20) (1992) 3088.

[9] R.C. Ashoori, H.L. Stormer, J.S. Weiner et al., Physica B 189 (1993) 117.

[10] A.T. Johnson, L.P. Kouwenhoven, W. de Jong et al., Phys. Rev. Lett. 69 (10) (1992) 1592.

[11] R.C. Ashoori, H.L. Stormer, J.S. Weiner et al., Phys. Rev. Lett. 71 (1993) 613.

[12] P.W. Anderson, Phys. Rev. Lett. 34 (1975) 953.

[13] Y. Wan, G. Ortiz, P. Phillips, Phys. Rev. Lett. 75 (1995) 2879.

[14] M.E. Raikh, L.I. Glazman, L.E. Zhukov, Phys. Rev. Lett. 77 (7) (1996) 1354.

[15] P.L. McEuen, E.B. Foxman, Jari Kinaret et al., Phys. Rev. B 45 (1992) 11 419.

[16] V.M. Bedanov, F.M. Peeters, Phys. Rev. B 49 (4) (1994) 2667.

[17] L. Levitov, private communication.

[18] N.B. Zhitenev et al., to be published.

[19] R.C. Ashoori, J.A. Lebens, N.P. Bigelow et al., Phys. Rev. Lett. 64 (6) (1990) 681.

[20] R.C. Ashoori, J.A. Lebens, N.P. Bigelow et al., Phys. Rev. B 48 (1993) 4616.

[21] H.B. Chan, P.I. Glicofridis, R.C. Ashoori et al., Phys. Rev. Lett. 79 (15) (1997) 2867.

[22] J.P. Eisenstein, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 69 (1992) 3804.

[23] F.R. Waugh, M.J. Berry, D.J. Mar et al., Phys. Rev. Lett. 75 (4) (1995) 705.

[24] C. de C. Chamon, X.G. Wen, Phys. Rev. B 49 (12) (1994) 8227.

[25] A.H. MacDonald, S.R. Eric Yang, M.D. Johnson, Aust. J. Phys. 46 (1993) 345.

Single-electron transistor probes two-dimensional electron system in high magnetic fields

J. Weis*, Y.Y. Wei, K. v. Klitzing

Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, D-70569 Stuttgart, Germany

Abstract

Strips of metal-like and insulator-like behavior, the so-called compressible and incompressible strips, are expected to develop with magnetic field within the depletion region at the edge of a two-dimensional electron system (2DES). Here, we summarize our recent experimental results in mapping out these strips by using a metallic single-electron transistor close to a gate electrode which redefines the edge of the 2DES by electrostatic depletion. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 73.40.Hm; 71.30.+h; 07.50.Ls
Keywords: Quantum Hall effect; Single-electron transistor; Chemical potential; Metal–insulator transition

1. Introduction

For the quantum Hall effect [1], the edges of the two-dimensional electron system (2DES) play an important role [2]. In the last years, different types of experiments have been performed on 2DES in high magnetic fields to investigate the potential distribution [3,4], the regions of dissipation [5], the current distribution [6,7] and the edges [8–11]. Theoretical works [12–15] predict strips of different electrical properties within the depletion region at the edge. In the experiments presented here, a single-electron transistor (SET), made of metal, is used as a local electrometer to investigate the bulk and the edge of the 2DES *without* a bias current. In contrast to a recently presented

* Corresponding author. Fax: +49-711-689-1010; e-mail: weis@klizix.mpi-stuttgart.mpg.de.

approach [16] of scanning the SET on a tip, here the SET is deposited on top of the heterostructure containing the 2DES and the edge is shifted towards the SET by electrostatic depletion with a gate electrode.

2. Edge strips

At the edges of the 2DES, within the depletion region the electron concentration n_s changes from zero to its bulk value over a typical distance of $1\,\mu\text{m}$. Increasing the electron concentration from the edge to the bulk requires locally populating more and more electronic states. In thermodynamical equilibrium, the electron concentration profile $n_s(r)$ is self-consistently calculated [14] within a Thomas–Fermi approximation (1) by the constraint that the electrochemical potential μ_{elch}, i.e. the local chemical potential $\mu_{\text{ch}}(r)$ plus the

Fig. 1. Sketches of the density of states $D(\varepsilon)$ versus single-particle energy ε, the chemical potential μ_{ch} versus sheet electron concentration n_s, and the electron concentration profile at the edge of an idealized 2DES with and without magnetic field B.

local electrostatic energy $-e\phi(\mathbf{r})$, is constant within the whole 2DES,

$$\mu_{elch} = \mu_{ch}(\mathbf{r}) - e\phi(\mathbf{r}) = \text{const}, \qquad (1)$$

and (2) by the Poisson equation which relates the local electrostatic potential $\phi(\mathbf{r})$ to the local electron concentration $n_s(\mathbf{r})$ by taking into account the adequate boundary conditions for $\phi(\mathbf{r})$ which are due to the sample.

Without applied magnetic field, for an idealized infinite extended 2DES, the density of state $D(\varepsilon)$ is constant versus the single-particle energy ε (see Fig. 1). Therefore, the chemical potential μ_{ch} increases linearly with the electron concentration n_s. Due to the constraints (1) and (2) discussed above, the profile $n_s(\mathbf{r})$ changes smoothly from the edge to the bulk as sketched in Fig. 1. With magnetic field applied perpendicular to the plane of the 2DES, the profile is different. The magnetic field forces the electrons on cyclotron orbits leading to a quantisation of the energy spectrum of the 2DES into Landau (and spin) levels. The density of states becomes discrete (see sketch in Fig. 1) and the chemical potential increases step-like whenever an energy level is completely filled. Therefore, whenever a level is occupied, instead of paying a large amount of chemical energy for increasing the electron concentration further towards the bulk, it is energetically favorable only to pay a small amount of

electrostatic energy, i.e. to keep the electron concentration constant for a region of finite width. Metal-like (constant electrostatic potential) and insulator-like (constant electron concentration) strips – so-called compressible and incompressible strips respectively – are formed in the depletion region of the 2DES [13–15]. The widths and positions of these strips depend on the imprinted potential profile at the edge and on the quantisation energy which increases with magnetic field.

3. The sample

To resolve the strip-like structure at the edge of a 2DES in high magnetic fields, a metallic single-electron transistor (SET) is deposited on top of an $Al_{0.33}Ga_{0.67}As$-GaAs heterostructure with the 2DES about 90 nm below the surface (the sample preparation is described in Ref. [18]; bulk electron concentration $n_s^{bulk} = 2.2 \times 10^{11}$ cm^{-2}, electron mobility 1.3×10^6 cm^2/Vs at the temperature $T = 4.2$ K). Fig. 2 shows a sketch of the arrangement and a scanning electron microscope image of the SET. The SET consists of a small aluminum island – 0.13 μm wide and 0.9 μm long – which is coupled by aluminum oxide tunneling barriers to the aluminum source and drain electrodes. Due to the small size of the electronic island and small area of the tunnel junctions (about 0.13 μm by 0.13 μm), the total capacitance C_Σ of the island is small. Adding an electron to the island requires the Coulomb charging energy $e^2/2C_\Sigma$ (about 0.07 meV here), which acts like an energy barrier and blocks electrical transport through the island at the temperature $T < 100$ mK used in our experiments. Alloyed ohmic contacts to the 2DES allow to use the 2DES as a gate electrode for the SET island. By changing V_{2DES}, the electrostatic potential of the island is shifted and the energy for adding an electron is lowered. The island is charged by another additional electron whenever the voltage is increased by $\Delta V_{2DES}^{pp} = e/C_{2DES}$ where C_{2DES} is the 2DES – island capacitance. As shown in Fig. 3, a sequence of conductance peaks is observed with the period ΔV_{2DES}^{pp} – the so-called Coulomb blockade oscillations (CBOs) [17].
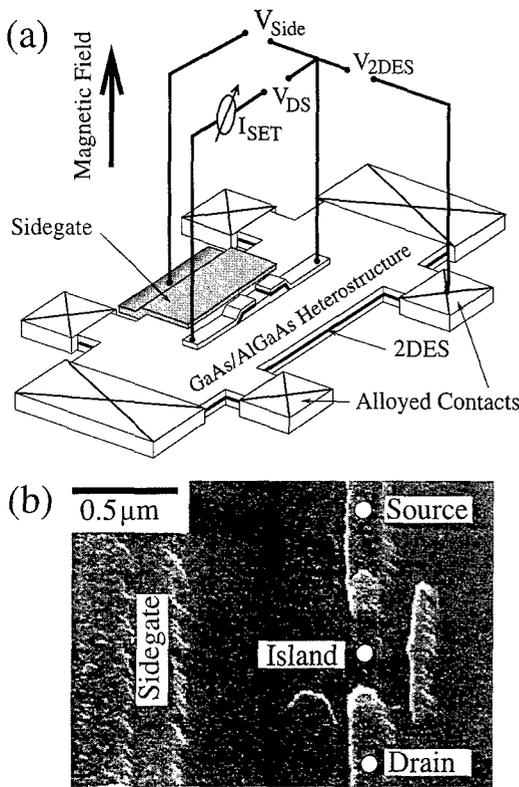
Fig. 2. (a) Sketch of the experimental setup and sample (not on scale): A SET is deposited on top of a Hall bar which has been etched into a GaAs/AlGaAs heterostructure containing the 2DES. Alloyed ohmic contacts give electrical connection to the 2DES. (b) SEM picture of the SET and the sidegate. All lengths along the vertical axis appear reduced by the factor $1/\sqrt{2}$. The 'shadows' of the source and drain leads and the isolated island are due to the fabrication process.

4. Probing the local chemical potential

By applying a magnetic field B perpendicular to the plane of the 2DES, the Coulomb blockade oscillations shift along the axis of the externally applied voltage V_{2DES}, as shown in Fig. 3. These shifts reflect the variation of the chemical potential of the 2DES at constant electron concentration with increasing magnetic field [18]. This becomes clear when taking into account the intrinsic contact voltage V_C which occurs in series to the external applied voltage V_{2DES}. In general, if two electrical materials of different chemical potential (workfunction) are in direct or indirect electrical contact, electrons are transferred from one material to the
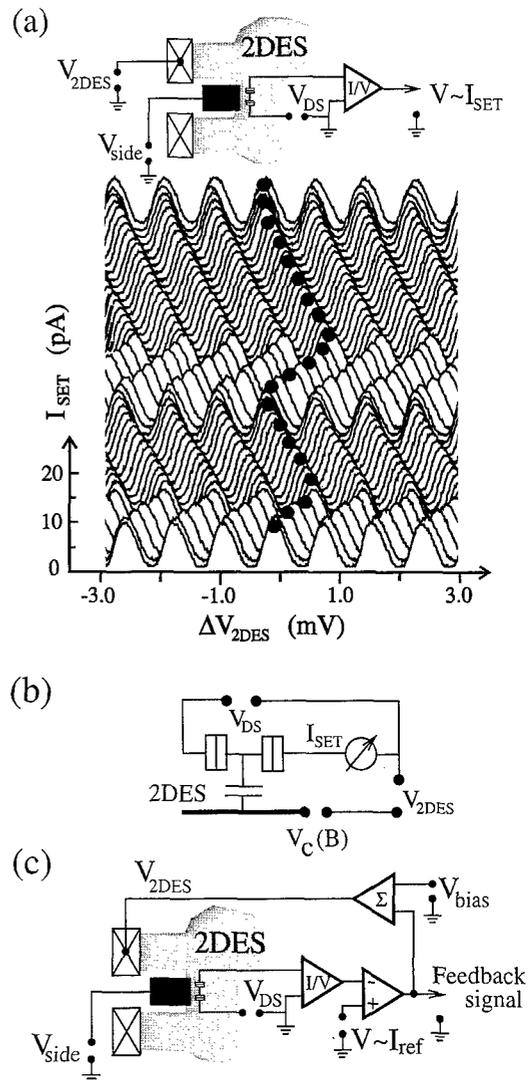
Fig. 3. (a) I_{SET} versus V_{2DES} for different magnetic field values B from 1 T (bottom curve) to 1.5 T (top curve) in steps of 0.01 T. For better presentation, offsets are added to the measured curves. Inset: circuit setup. (b) Modified circuit diagram taking into account the contact voltage V_C. (c) Feedback circuit which keeps I_{SET} constant.

other material in order to get thermodynamical equilibrium. Both materials are electrically charged, i.e. an electrostatic potential difference (the contact voltage) builds up which is given by the difference of both chemical potentials (see relation (1)). By changing the homogeneously applied magnetic field, the electronic states of both materials are affected causing changes

in both chemical potentials, and therefore a variation in the contact voltage between both materials. Applied to our system here, the change in the contact voltage is determined by the variation in the chemical potential of the SET leads and the variations in the chemical potential of the 2DES *locally* below the SET island [18]:

$$ -e \cdot \frac{\mathrm{d}V_C}{\mathrm{d}B} = \frac{\mathrm{d}\mu_{\mathrm{ch}}(\mathrm{SET})}{\mathrm{d}B} - \frac{\mathrm{d}\mu_{\mathrm{ch}}(\mathrm{2DES})}{\mathrm{d}B}. \qquad (2) $$

Since the effect on $\mu_{\mathrm{ch}}(\mathrm{SET})$ is negligible small in comparison to the effect on $\mu_{\mathrm{ch}}(\mathrm{2DES})$ – the aluminum is not in the superconducting state at such high magnetic fields – the contact voltage variations reflect mostly the change of $\mu_{\mathrm{ch}}(\mathrm{2DES})$. Since V_C contributes to the electrostatic potential difference between the 2DES and the SET island, which determines resonance for electrical transport through the SET island,[1] the shift of the CBOs shown in Fig. 3 reflects directly $-\mathrm{d}V_C/\mathrm{d}B$. Therefore, to keep the current I_{SET} through the SET island constant, the variations of the contact voltage V_C have to be compensated by the externally applied voltage: $\mathrm{d}V_{\mathrm{2DES}} = -\mathrm{d}V_C$. This can be done automatically by using the feedback circuit shown in Fig. 3c.

It was shown [19] that in the low magnetic field regime, the depopulation of Landau levels with increasing magnetic field is nicely observed. This is different at high magnetic fields in the regime of well-developed quantum Hall plateaus where the Fermi level is located in the mobility gap, i.e. between two Landau levels. The 2DES no longer works as a gate electrode for the SET. The 2DES below the SET island is electrically decoupled from the edge where the voltage V_{2DES} is applied. Instead, relaxation processes and charge fluctuations versus time become visible in the vicinity of the SET island [20]. The 2DES in the bulk has lost its good conductivity and the screening properties of a metallic layer. This interpretation is verified by changing the voltage applied to the metal electrode on the backside of the heterostructure. Whenever the 2DES, which lies between the backgate and the SET, shows metal-like behavior in the bulk, Coulomb blockade oscillations as a function

of *backgate voltage* are not observable. But CBOs as a function of the backgate voltage become visible *within* the quantum Hall regime which demonstrates that under this condition the 2DES behaves like an insulator and cannot screen the voltage variations of the backgate for the SET island [20].

5. Avoiding the inhomogenity in the electron concentration

The presence of the SET on the surface of the heterostructure usually affects the electron concentration in the 2DES directly below the SET. This might be due to the electrostatic potential difference (contact voltage) between the SET and the 2DES which already exists at $V_{\mathrm{2DES}} = 0$ caused by the different chemical potentials. In addition, local stress, surface charges or effects due to the fabrication process might have created inhomogenities in the electron concentration in the vicinity of the SET. To minimize an inhomogenity in the electron concentration of the 2DES, which otherwise would cause undesired effects, a bias voltage V_{bias} (see Fig. 3c) is applied to the 2DES allowing us to adjust the electron concentration below the SET. To determine the correct value for V_{bias}, the following two methods have been used: In the first method, the local electron concentration is obtained by measuring the local chemical potential variation with the SET which is then compared with the overall electron concentration obtained from Hall measurements. More reliable – because local – seems the second method: The variation of the contact voltage is measured for different bias voltages V_{bias}. As sketched in Fig. 4a, an inhomogeneous electron concentration leads to Landau level depopulation at different magnetic field values and therefore locally to different variations in the electrostatic potential with increasing magnetic field. All these variations contribute to the electrostatic potential of the SET island weighted by their capacitive coupling. Therefore, the feedback signal to keep the current through the SET island constant, contains all these contributions, i.e. the signal is smeared out in comparison to the signal obtained from the Landau level depopulation in the case of a homogeneous n_{s}. Fig. 4b shows the measured feedback signal obtained for different bias voltages. The sharpest transitions with magnetic field are obtained for $V_{\mathrm{bias}} = 10\,\mathrm{mV}$

---

[1] V_{2DES} determines the difference of electrochemical potentials and not per se the difference in the electrostatic potentials between SET and 2DES.
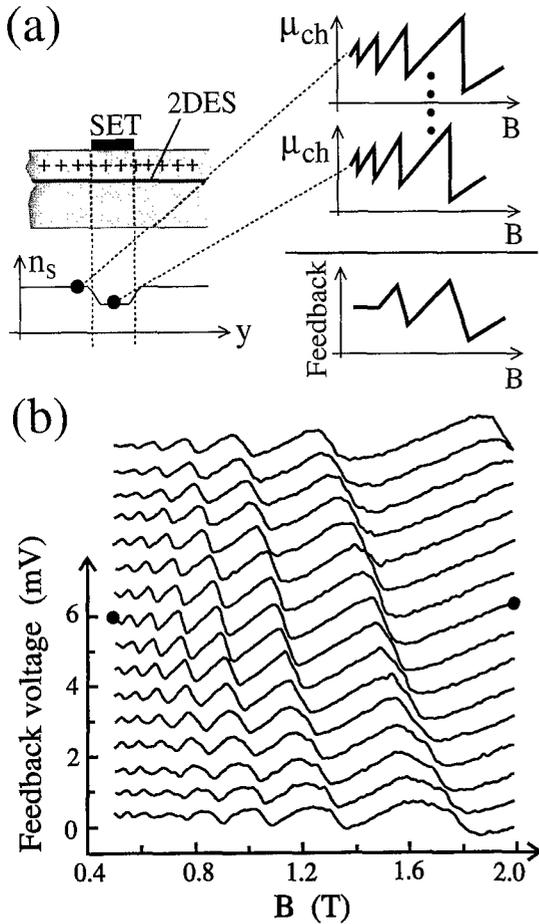
Fig. 4. (a) Sketch of what is happening on the feedback signal which keeps I_{SET} constant (see Fig. 3c) versus magnetic field B in the case of an inhomogeneous electron concentration n_s in the vicinity of the SET. (b) Feedback signal versus magnetic field B for different bias voltages V_{bias} from 80 mV (top curve) to −70 mV (bottom curve) in steps of 10 mV. The curves are plotted with offsets. The dots mark the curve for $V_{bias} = 10$ mV.

which is also obtained by the first method. Therefore this value is chosen as the optimum for countering an inhomogenous n_s.

6. Feedback signal at the edge

By electrostatically depleting the 2DES below the gate electrode which is shown in Fig. 2, the edge of the 2DES is redefined at a distance of about 0.9 μm from the SET island ($V_{side} \approx -0.26$ V). By applying
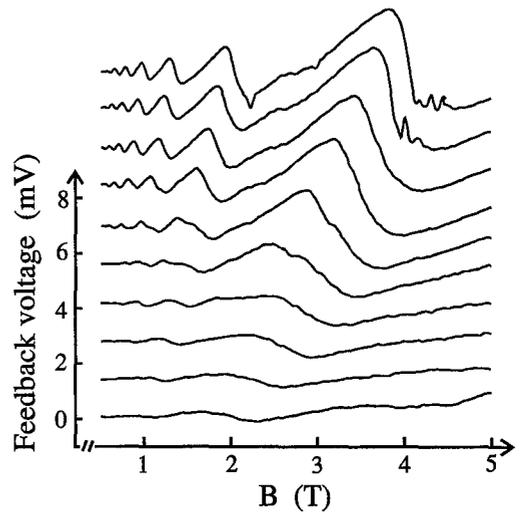
Fig. 5. Feedback signal versus magnetic field B for $V_{side} = -1$ V (top) to $V_{side} = -5$ V (bottom) in steps of 0.5 V. The curves for different V_{side} are plotted with offsets.

a more negative sidegate voltage V_{side}, the edge of the 2DES is shifted closer to the SET island. Fig. 5 shows the measured feedback signal versus magnetic field for different gate voltages V_{side}. Shifting the edge closer towards the SET island, the transitions due to Landau level depopulation scale towards lower magnetic field values reflecting the decrease in the electron concentration below the SET island due to V_{side}. In addition, the feedback signal smears out with more negative V_{side}, which indicates (see discussion to Fig. 4a) that the gradient in the electron concentration closer to the edge becomes steeper.

7. Resolving the edge strips

It was already mentioned that within the quantum Hall regime of the 2DES the SET cannot be controlled by the voltage V_{2DES} applied at the alloyed contact at the edge of the 2DES. This is also true for the case that the edge is redefined at 0.9 μm distance to the SET island: Fig. 6a shows the current I_{SET} versus V_{2DES} for different values of V_{side} at $B = 9$ T, i.e. at Landau level filling factor $v = 1$ ($v = hn_s^{bulk}/eB$).

(a)

(b)
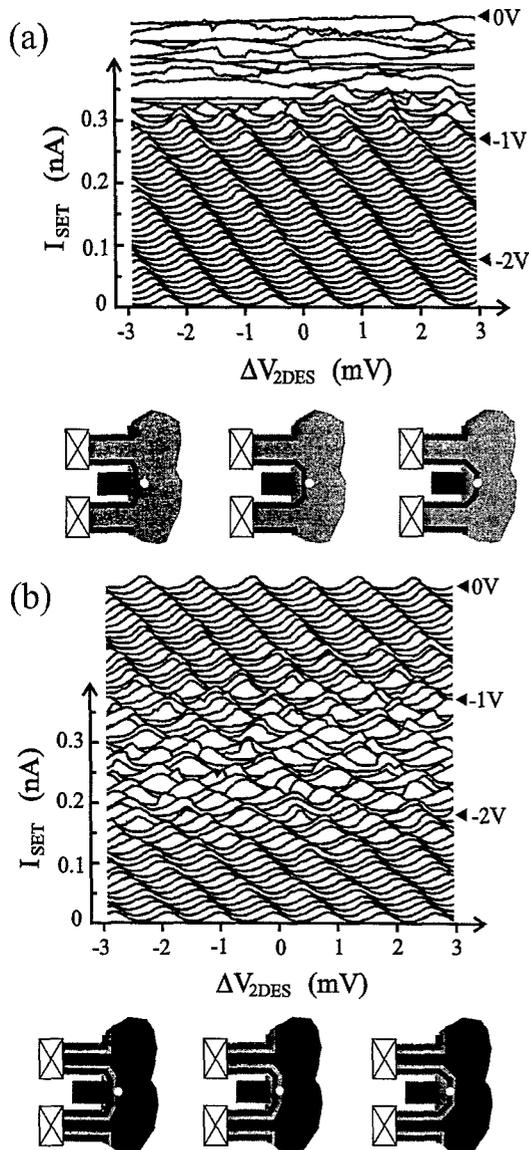
Fig. 6. (a) I_{SET} versus V_{2DES} measured for different sidegate voltage V_{side} varied between $0\,V$ and $-3\,V$ in the step of $\Delta V_{side} = 0.05\,V$ at $B = 9\,T$. For the presentation, offsets are added to the measured curves. Inset: sketch of the edge shifting closer to the SET (from left to right). Dark regions represent compressible strips, light regions incompressible strips. (b) Same as (a), but for $B = 7.5\,T$.

For $V_{side} > -0.7\,V$, only charge fluctuations are visible which are not screened due to the lack of good conductivity in the 2DES below the SET island. This is different for $V_{side} < -1\,V$ where well-developed
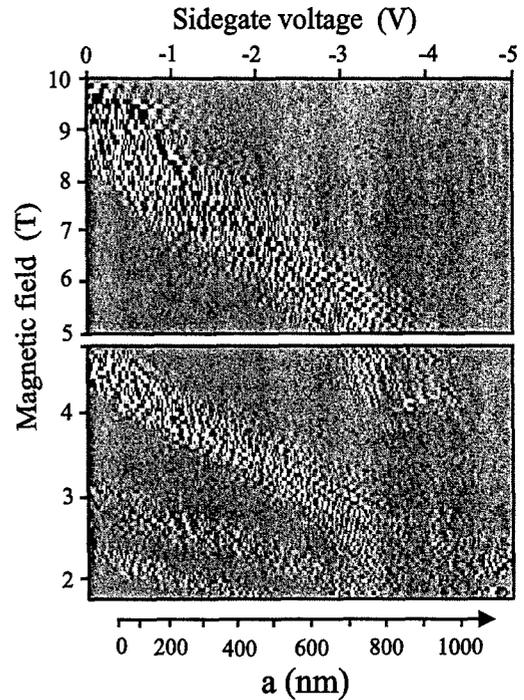
Fig. 7. Grey-scale plot of the fluctuations observed in the SET current as a function of the sidegate voltage and of the magnetic field. With more negative sidegate voltage, the edge is shifted closer to the SET.

CBOs are observable. The inset in Fig. 6a shows the sketch of the edge for both situations: At less negative depletion voltage, the SET is situated on top of the insulating region of the 2DES. At more negative depletion voltage, the metal-like region at the edge is shifted below the SET island. But what happens at a filling factor $1 < \nu < 2$? As shown in Fig. 6b, the bulk behaves metal-like and the SET island is well controlled by V_{2DES} for $V_{side} > -0.8\,V$. Then for $-0.8\,V > V_{side} > -2\,V$, charge fluctuations are visible. For $V_{side} < -2\,V$, again well-pronounced and stable CBOs are observable. The interpretation is sketched in Fig. 6b: With more negative V_{side} an insulator-like (incompressible) strip which does not screen charge fluctuations is pushed under the SET island, followed by the outer metal-like (compressible) strip.

Instead of measuring the CBOs, the feedback circuit (Fig. 3c) can be used to monitor the regimes of charge fluctuations as a function of sidegate voltage and magnetic field [22]. Fluctuations in the feedback signal shown in Fig. 7 are interpreted as in-

compressible (insulating) regions at the position of the SET. At low magnetic fields, several strips of different screening properties become visible as expected from the compressible–incompressible strip model [14,13]. Since the sidegate voltage moves the electrostatically defined edge, V_{side} can be directly used as a measure of the distance a of the 2DES's edge to the sidegate's edge. Simple model calculations [21] show that the relation is almost linear (see bottom of Fig. 7).

8. Conclusion

The chemical potential variations of the 2DES is measured at low magnetic field directly by a metal SET. At high magnetic field, i.e. low Landau level filling factors, charge fluctuations become visible in the quantum Hall regime where the bulk of the 2DES is decoupled from the edge of the 2DES. By using a sidegate at a distance of about 0.9 µm from the SET island, fingerprints of strips with different screening properties are clearly detected. The number of resolved strips and their evolution is qualitatively in agreement with the theoretical predictions.

References

[1] K.v. Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. 45 (1980) 494.

[2] See the reviews: R.J. Haug, Semicond. Sci. Technol. 8 (1993) 131; K.v. Klitzing, Physica B 184 (1993) 1; K.v. Klitzing, Physica B 204 (1995) 111.

[3] G. Ebert, K.v. Klitzing, G. Weimann, J. Phys. C 18 (1985) L257; H.Z. Zheng, D.C. Tui, A.M. Chang, Phys. Rev. B 32 (1985) 5506; E.K. Sichel, H.H. Sample, J.P. Salerno, Phys. Rev. B 32 (1985) 6975; Ch. Simon, B.B. Goldberg, F.F. Fang, M.K. Thomas, S. Wright, Phys. Rev. B 33 (1986) 1190.

[4] R.J.F. Haren, F.A.P. Blom, J.H. Wolters, Phys. Rev. Lett. 74 (1995) 1198; R. Knott, W. Dietsche, K. von Klitzing, K. Ploog, Semicond. Sci. Technol. 10 (1995) 117.

[5] U. Klass, W. Dietsche, K.v. Klitzing, K. Ploog, Surf. Sci. 263 (1992) 97.

[6] S.W. Hwang, D.C. Tsui, M. Shayegan, Phys. Rev. B 48 (1993) 8161.

[7] E. Yahel, D. Orgad, A. Palevski, H. Shtrikman, Phys. Rev. Lett. 76 (1996) 2149.

[8] A.J. Kent, D.J. McKitterick, L.J. Challis, P. Hawker, C.J. Mellor, M. Henini, Phys. Rev. Lett. 69 (1992) 1684; D.J. McKitterick, A. Shik, A.J. Kent, M. Henini, Phys. Rev. B 49 (1994) 2585.

[9] R. Merz, F. Keilmann, R.J. Haug, K. Ploog, Phys. Rev. Lett. 70 (1993) 651.

[10] S. Takaoka, K. Oto, K. Murase, K. Gamo, S. Nishi, Phys. Rev. Lett. 72 (1994) 3080; S. Takaoka, K. Oto, K. Murase, Int. J. of Modern Phys. B 11 (1997) 2593.

[11] N.B. Zhitenev, R.J. Haug, K.v. Klitzing, K. Eberl, Phys. Rev. Lett. 71 (1993) 2292; G. Ernst, N.B. Zhitenev, R.J. Haug, K.v. Klitzing, K. Eberl, Phys. Rev. Lett. 79 (1997) 3748.

[12] P.L. McEuen, E.B. Foxman, J. Kinaret, U. Meirav, M.A. Kastner, N.S. Wingreen, S.J. Wind, Phys. Rev. B 45 (1992) 11419.

[13] D.B. Chklovskii, B.I. Shklovskii, L.I. Glazman, Phys. Rev. B 46 (1992) 4026; D.B. Chklovskii, K.A. Matveev, B.I. Shklovskii, Phys. Rev. B 47 (1993) 12605.

[14] K. Lier, R.R. Gerhardts, Phys. Rev. B 50 (1994) 7757.

[15] J.H. Oh, R.R. Gerhardts, Phys. Rev. B 56 (1997) 13519.

[16] M.J. Yoo, T.A. Fulton, H.F. Hess, R.L. Willett, L.N. Dunkleberger, R.J. Chichester, L.N. Pfeiffer, K.W. West, Science 276 (1997) 579.

[17] See for review: H. Grabert, M.H. Devoret (Eds.), Single Charge Tunneling, NATO ASI Ser. B, vol. 294, Plenum Press, New York, 1992.

[18] Y.Y. Wei, J. Weis, K.v. Klitzing, K. Eberl, Appl. Phys. Lett. 71 (1997) 2514.

[19] Y.Y. Wei, J. Weis, K.v. Klitzing, K. Eberl, Physica E 1 (1997) 135.

[20] Y.Y. Wei, J. Weis, K.v. Klitzing, K. Eberl, Proc. 12th Int Conf. on Electronic Properties of Two-Dimensional Systems, Tokyo, 1997, Physica B 249–251 (1998) 496.

[21] I.A. Larkin, J.H. Davies, Phys. Rev. B 52 (1995) R5535.

[22] Y.Y. Wei, J. Weis, K.v. Klitzing, K. Eberl, Phys. Rev. Lett., to be published.

A new model for the transport regime of the integer quantum Hall effect: The role of bulk transport in the edge channel picture

Josef Oswald *

Institute of Physics, University of Leoben, Franz Josef Str. 18, A-8700 Leoben, Austria

Abstract

Based on a current balance between edge and bulk we obtain a modification of the Landauer–Büttiker formalism. The new aspect of this approach is an interpretation of edge channel (EC) backscattering in terms of a bulk current which couples the edges. This coupling is described by a novel backscattering parameter P, which is a function $P(\Delta v)$ of the Landau level filling. We show, that the most important features of transport can be modeled already without requesting a specific function for $P(\Delta v)$. In addition, a number of trends in R_{xx} and R_{xy}, from which most of them have been studied experimentally in the recent work of Shahar et al., can be reproduced by using a pure exponential function for $P(\Delta v)$. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Two-dimensional electronic system; Quantum Hall effect; Edge channels; Backscattering

1. Introduction

Even more than 15 y after the discovery of the integer quantum Hall effect (IQHE) in two-dimensional electronic systems [1], the QHE is still one major topic of every conference on low dimensional electronic systems. Particularly the nature of the transitions between adjacent QH plateaus is still a controversial question and a quantitative modeling of the complete transport regime of the IQHE has not been given so far. While the quantized values of the Hall resistance are well described by the edge channel (EC) picture [2], it is widely believed that the EC-picture is insufficient to describe also the transport regime between

* Tel.: +43 3842 402 265; fax: +43 3842 402 760; e-mail: oswald@unileoben.ac.at.

the IQHE plateaus. However, non-local experiments for the IQHE [3] as well as for the fractional QHE [4] have demonstrated, that in the transition regime between QH-plateaus EC- and bulk transport exist simultaneously. In this regime dissipative transport appears in the partly filled top Landau level (LL) while the contribution of the lower LLs is represented by dissipationless EC-transport. A similar picture in terms of a coexistence of edge and bulk transport has been used to explain the results of magnetotransport experiments in PbTe wide quantum wells. In that case the EC-system and the dissipative bulk system have been explained to result from electrons in different valleys of the many valley semiconductor PbTe [5].

It is widely accepted that in the plateau transition regime backscattering in the top LL enables dissipative bulk conduction if the Fermi level E_F is near the

center of the broadened LL. In the bulk region a transition to an insulating state occurs if E_F moves out of the center on either side [6,7]. In fact, this symmetry of the transport behavior with respect to the LL center makes ECs unlikely to be responsible for the characteristics of the inter plateau regime. The problem can be summarized as follows: By starting with E_F above a completely filled top LL, ECs are formed and the transport can be described by the EC-picture without backscattering. With E_F approaching the center of the top LL dissipation because of backscattering becomes possible and finally with E_F moving below the center of the top LL the associated pair of ECs disappears. Consequently, the contribution of the top LL should be different if E_F is above or below the center. In contrast, the experimental results of the plateau transitions indicate a symmetric behavior of the top LL like one expects for pure bulk transport.

However, there can be found one particular regime, where an asymmetric transport behavior in a LL is also obtained in the experiments, namely the regime of the Hall insulator (HI): Using QHE samples with not too low disorder, the HI regime is entered directly from the $v = 1$ integer QHE regime without observing the fractional QHE and recent results [8–10] are presently stimulating also the ongoing studies of the inter-plateau transitions. An analysis of the transport ranging from the HI to the adjacent QH liquid regime suggests the existence of a close relation between the transport behavior in the two regimes [10]: By defining a critical filling factor v_c it is possible to distinguish two regimes that are coupled by the relation $\rho_{xx}(\Delta v) = 1/\rho_{xx}(-\Delta v)$, where Δv is the filling factor relative to v_c. Another important experimental fact is the existence of a critical longitudinal resistivity ρ_{xx}^c, which appears at the transition point from the QH-liquid to the HI regime [8]. This critical point v_c is indicated by the crossing of the temperature dependent ρ_{xx} traces and the value of ρ_{xx}^c was found to be close to h/e^2. Using a tensor based analysis of the experimental data, Shahar et al. [11] have been able to extract the contribution of the top LL (referred to as ρ_{xx}^{top}) to the total ρ_{xx} in the transition regime between the 1st and 2nd QH-plateau. They found that ρ_{xx}^{top} shows the same behavior like ρ_{xx}^{ins} in the HI regime, namely a monotonous increase with increasing magnetic field without any peak-like behavior. It was possible to collapse all temperature dependent traces onto

each other by plotting ρ_{xx}^{ins} as well as ρ_{xx}^{top} with respect to $(v - v_c)T^{-\kappa}$ using the same $\kappa = 0.45$. Another experimental fact is that ρ_{xy} remains quantized on the $v = 1$ plateau also in the HI regime below the critical filling factor, while ρ_{xx}^{ins} already steeply rises. Furthermore, Shahar et al. demonstrated, that the conductivity components in the HI regime (σ_{xx}^{ins} and σ_{xy}^{ins}) as well as the extracted components for the top LL (σ_{xx}^{top} and σ_{xy}^{top}) for the $1 \rightarrow 2$ plateau transition fulfill a semicircle relation ($\sigma_{xx}^2 + \sigma_{xy}^2 \propto \sigma_{xy}$). Such a semicircle relation has been interpreted as being the consequence of a close relation between σ_{xx} and σ_{xy} [12].

2. Modeling of backscattering

Several attempts for modeling a four-terminal experiment with a discrete backscattering barrier or a disordered region between ideal conductors have been made already in the past [2,13–15]. As an example, for the case of a four-terminal arrangement Büttiker [2,13] obtains

$$R_{xx} = (h/e^2)[R/(NT)], \qquad (1)$$

where N is the number of channels, R and T are the reflection and transmission coefficients of the barrier. The Landauer–Büttiker (LB) formalism is in principle general and the transmitting channels are not necessarily edge channels. However, for the explanation of the quantized Hall resistance values Büttiker uses the EC-picture and therefore, the transmitting channels are ECs and N therefore has to be identified with the filling factor v.

Fig. 1 shows a typical configuration for studying the effect of EC-backscattering. By applying a gate voltage the two-dimensional electron gas (2DEG) can be partly depleted, which allows to achieve a controllable reflection of ECs. An extensive experimental study, in which a number of different gates are used, has been performed e.g. by Müller et al. [16]. All these cases had been well described within the framework of the LB formalism.

However, there are some problems associated with the application of the LB formalism to the plateau transitions of the QHE: It describes the scattering region in terms of global transmission and reflection coefficients between ideal (scattering free) conductors. In addition, the LB formalism does not care about where the
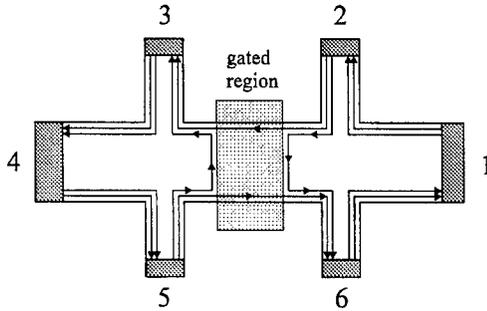
Fig. 1. Schematic sample configuration for an experimental study of EC-backscattering. The reflection (backscattering) is forced by a gated region between ideal conductors. The ungated parts are assumed to be in the plateau regime, where there is no backscattering. The edge channels are indicated by the arrows.

Fig. 2. Schematic sample configuration for representing backscattering by a dissipative current across a resistor R, which connects both edges. As compared to Fig. 1, the gate is replaced by an additional Hall contact pair with an ohmic resistor connected to it. Consequently, the backscattering process is no longer represented by the reflection of ECs. The edge channels are indicated by the arrows.

associated dissipation occurs. In the example of Fig. 1 the additional dissipation due to the EC reflection occurs directly at the Hall contacts. In contrast, in the real situation of the QHE the whole sample enters the regime of finite ρ_{xx} at once and the dissipation occurs directly in the bulk region of the sample. However, dissipation in the bulk implies automatically the existence of a dissipative current in the bulk, which must be related to the local bulk conductivity and hence to the particular bulk transport mechanism. The existence of such a dissipative bulk current is not addressed by the LB formalism. On this background the above referenced EC-approach together with the LB formalism leaves a missing link between the EC-picture and the local bulk transport properties and therefore, it seems not to be applicable to the plateau transitions of the QHE.

In the following, we are going to develop an alternative approach for EC-backscattering, which directly addresses the simultaneous presence of an edge current and a dissipative current in the bulk. We consider a simplified situation in terms of an equivalent circuit, which allows us to present a short version of the more extended treatment presented in Ref. [17]. The situation can be understood to be obtained from Fig. 1 by replacing the gate by an additional Hall contact pair. Dissipation can now be introduced by connecting a resistor across this additional contact pair (Fig. 2).

The sample itself is supposed to be in the plateau regime, which means that without the artificial (bulk) resistor R_b there is no backscattering in the system, which means that $R_{xx} = 0$ between any longitudinal

contact pairs at the same edge. The Hall voltage will be the same at all three Hall contact pairs, resulting in a Hall resistance of $R_{xy} = (1/v)(e^2/h)$, where v is the number of EC-pairs. Now we are going to consider the effect of a dissipative current I_b across the contact pair 7–8 and calculate the resulting longitudinal resistance R_{xx} between the outer voltage probes 2 and 3. In order to preserve current conservation, the current I_b must be supplied through the contact arms by the ECs. This means that there must exist a potential difference (Hall voltage) between incoming and outgoing ECs in the contact arms of the middle Hall contacts 7 and 8. Since the incoming ECs carry the potential of the previous contact 2 and the potential of contact 7 is transmitted to contact 3 by the outgoing EC, the Hall voltage U_{xy}^c in the contact arm 7 must appear directly as the longitudinal voltage drop U_{xx} between contact 2 and 3. The situation at the lower edge is similar and we obtain:

$$U_{xx} = U_{xy}^c = I_b R_{xy} = I R_{xx}. \tag{2}$$

In this way, we get a coupling between R_{xx} and R_{xy} by $R_{xx} = (I_b/I)R_{xy}$. We choose a dimensionless parameter $P = (I_b/I)$, which serves now as an alternative representation of backscattering as compared with the LB formalism:

$$R_{xx} = PR_{xy}. \tag{3}$$

For the case of a 2D system R_{xy} has to be represented by $R_{xy} = h/(ve^2)$. The ratio of the currents $P = (I_b/I)$ can be interpreted as the ratio of the probability for an edge electron of being scattered to the opposite edge or not. At this point our approach formally meets the LB formula for $R_{xx} = (h/e^2)[R/(vT)]$ and a comparison yields the result that P corresponds formally to the ratio R/T of the LB formula. However, the main difference is that our result is directly obtained from considering the presence of a dissipative current which is not an edge current. In a more detailed treatment, which is presented in Ref. [17], it is shown that the same result is obtained if one considers EC-backscattering to appear continuously all along the conductor like one has to expect in a realistic case and therefore we ask to refer to Ref. [17] for more details. It is easily checked that Eq. (3) also obeys energy conservation: The dissipation due to the sample current I is given by $I^2 R_{xx}$ which is the same as $I^2 P R_{xy} = I_b I R_{xy} = I_b U_{xy}$.

3. Modeling of transport

For the case of a *single* LL, which is represented by a single pair of ECs, we substitute R_{xy} by h/e^2 and get

$$R_{xx} = \frac{h}{e^2} P. \tag{4}$$

For a standard QH-system, as in e.g. AlGaAs/GaAs, backscattering appears only in the top LL in the regime between plateaus, while the transport in the lower LLs remains dissipationless. For a transport model in the EC-picture one has therefore to combine one pair of ECs with non-zero backscattering ($P > 0$) and a set of EC pairs without backscattering ($P = 0$). R_{xx} and R_{xy} of the complete system must finally result from the current distribution between both EC systems. [1] For treating these parallel systems we use the components of the conductance tensor G_{xx} and G_{xy}, which

[1] In what follows we present an analytical version of our model. The results are fully in agreement with those of a numerical model which does not use the tensor relations. In the numerical model the sample current is allowed to flow via the two parallel EC systems, which are connected at the metallic contacts only. The potential differences at the contacts are then obtained in an iterative way from current conservation considerations. The numerical version of our model is able to give correct results for non-local contact configurations as well (considered for publication elsewhere).

can be obtained from the components of the resistance tensor R_{xx} and R_{xy} by the well known relations $G_{xx,xy} = R_{xx,xy}/(R_{xx}^2 + R_{xy}^2)$. The use of these equations means that we restrict our analysis to the case of a symmetric behavior where $R_{xx} = R_{yy}$. In comparison with classical transport this corresponds to the case of a quadratically shaped conductor. Consequently the equations are formally identical with the equations for the resistivity ρ_{xx}, ρ_{xy} and conductivity σ_{xx}, σ_{xy}. In order to point out that the resulting quantities are not necessarily local quantities, we continue to use the symbols for global conductances $G_{xx,xy}$ and resistances $R_{xx,xy}$. Using Eq. (3) we get for the top LL:

$$G_{xx}^{top} = \frac{e^2}{h} \frac{P}{1+P^2}. \tag{5}$$

Due to the absence of backscattering in the lower LLs we have $G_{xx}^{low} = 0$ and therefore the total G_{xx} is given by G_{xx}^{top}. In an analogous way we calculate the Hall components,

$$G_{xy}^{top} = \frac{e^2}{h} \frac{1}{1+P^2}, \tag{6}$$

$$G_{xy}^{low} = \frac{e^2}{h} \bar{v}, \tag{7}$$

where $\bar{v}$ is the number of filled LLs below the top LL. The total Hall conductance G_{xy} is given by the sum of Eqs. (6) and (7).

Now, using $R_{xx,xy} = G_{xx,xy}/(G_{xx}^2 + G_{xy}^2)$ we obtain Eqs. (8) and (9)

$$R_{xx} = \frac{h}{e^2} \frac{P}{(\bar{v}+1)^2 + (\bar{v}P)^2}, \tag{8}$$

$$R_{xy} = \frac{h}{e^2} \left[\bar{v} + \frac{1}{1+P^2} \right] \left[\bar{v}^2 + \frac{2\bar{v}+1}{1+P^2} \right]^{-1}, \tag{9}$$

where P will be a function of the partial filling v^{top} of the top LL. Even though we can obtain already quite a number of important results without knowing the specific function $P(v^{top})$, we will now continue to derive an appropriate function which is required for generating complete transport traces. A detailed discussion of results, which can be obtained already without
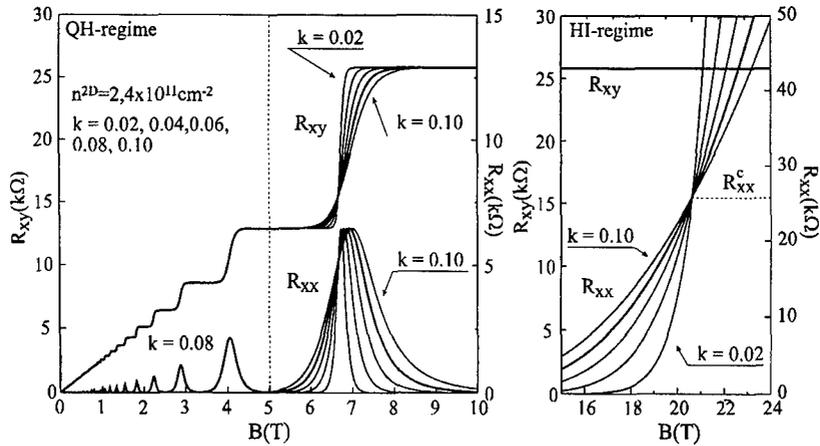
Fig. 3. R_{xx} and R_{xy} calculated according to Eqs. (8), (9) and (10) for a sheet carrier density of $n^{2D} = 2.4 \times 10^{11}$ cm^{-2} and different factors k in the exponent of $P(\Delta v)$. The range below $B = 5$ T shows just the traces for $k = 0.08$, the range above $B = 5$ T shows the traces for all different k values as given in the figure. The HI-regime is shown separately with a different R_{xx} scale on the right.

knowing this function is given in the discussion section below.

Considering Eq. (5) one can see that G_{xx} is proportional to P for $P \ll 1$, while it changes to a reciprocal dependence on P for $P \gg 1$. This implies that there exists a maximum of G_{xx}, which has to be identified with the experimentally obtained G_{xx}-peaks. If we now request a symmetric form of G_{xx}, we have to look for a suitable monotonous function $P(v^{top})$. To get perfect symmetry, the form of Eq. (5) requires a function which fulfils the relation $P(\Delta v) = 1/P(-\Delta v)$ where Δv is the filling factor of the top LL relative to the center. It is easily seen, that the only function which is also in agreement with the experimental observations [10] is of the form

$$P(\Delta v) = \exp(-\Delta v/k), \tag{10}$$

with k being a constant but possibly temperature dependent factor $k(T)$. In order to get a curve without a point of inflection at $\Delta v = 0$, like experimentally observed, Δv must appear linearly in the exponent. Since the maximum of G_{xx} is identified with the center of the top LL, Δv is the filling factor relative to half filling. From Fig. 3 one can see that the calculation based on Eqs. (8), (9) and (10) reproduces very well the typical traces known from the experimental curves at different temperatures.

4. Results and discussion

Although we have already given an appropriate function $P(\Delta v)$ which allows us to generate complete QHE traces, we will show that a number of fundamental results can be obtained already without needing a particular function for $P(v^{top})$. As a first point one can directly see that Eqs. (5) and (6) fulfill the semicircle relation, $\sigma_{xx}^2 + \sigma_{xy}^2 \propto \sigma_{xy}$, which is valid also for the complete system. This semicircle relation was experimentally found to be valid for the top LL as well as for the HI-regime [11]. Based on the special form of Eq. (5) we can distinguish two regimes which are divided by the point at which $P = 1$. In the following, we will show that the regime $0 < P < 1$ can be attributed to the situation where the Fermi level E_F is located above the center of the top LL, while $P > 1$ corresponds to E_F below the center of the top LL.

Fig. 4a and b show schematically the situation in the two regimes: While E_F is moving towards the center of the broadened top LL (Fig. 4a), localized magnetic bound states are created in the bulk region in addition to the associated pair of ECs. Therefore, some transport across those loops by e.g. tunneling becomes possible, which finally acts as EC-backscattering ($0 < P < 1$). According to Eq. (4), R_{xx}^{top} is directly proportional to the backscattering rate and hence proportional to the bulk conductivity in
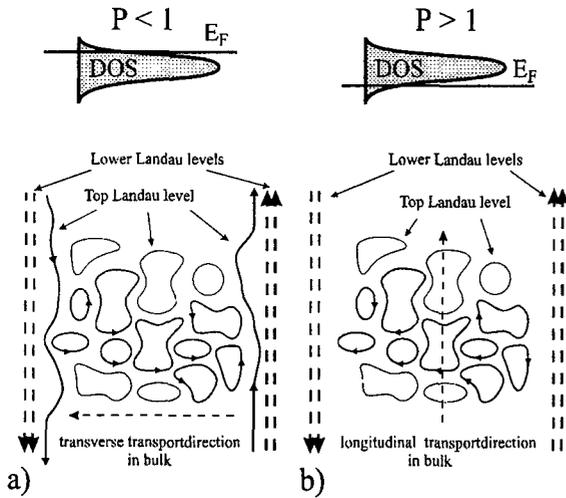
Fig. 4. (a) Edge-channel conduction in the top-LL in the presence of localized magnetic boundstates. The transport across the loops appears as a transverse current, which acts as a backscattering process. (b) Conduction in the top-LL in the presence of localized magnetic boundstates but in the absence of an associated EC-pair. In contrast to the situation sketched in (a), the transport across the loops appears now as a longitudinal current. The ECs of the lower LLs are indicated by the dashed arrows and are considered to be completely de-coupled from the top LL. The relative position of the Fermi level with respect to the LL is indicated at the top of the figure.

this regime. For describing this type of transport in the bulk region, basically a network model such as e.g. that one of Chalker and Coddington [18] would be suitable. A situation with E_F below the LL center is schematically shown in Fig. 4b with one major difference to Fig. 4a, namely that the associated EC-pair is not present, while the transport mechanism in the bulk itself may remain the same like in the regime of Fig. 4a. Consequently, the transport in the bulk does no longer act as a coupling between opposite edges, but may contribute now via a current in the longitudinal direction instead. This will lead basically to a reciprocal dependence of R_{xx} on the bulk conductivity in the associated LL. One can interpret the two regimes as two different phases of the top LL with perpendicular directions of the dissipative bulk current. This is in striking agreement with Ruzin et al. [12] who also found, that for a correct description of the transport behavior the bulk current directions in both phases must be perpendicular to each other.

Characterizing the dissipative transport through the bulk by a conductivity σ_{bulk}, we get basically $R_{xx}^{top} \propto \sigma_{bulk}$ for E_F above the LL center and $R_{xx}^{top} \propto \sigma_{bulk}^{-1}$ for E_F below the LL center. Consequently, any influence of an eventually existing temperature dependence of σ_{bulk} on the longitudinal transport properties must appear with opposite sign in the two regimes. This implies that there must be a crossover of the two regimes where the temperature dependence of R_{xx} is canceled. In this way our model indicates correctly the existence of metallic like and insulator like regimes. It is interesting to note that if one assumes an insulator like temperature dependence for the bulk conductivity σ_{bulk} in the whole transport regime, the temperature dependence of R_{xx} will appear metallic-like for $P < 1$ and insulator-like for $P > 1$, which is in agreement with the experimental observations. It is easily found that the critical (temperature invariant) point in the crossover regime occurs at $P = 1$. According to Eq. (4) this means that at the critical point R_{xx}^{top} approaches the quantized value h/e^2. $P = 1$ also means that for the transport in a single LL $G_{xx} = G_{xy} = 0.5e^2/h$, in agreement with Ref. [19]. In Ref. [11] also the peak between the 1st and 2nd plateau has been analyzed. It has been found that the maximum value is $h/4e^2$ while R_{xx}^c at the critical point appears as $h/5e^2$. In our model the critical point appears at $P = 1$, for which we get a value of $R_{xx}^c = h/5e^2$, in agreement with Ref. [11]. Considering the maximum of Eq. (8) for $\bar{v} = 1$, we find $P = 2$, which leads to $R_{xx}^{max} = h/4e^2$, also in agreement with Ref. [11].

Using the particular function $P(\Delta v)$ according to Eq. (10), we can go a step further: With the help of Eq. (4) we obtain $R_{xx}^{top} = (h/e^2) \cdot \exp(-\Delta v/k)$, which is a monotonous function and covers both regimes $P > 1$ and $P < 1$. Now, we can also consider the principal behavior of σ_{bulk} in the tails of the LL ($P \gg 1$ and $P \ll 1$) by using $\sigma_{bulk} \propto R_{xx}^{top}$ for $\Delta v > 0$ and $\sigma_{bulk} \propto 1/R_{xx}^{top}$ for $\Delta v < 0$. As one expects for pure bulk transport, we obtain a symmetric function for σ_{bulk} around the LL-center $\sigma_{bulk} \propto \exp(-|\Delta v|/k)$. Thus, it is demonstrated, that our model provides the correct framework to include also dissipative bulk transport.

The experimental evidence for the non-symmetric transport behavior of R_{xx}^{top} comes with E_F in the lowest LL ($\bar{v} = 0$, see Eq. (8)). There R_{xx} is identical to R_{xx}^{top} and increases monotonously with decreasing

filling factor. This is exactly the very well experimentally investigated regime of the HI: R_{xx}^{ins} has been found to be monotonously increasing without any peak behavior and R_{xy} stays at the quantized value h/e^2, in agreement with Eq. (9) for $\bar{v} = 0$. Therefore, we can interpret the behavior in the HI regime to be a direct consequence of the asymmetric transport behavior of a single LL. Since in our model the transition to the HI as well as the inter-plateau transitions are described by the same function $P(\Delta v)$, the experimentally observed equivalent behavior of R_{xx}^{top} and R_{xx}^{ins} [11] is an inherent property of our model.

The fact, that the temperature dependence disappears at a certain point, suggests that the temperature T enters only the factor k in the exponent of Eq. (10). Moreover, $\Delta v = 0$, in Eq. (10) means that $P = 1$ and therefore $R_{xx}^c = h/e^2$ for $\bar{v} = 0$, in agreement with Ref. [8]. This is also evident from Fig. 3, where the traces cross each other at $R_{xx}^c = h/e^2$ at ($B = 20$T).

A widely used basis for the discussion of experimental data is the plot of the ρ_{xx} peak width ΔB as a function of temperature. In this context, we analyze the width of the G_{xx} peak, which is described by Eq. (5): $G_{xx} \propto 1/(P + 1/P)$ is symmetric in P with respect to $P = 1$ and the maximum appears at $P = 1$. On the basis of this symmetry we choose a point on each side of the G_{xx} maximum. The associated values of the backscattering function $P(\Delta v)$ are P_1 and $P_2 = 1/P_1$, with P_1 being a constant, except unity. We can write $P_1 = \exp(\Delta v_1/k)$ and $P_2 = \exp(-\Delta v_1/k)$ and obtain an invariant expression by considering the relation $P_1/P_2 = P_1^2 = \exp(2\Delta v_1/k)$, where $2\Delta v_1$ can be identified as the width of the G_{xx} peak on the filling factor scale. Applying the logarithm results in $\ln(P_1^2) = 2\Delta v_1/k = $ constant, which means that the temperature dependence of $2\Delta v_1(T)$ and $k(T)$ must be the same, regardless of any particular form of $k(T)$. One gets temperature independent traces if one plots all temperature dependent traces for R_{xx} versus $\Delta v/k(T)$. In this way such "scaling plots" of the whole R_{xx} can also be used to obtain the unknown function $k(T)$ from the experimental data. This was done by Shahar et al. and the fact that all experimentally obtained traces of R_{xx}^{top} and R_{xx}^{ins} of Ref. [11] collapse onto a single trace, if plotted with respect to $(v - v_c)T^{-\kappa}$, suggests that the argument of the exponential function should have the form

$\alpha(v - v_c)T^{-\kappa}$ with α being a constant. However, as evident from above, also any alternative temperature dependence $k(T)$ can be used in our model, such as e.g., $k(T) = \alpha + \beta T$, which has been suggested recently by Shahar et al. [20] have interpreted the appearance of this new temperature dependence as the indication of a new transport regime for the QHE. Since we have shown that many of the features can be modeled without needing a special form of the backscattering function, it might indeed happen, that different dominating bulk transport mechanisms cause similar experimental features.

5. Summary and conclusion

We have presented a model for the IQHE which makes use of a novel representation of backscattering. It successfully describes the full transport regime between the plateaus as well as the transitions between the QH-liquid and insulator regimes. Even though we use the edge channel approach for the IQHE as an input for our model, the results are more general and the model provides the correct framework to include also dissipative bulk transport. We have shown that already the interplay between edge and bulk leads to a number of important features, which seem to be insensitive to the nature of a particular bulk transport mechanism. Quite a number of well-known facts can be obtained without needing the particular function of backscattering versus Landau level filling $P(\Delta v)$: (i) the semicircle relation between σ_{xx} and σ_{xy} for the complete QH regime as well as for the Hall insulator (HI) regime, (ii) the critical value $\rho_{xx}^c = h/e^2$ in the HI regime, (iii) the value $\sigma_{xx} = \sigma_{xy} = 0.5e^2/h$ at the critical point for a single Landau level, (iv) the maximum value $\rho_{xx}^c = h/4e^2$ for the $1 \to 2$ transition, (v) the critical value $\rho_{xx}^c = h/5e^2$ for the $1 \to 2$ transition. Using an exponential function for $P(\Delta v)$ we obtain further: (vi) the validity of the relation $\rho_{xx}(\Delta v) = 1/\rho_{xx}(-\Delta v)$ between the HI and the adjacent QH-liquid regime, (vii) the equivalence of the temperature scaling of ρ_{xx} in the HI regime, of ρ_{xx} of the top LL and of the ρ_{xx} – peak width. (viii) regarding the temperature dependence (using any $k(T)$ monotonously increasing with T), the model indicates correctly the existence of metallic like ($P < 1$) and insulator like ($P > 1$) regimes.

Acknowledgements

References

[1] K. von Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. 45 (1980) 494.

[2] M. Büttiker, Phys. Rev. B 38 (1988) 9375.

[3] P.L. McEuen et al., Phys. Rev. Lett. 64 (1990) 2062.

[4] J.K. Wang, V.J. Goldman, Phys. Rev. B 45 (1992) 13 479.

[5] J. Oswald et al., Solid State Commun. 102 (1997) 391.

[6] For a recent review see S.L. Sondhi, S.M. Girvin, J.P. Carini, D. Shahar, Rev. Mod. Phys. 69 (1997) 315.

[7] See also B. Huckestein, Rev. Mod. Phys. 67 (1995) 357.

[8] D. Shahar et al., Phys. Rev. Lett. 74 (1995) 4511.

[9] M. Shayegan, Solid State Commun. 102 (1997) 155.

[10] D. Sahar et al., Solid State Commun. 102 (12) (1997) 817.

[11] D. Shahar et al., Phys. Rev. Lett. 79 (1997) 479.

[12] I. Ruzin, Shechao Feng, Phys. Rev. Lett. 74 (1995) 154.

[13] M. Büttiker, Semiconductors and Semimetals, vol. 35, Academic Press, New York, 1992, p. 191.

[14] J.K. Jain, S.A. Kivelson, Phys. Rev. Lett. 60 (1988) 1542.

[15] P. Streda, J. Kucera, A.H. MacDonald, Phys. Rev. Lett. 59 (1987) 1973.

[16] G. Müller et al., Phys. Rev. B 42 (1990) 7633.

[17] J. Oswald, G. Span, Semicond. Sci. Technol. 12 (1997) 345.

[18] J.T. Chalker, P.D. Coddington, J. Phys. C 21 (1988) 2665.

[19] Y. Huo, R.E. Hetzel, R.N. Bhatt, Phys. Rev. Lett. 70 (1993) 481.

[20] D. Shahar et al., preprint cond-mat/9706045.

Composite fermions at different levels

H.L. Stormer [a,b,*], A.S. Yeh [c], W. Pan [c], D.C. Tsui [c], L.N. Pfeiffer [a], K.W. Baldwin [a], K.W. West [a]

[a] *Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA*
[b] *Department of Physics and Department of Applied Physics, Columbia University, New York, NY 10022, USA*
[c] *Princeton University, Princeton, USA*

Abstract

This article provides a short review of recent developments in the area of the fractional quantum Hall effect. It attempts to provide some intuitively appealing images of the highly correlated electron behavior and the new quasi-particles it creates at half-filled Landau levels. Finally, it surveys in a cursory manner our experimental knowledge and lack of knowledge of composite fermions at fractions different from half-filling. Recent theoretical calculations promise yet more bizarre electronic states to be discovered. © 1998 Elsevier Science B.V. All rigths reserved.

Keywords: Quantum Hall effect; Composite fermions

1. The guiding principle

Vortex attachment has become the unifying principle underlying the multiple novel electronic phases of two-dimensional electrons exposed to a high magnetic field [1,2]. Laughlin's original wave function [3], for the fractional quantum Hall state [4] (see Fig. 1) at Landau level filling factor $v = \frac{1}{3}$ is the prime example for this principle at work. The presence of the magnetic field requires the electron system to assume as many zeroes within a given area as there are magnetic flux-quanta penetrating it [5]. These zeroes "heal" on the scale of a magnetic length ($l_0^2 = h/(2\pi eB)$) and they impart a 2π phase twist to the many-particle wave function at each of their positions. Analogies with

topological excitations in superfluid He establishes the name "vortex" for these objects.

The integrated charge deficit created by a vortex corresponds to a fraction v of the charge of an electron, where v represents the Landau level filling factor which is the degree to which the Landau levels are being filled by electrons. At $v = 1$ the lowest Landau level is totally filled (neglecting spin) and at $v = \frac{1}{3}$ it is filled only to $\frac{1}{3}$ of its capacity. The filling factor also determines the ratio of electrons to flux quanta and, due to the one-flux-quantum-one-vortex correspondence, it establishes the ratio of electrons to vortices. At $v = 1$ the number of vortices equals the number of electrons, whereas vortices outnumber electrons by a factor of 3 to 1 at $v = \frac{1}{3}$ filling. Vortices appear as positive charges to the negative electron. At their position, the positive charges of the neutralizing background of donors are locally uncompensated and the charge deficit of $\frac{1}{3}e$ appears as a positive charge of $+\frac{1}{3}e$.

* Corresponding author. Tel.: +1 908 582 3380; fax: +1 908 582 3260; e-mail: horst@physics-bell-labs.com.
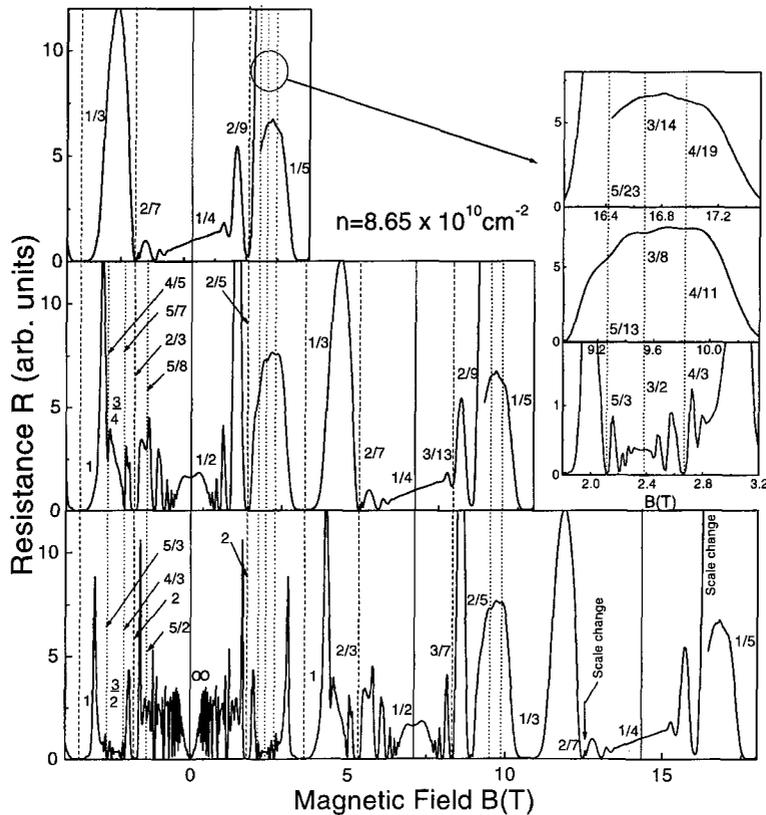
Fig. 1. Magneto-resistance of a high-mobility, two-dimensional electron gas at $T \sim 25$ mK. The middle and top sections of the main figure are offset by one and two-times the magnetic field corresponding to $v = \frac{1}{2}$ and $v = \frac{1}{4}$. This figure shows many self-similarities inherent in such data. Vertical lines indicate but a few of such corresponding filling fractions. The inset shows weak features that are just becoming visible in the most recent specimens and their relationship to previously observed fractions.

Just as electrons, vortices are delocalized in the plane. However, one object representing a charge accumulation, the other a charge deficit, they attract each other. Considerable Coulomb energy can be gained by placing vortices on to electrons.

Of course, each electron must always be "dressed" by at least one vortex (one zero) in order to observe the Pauli principle. Additional vortex attachment is "optional", driven by the inherent Coulomb gain. Vortex attachment to an electron, implying a local displacement of companion electrons, is energetically always beneficial. The situation is somewhat reminiscent of the screening cloud around an electron in a regular metal, although in the present case, such "screening" is very rigid and quantized. As has long been recognized, the attachment of exactly three vortices of charge deficit $\frac{1}{3}$e to each electron is at the

origin of the $v = \frac{1}{3}$ FQHE state, so elegantly expressed by Laughlin's wave function.

Electrons adorned by vortices can be viewed as new, composite particles. The 2π phase change per vortex which they impart to the many-particle wave function, can be regarded as a property of these composites and it dramatically alters their statistics. The additional odd-numbered or even-numbered 2π phase twists around each composite particle causes the total wave function to either invert sign or maintain sign under the exchange of two of their constituent particles. Therefore, electrons endowed with an odd number of vortices obey Bose statistics whereas electrons dressed with an even number of vortices obey Fermi statistics.

Hence, one view of the 1/3 FQHE ground state is that of composite bosons, consisting of electrons

plus three vortices, condensing into a Bose condensate and creating a finite gap to any excitation [6,7]. Equivalently, the state at $v = \frac{1}{5}$ (see Fig. 1) consists of 5-vortex electrons condensing into such a bosonic ground state at this particular filling factor. How do the new particles interact with the external magnetic field? Different views can be taken.

A negatively charged electron dressed by three vortices of charge $+\frac{1}{3}e$ appears neutral and is oblivious to the presence of the field. The condensation is a condensation of "neutralized" particles in the presence of large, external magnetic field. Alternatively, one may think of the phase twist around each composite as being generated by the direct attachment of magnetic flux quanta to the electron. Each quantum creates one 2π phase change and exactly all external field is required to generate all phase twists around all electrons. Consequently, in a "reverse mean-field sense", all magnetic field has been "used up" to produce all required phase twists around all particles. In this picture, the modification of the statistics is particularly transparent, since exchange of two particles adds an Aharanov–Bohm phase due to the attached magnetic flux. The resulting particles are charged (electron plus magnetic flux) and are moving in an apparently vanishing magnetic field [8]. When vortices unbind, either by thermal agitation or by a small change in the ratio of vortices to electrons, they become quasi-particles in the sea of condensate, carrying with them a fractional charge of $\frac{1}{3}e$. These are the $\frac{1}{3}$-charged quasiparticles whose existence has been deduced from several experiments [9–12], most recently in measurements of the shot noise in a tunneling geometry [13,14]. Other fractions are expected to carry other fractional charges such as $\frac{1}{5}e$, $\frac{1}{7}e$, etc.

2. Even-denominators

What about the states at even-denominator filling, where an even number of vortices is available to each electron? These states at $v = \frac{1}{2}$ or $v = \frac{1}{4}$ (see Fig. 1) had been regarded simply as the transition regions from one FQHE state to the next. Compared with the odd-denominator FQHE states, even-denominator states appeared uninspiring, showing hardly any variation in temperature-dependent transport. As it turns out, the even-denominator regions are every bit as fascinating as their odd-denominator counterpart.

Vortex attachment, always being beneficial, is taking place here too. However, the resulting composite particles are fermions and not bosons. Therefore, instead of condensing, CFs populate a Fermi sea up to a Fermi energy and a maximum wave vector k_F. During the past few years, there has been ample experimental support for the existence of such a Fermi wave vector [15–19]. What are these CFs? How can one visualize them? Recently, a very interesting view has emerged [20–23].

Just as at $v = \frac{1}{3}$, so also at $v = \frac{1}{2}$ vortices are attaching themselves to electrons in order to reduce Coulomb interaction. The simplest and energetically most beneficial way to achieve this would be to place the two vortices per electron at $v = \frac{1}{2}$ exactly on to the electron's position. Such a configuration represents a straight generalization of Laughlin's wave function for the $v = \frac{1}{3}$ FQHE state, lowering the exponent from 3 to 2. However, such a state is not allowed, since it violates the anti-symmetry requirement for an electronic wave function.

3. Fermi sea of dipoles

Barring exact vortex attachment, vortex proximity still remains beneficial. The electron system can maintain anti-symmetry while reaping considerable Coulomb gain by playing a slight variation on the vortex attachment scheme. It centers one vortex directly on to each electron (Pauli principle) and keeps the second vortex as close as possible, although somewhat off center. The electron system seems to be able to succeed in this balancing act by proceeding in a fashion that is analogous to the filling of a Fermi sea.

Filling a Fermi sea, one commences by placing the first particle into the $k = 0$ state, progressively filling larger k-states up to k_F. The CF system apparently follows an equivalent route by successively increasing the separation between electron and vortex. On the first electron the second vortex is exactly centered (equivalent $k = 0$). It is placed somewhat off-center on the second and attached at yet a further distance on the third electron. Successively increasing this displacement in additional electron–vortex pairs, finally a maximum r_F is reached (equivalent k_F). The resulting particles are no longer monopoles as in the $v = \frac{1}{3}$ state, but dipoles. They consist of two charges ($-e + 1/2e = -1/2e$ and

+1/2e) each spread out over a magnetic length, which are separated by a distance r. In this model, filling of the Fermi sea of CFs becomes quite transparent.

The picture is reminiscent of the motion of excitons in the presence of a magnetic field. The opposite charges of a dipole experience opposite Lorentz forces when traversing a magnetic field. In equilibrium, the "unbinding" Lorentz force must be exactly compensated by the Coulomb attraction between the oppositely charged partners. In this classical image, the stable state of an electric dipole in a magnetic field is a dipole in motion and different separations correspond to different momenta. These states of increasing momenta are filling up a Fermi sea in momentum space which corresponds to the filling up of an equivalent Fermi sea in dipole moment space. Amazingly, balancing Coulomb attraction with Lorentz repulsion one can even derive a dispersion relation for such particles and hence a mass at k_F. This mass is born exclusively from electron–vortex interaction and hence is of purely electron–electron origin and unrelated to any intrinsic mass of the carriers. As expected from a simple dimensionality argument, the mass is only dependent on the square root of the magnetic field and therefore only on electron density [8].

Even the Hall current can be visualized in such a dipole picture: An electric field favors aligned dipoles as compared to opposing dipoles. Keeping in mind the 90° angle between dipole moment and momentum of each dipole, an electric field in the Hall direction creates a current perpendicular to it, as required in a Hall geometry. Accounting for the charge transported by such a current requires some argumentation, since the dipoles of the model should have zero net charge. In analogy to a metal in which the screening charge does not modify the charge transported per electron, here too, the screening charge stemming from the vortices does not enter the charge e, which is ultimately carried for each electron in an electrical current.

A model for CFs based on the existence of such dipoles is presently intensely debated. It remains unclear to which degree it is valid and what its implications are for such properties as the density of states in the vicinity of the Fermi energy [8]. The model derives from a trial wave function proposed by Rezayi and Read [24] which represents the logical limit of the wave functions introduced by Jain for the higher-order FQHE states. However, like all such trial wave

function, ultimately, although highly educated, they remain guesses and should remain subject to further scrutiny. Derivation of such wave functions from a model Hamiltonian is presently being pursued [25].

4. Off even-denominators

What happens as the magnetic field deviates from exactly even-denominator filling? Again, analogies to regular electrons in the vicinity of zero-magnetic field are being drawn.

For small deviations, the particles experience the residual (or effective) magnetic field which bends the initially straight trajectories into classical cyclotron orbits. In this flux-attachment view, the charge of the composite is exactly e and it is affected by the small residual magnetic field, the deviation from half-filling. Alternatively, one may take the vortex-charge-compensation view: As the magnetic field deviates from half filling the charge of the vortices deviate from $+\frac{1}{2}e$ and the compensation of each electronic charge by two vortices is no longer exact. The CF gains a small, although finite net charge which is deflected by the full external magnetic field. The classical cyclotron radius calculated from both perspectives turns out to be identical. It is the resonance of this semi-classical orbit with an externally imposed length which is at the heart of all geometrical resonance experiments that support the existence of a Fermi system at half-filling [15,16,18,19].

5. CF Landau-levels

For large deviations of magnetic field from half-filling, CFs are thought of being quantized into Landau levels of CFs by the residual magnetic field, in analogy to the way electrons are quantized into electron Landau levels by the total field [6,7,26–28]. This notion originates from an ansatz due to Jain [26] who, starting from Laughlin's $\frac{1}{3}$ wave function, proposed trial wave function for higher-order FQHE states (e.g. $\frac{2}{5}$, $\frac{3}{7}$, $\frac{4}{9}$ ...) which are analogous to fully filled electron Landau levels. However, the particles filling such levels are not regular electrons but CFs whose origin of magnetic field resides at $v = \frac{1}{2}$ and not at true zero. The so constructed Landau levels of CFs coincide with

the primary sequence of the FQHE, $v = \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \ldots$ and $v = \frac{2}{3}, \frac{3}{5}, \frac{4}{7} \ldots$ around $v = \frac{1}{2}$. Therefore, the FQHE for electrons has been described as the integral quantum Hall effect (IQHE) of CFs, bringing about intense debate.

6. Pros and cons

A major objection to such a view has been that it trivializes a highly complex many-body effect. More appropriately, one may argue that it transforms the challenge of capturing the highly correlated motion of electrons in the FQHE states into the challenge of creating new particles having little mutual interaction. Both schemes emerge from the same principle of vortex (or equivalently flux quantum) attachment. Some properties are more readily comprehensible from one perspective, while others are more palpable from the alternate vista. The predominant sequence of higher order FQHE states around $v = \frac{1}{2}$ finds an aesthetically pleasing interpretation within the CF picture, whereas the previous hierarchy model [29,30] does not exhibit the same elegance. Furthermore, the CF trial wave functions seem to lend themselves more readily to numerical calculations [31] than previous such attempts. The CF approach also has been instrumental for tackling the previously enigmatic states at exactly even-denominator filling [8] and its limiting case has provided us with a very interesting trial wave function at these filling factors [24]. In this context, one should note that at the primary fractions, $v = \frac{1}{3}, \frac{1}{5}, \frac{1}{7} \ldots$, wave functions derived from both approaches are identical, muting the argument for these cases.

On the other hand, the appearance of a fractional charge, and of fractional quantum numbers in general, becomes less transparent in the CF model. Following the Landau level analogy, quasiparticles in the FQHE states represent CFs promoted to the next CF Landau level. While, admittedly, this provides an attractive image for the nature of such quasiparticles, it misses, at least initially, the fact that these have exact fractional charge. A second step, characterized as "screening" of the initially full electron charge of the CF to the requisite fractional charge, needs to be invoked [32]. Yet, traditional screening would not account for the precision to which an exact fraction of an electron charge is being created except if such

screening is of the vortex-attachment type which ensures the required quantized precision. Here, the plain creation of additional or thermally excited fractionally charged vortices in the ground state is very powerful. It leads in one step to fractionally charged quasiparticles, the true hallmark of the FQHE.

7. CFs and FQHE gaps

Apart from providing good trial wave functions for the higher-order FQHE states around $v = \frac{1}{2}$, the CF model also proved instrumental for rationalizing the relative strength of the energy gaps of these fractions. The apparent analogy between electron Landau levels and CF Landau levels in the FQHE and the existence of a Fermi sea at $v = \frac{1}{2}$, led Halperin, Lee and Read [8] to postulating an analogy between CF gap energies (i.e. FQHE gap of higher-order fractions) and cyclotron energy gap in the electron case. Similar to regular electrons, for which a cyclotron energy gap is opening linearly with real magnetic field, the FQHE gap energies are conjectured to open linearly with effective magnetic field. (A smoothly varying prefactor needs to be added to maintain electron–hole symmetry.) This conjecture finds very good support in several numerical few particle calculations based on different wave functions. In particular, wave functions derived from the CF concept have recently demonstrated a close to linear relationship in energy gap [33] for the series of fractions of $v = \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{4}{9}$. However, it is important to realize that, while opening up almost linearly and hence being similar to the case of regular electrons, the mass value derived from such gaps is unrelated to the cyclotron mass of electrons nor is it dependent on any other material parameter. It is exclusively a function of the square-root of magnetic field and hence of electron density.

There have been several experiments [34–36] measuring the energy gaps of higher-order FQHE states around $v = \frac{1}{2}$. All measurements agree in furnishing an energy gap that is increasing with increasing effective magnetic field away from $v = \frac{1}{2}$. All show a largely linear increase supporting an interpretation in terms of a mass of a particle. In detail, there remain several discrepancies between groups, methods and with theory. Does the mass diverge as one approaches $v = \frac{1}{2}$? Is the dependence further away from $v = \frac{1}{2}$ truly linear

(even after removal of the prefactor, required to satisfy electron-hole symmetry)? Is a Shubnikov–deHaas interpretation of the FQHE oscillation valid? Can one derive a Dingle temperature from the data (a measure for the scattering time) in the traditional fashion? Are activation energy data affected by the lifetime of the particles and to what degree? How does disorder affect theory and experiment?

These are but a few of the questions that need to be addressed in a much more systematic fashion. However, the general features of the existing data support the notion that for sufficient distance from exactly $v = \frac{1}{2}$ filling the energy gaps of the FQHE show a largely linear dependence on effective magnetic field, consistent with an interpretation in terms of the mass of CFs.

Self-similarities are another aspect of transport data in the IQHE/FQHE regime, see Fig. 1. They reflect the many different views that can be taken in characterizing a given fractional state. For example, the $v = \frac{1}{3}$ state can be taken as the original Laughlin FQHE state, directly derived from electron–electron correlation. One may also regard it (see middle section of Fig. 1) as the $v = \frac{1}{3}$ FQHE state of CFs from $v = \frac{1}{2}$. Finally (see top section of Fig. 1), one may take it as the $v = 1$ IQHE state of CFs originating from $v = \frac{1}{4}$. The $v = \frac{1}{4}$ state itself can be viewed as either a CF with four attached vortices (^{4}CF) or as the $v = \frac{1}{2}$ CF state, not of electrons, but of CFs originating from the electron $v = \frac{1}{2}$ state. Many further such self-similarities are indicated by vertical dashed and dotted lines in Fig. 1.

8. What about those other CFs?

Filling factor $v = \frac{1}{2}$ is the most prominent of even denominator fractions. However, CF formation is expected around all even-denominator fractions. Compared to $v = \frac{1}{2}$ we know much less about these other fractions.

9. $v = \frac{3}{2}$

Apart from $v = \frac{1}{2}$, $v = \frac{3}{2}$ is the best known of the even-denominator states. The FQHE states in the vicinity of this filling factor ($v = \frac{4}{3}, \frac{7}{5}, \frac{10}{7} \cdots \frac{5}{3}, \frac{8}{5}, \frac{11}{7} \cdots$) have become the prime example for the

impact of spin on the CF level structure. At these filling factors, the lowest spin level of the lowest Landau level is fully occupied ($v = 1$) and the upper spin level of this same Landau level is only partially occupied. Different from the behavior around $v = \frac{1}{2}$, where the strength of the FQHE features increases with increasing deviation from $v = \frac{1}{2}$, these equivalent transport features around $v = \frac{3}{2}$ show a non-monotonic variation.

Tilting the magnetic field away from normal has a dramatic impact on these FQHE states, different from equivalent experiments around $v = \frac{1}{2}$ (at least at the common electron density of $\sim 2 \times 10^{11} \text{cm}^{-2}$). All disappearance and reappearance of energy gaps around $v = 3/2$ find a beautifully simple interpretations in terms of CFs with a spin and a g-factor [37–39]. As a function of tilt angle, the Zeeman energy, which depends on total magnetic field, and the CF cyclotron energy (FQHE energy gap), which depends on perpendicular magnetic field, repeatedly coincide. This results in the collapse of the gap at such v. A simple intersecting fan diagram is very successful in identifying the location of all the experimentally observed positions of vanishing excitation gap. Furthermore, the diagram identifies the spin polarization of all FQHE states at any angle. A quantitative analysis leads to a g-factor of $g \sim 0.6$ with a slight, linear dependence on filling factor. Such a simple dependence is surprising since one expects carriers in different FQHE states and of different spin polarizations to have rather different exchange interactions. Since exchange is a major contributor to the effective g-factor one would expect discrete jumps in g which are presently not apparent in the data.

In case of the FQHE states around $v = \frac{3}{2}$, the CF model has provided us with a conceptual framework in which to analyze the experimental observations and simplify a previously complex pattern. Most recent experiments on specimen with a variation in electron density by a factor of four reinforce the previous interpretation [40].

10. $v = \frac{1}{4}$

The state at filling factor $v = \frac{1}{4}$ are expected to consist of CFs composed of electrons with 4 attached vortices. Alternatively, one could regard $v = \frac{1}{4}$ as the

$v = \frac{1}{2}$ state of CFs emanating from $v = \frac{1}{2}$, see Fig. 1. Since electrons, which are filling half of an electron Landau level at $v = \frac{1}{2}$ are forming ^{2}CFs (two attached vortices), ^{2}CFs filling half of a ^{2}CF Landau level at $v = \frac{1}{4}$ are forming ^{2}(^{2}CFs) i.e. ^{4}CFs. In analogy to $v = \frac{1}{2}$, the FQHE states at $v = \frac{1}{5}, \frac{2}{9}, \frac{3}{13} \cdots \frac{2}{7}, \frac{3}{11}, \frac{4}{15} \cdots$ derive from the Landau levels of such ^{4}CFs. Compared with the $v = \frac{1}{2}$ sequence, we know very little about the states around $v = \frac{1}{4}$. The proximity of such states at small v to filling fractions of diverging magneto-resistance, which is believed to be associated with electron crystal formation, complicates matters. This is apparent in recent SdH data [41] which show strong asymmetries of the results for v higher than $\frac{1}{4}$ as compared to v smaller than $\frac{1}{4}$. Since for a given specimen, experiments around $v = \frac{1}{4}$ require much higher magnetic fields than $v = \frac{1}{2}$ and since Zeeman splittings are generally much larger than around $v = \frac{1}{2}$ or $v = \frac{3}{2}$, no spin gap/Landau level gap coincidences have been observed.

11. $v = \frac{3}{4}$

The states at $v = \frac{3}{4}$ are expected to be symmetric to the states at $v = \frac{1}{4}$ due to electron-hole symmetry. Alternatively, one may regard it as the $v = -\frac{3}{2}$ state of ^{2}CFs, where the minus sign denotes the negative effective field direction. However, comparison with the $\frac{3}{2}$ state is somewhat misleading since the $\frac{3}{2}$ state (at typical densities) consist of a fully filled lower spin level and a half-filled upper spin level. The $\frac{3}{4}$ state, on the other hand, is fully spin-polarized and, in the CF model, consists of a fully filled lowest CF Landau level with the next CF Landau level half-filled. In this sense, it is more appropriate to compare the $\frac{3}{4}$ state with the $\frac{5}{2}$ state in the second Landau level, which has very unusual properties, as discussed below. FQHE states around $v = \frac{3}{4}$ are typically much less developed and it is hard to quantify their gap energies. Data on the sequence of FQHE states around $v = \frac{3}{4}$ are just emerging [40]. In tilt experiments, which vary the Zeeman energy of the electrons, the FQHE states to both sides of $\frac{3}{4}$ show features very similar as data to both sides of $v = \frac{3}{2}$. These data support and emphasize the notion of CFs with a given mass and g-factor even for the case of ^{4}CFs.

12. $v = \frac{5}{2}$

At this filling factor, the lowest Landau level is totally filled and the lowest spin level of the second Landau level is half-filled. The state at $\frac{5}{2}$ has always been puzzling. It has all the properties of a FQHE state, including energy gap and quantized Hall resistance, in spite of its even-denominator classification. Discovered more than a decade ago [42,43], its origin had always been associated with the peculiarities of the second Landau level, whose wave functions have an additional node. The increased freedom provided by such additional nodes to reduce Coulomb interactions, combined with exchange energy reducing spin flips were conjectured to create a circumstance which favors condensation.

With the advent of the CF model and the experimental observation of a Fermi wave vector at half-fillings the previous interpretation of the origin of the $\frac{5}{2}$ state has recently been revisited. An earlier proposal for the states at half-filling, originally proposed by Greiter and Wilczek [44] and Moore and Read [45] is now being examined as a contender for the state at $v = \frac{5}{2}$. This so-called "Pfaffian-state", shows a very good overlap with the result of finite number exact diagonalizations by Morf [46], although other trial wave functions cannot be ruled out. The "Pfaffian", which has been characterized as a state of paired CFs, represents a very attractive option. If applicable, the state at $v = \frac{5}{2}$ can be thought of as arising in a two step process [47]. First, electrons in the second Landau level at half-filling would form the "familiar" Fermi sea of CFs. These CFs then pair in a BCS-like fashion and condense into a novel "superconducting" ground state of CF. Although this last analogy applies only in a very lose sense, the concept is truly exciting and one should remain hopeful that further investigations into the state at $v = \frac{5}{2}$ may indicate such an extraordinary new phase of two-dimensional electron systems. It is even speculated that such a state would form the true ground state at $v = \frac{1}{2}$ and $\frac{3}{2}$ which remains unobserved due to disorder [23].

13. $v = \frac{7}{2}, \frac{9}{2}, \frac{11}{2}, \frac{13}{2} \cdots$

Such states at yet higher levels of half-filling remain hardly investigated. Features similar to the state

at $v = \frac{5}{2}$ have been observed, although some of the minima show an extraordinary dependence on the current direction in the plane [48]. While the $\frac{7}{2}$ state is expected to mirror the state at $\frac{5}{2}$, yet higher half-fillings reside in yet higher Landau levels with yet more nodes in the single particle wave function. There is much opportunity for yet other electron correlations to take place.

14. Other even-denominator fractions

Is there anything beyond half and quarter-fillings? From a theoretical point of view, all even-denominators should form a Fermi sea as long as they are not being suppressed by disorder. On the experimental side, there is no solid evidence for the existence of any such CF state. This lack of evidence may largely derive from the relatively weak features the Fermi liquid state exhibits as compared with the strong minima of the FQHE. The inset of Fig. 1 shows hints of features which exist at positions where such higher-order states are expected. The $\frac{3}{8}$ state, situated between $v = \frac{1}{3}$ (lowest ^{2}CF Landau level) and $v = \frac{2}{5}$ (second ^{2}CF Landau level) represents the $v = \frac{3}{2}$ state of ^{2}CFs. However, because of the high field which polarizes all spins, such a ^{2}CF $\frac{3}{2}$ state resides in the second ^{2}CF Landau level rather than in the lowest Landau level as it is the case for electrons. Similar to the $\frac{3}{4}$ state discussed above, the $\frac{3}{8}$ state may need to be considered as deriving from the $\frac{5}{2}$ state for electrons. This may be the origin for the slight depression at $v = \frac{3}{8}$. However, this feature may also derive from a generally lower resistance around half-filling as it had been observed in early data around $v = \frac{1}{2}$ and $v = \frac{3}{4}$. The top section of the inset shows features similar to the $\frac{3}{8}$ modulation but resulting from equivalent states in ^{4}CFs, i.e. CFs with four attached vortices. Following the CF scheme in a strictly combinatorial sense, the slight minima at $\frac{5}{13}$ and $\frac{4}{11}$ (around $\frac{3}{8}$) and similar depressions at $\frac{5}{23}$ and $\frac{4}{19}$ (around $\frac{3}{14}$) represent the first CF Landau levels of these higher-order CF liquids.

15. Conclusion

The generalization of the principle of electron-vortex attachment in 2D electron systems from Laughlin's original wave function to even-denominator states has provided us with a concept for the formation of new Fermi systems in the presence of high magnetic fields. Several of the characteristics of these Fermi systems have now been observed. The CF model has also provided us with a concept for generating excellent wave functions for many of the so-called higher-order FQHE states. It has given us a rationale for their sequence of appearance and the relative size of their energy gaps in terms of Landau levels of CFs. Furthermore, the CF model very naturally clarifies much of the self-similarities observed in the magneto-resistance data of 2D specimens. Beyond $v = \frac{1}{2}$ and $\frac{3}{2}$ and associated FQHE states, little is known about even-denominator states and their neighboring FQHE states. There is ample opportunity for novel electron correlations to be discovered. In particular, the $v = \frac{5}{2}$ state has been postulated as consisting of pairs of CFs which condense into a highly correlated state, akin to superconductivity. If the past, which has dazzled us with new correlated electron states in 2D is any guide to the future, this very bizarre Pfaffian-state ought to show up somewhere.

Acknowledgements

References

[1] S. DasSarma, A. Pinczuk, (Eds.), Perspectives in Quantum Hall Effects, Wiley, New York, 1997.
[2] T. Chakraborty, P. Pietilainen, The Quantum Hall Effects, Springer Series in Solid State Sciences (85) 1995.
[3] R.B. Laughlin, Phys. Rev. Lett. 50 (1983) 1395.
[4] D.C. Tsui, H.L. Stormer, A.C. Gossard, Phys. Rev. Lett. 48 (1982) 1559.
[5] N. Read, Semicond. Sci. Technol. 9 (1994) 1859.
[6] S.M. Girvin, A.H. MacDonald, Phys. Rev. Lett. 58 (1987) 1252.

[7] S.C. Zhang, H. Hanson, S.A. Kivelson, Phys. Rev. Lett. 62 (1989) 82.

[8] B.I. Halperin, P.A. Lee, N. Read, Phys. Rev. B 47 (1993) 7312.

[9] R.G. Clark, J.R. Mallett, S.R. Haynes, J.J. Harris, C.T. Foxon, Phys. Rev. Lett. 60 (1988) 1747.

[10] V.J. Goldman, B. Su, Science 267 (1995) 1010.

[11] J.A. Simmons, H.P. Wei, L.W. Engel, D.C. Tsui, M. Shayegan, Phys. Rev. Lett. 63 (1989) 1731.

[12] A.M. Chang, J.E. Cunningham, Surf. Sci. 229 (1990) 216.

[13] R. de-Picciotto, M. Reznikov, M. Heiblum, V. Umansky, G. Bunin, D. Mahalu, Nature 389 (1997) 162.

[14] L. Saminadayar, D.C. Glattli, Y. Jin, B. Etienne, Phys. Rev. Lett. 79 (1997) 2526.

[15] R.L. Willett, M.A. Paalanen, R.R. Ruel, K.W. West, L.N. Pfeiffer, D.J. Bishop, Phys. Rev. Lett 54 (1990) 112.

[16] R.L. Willet, R.R. Ruel, K.W. West, L.N. Pfeiffer, Phys. Rev. Lett. 71 (1993) 3846.

[17] W. Kang, H.L. Stormer, L.N. Pfeiffer, K.W. Baldwin, K.W. West, Phys. Rev. Lett. 71 (1993) 3850.

[18] V.J. Goldman, B. Su, J.K. Jain, Phys. Rev. Lett. 72 (1994) 2065.

[19] J.H. Smet, D. Weiss, R.H. Blick, G. Lutjering, K. von Klitzing, R. Fleischmann, R. Ketzmerick, T. Geisel, G. Weimann, Phys. Rev. Lett. 77 (1996) 2272.

[20] N. Read, Semicond. Sci. Technol. 9 (1994) 1859.

[21] N. d'Ambrumenil, private communication.

[22] G. Baskaran, Physica B 212 (1995) 320.

[23] V. Pasquier, F.D.M. Haldane cond-mat/9712169 unpublished.

[24] E. Rezayi, N. Read, Phys. Rev. Lett. 71 (1994) 900.

[25] P. Shankar, G. Murthy, Phys. Rev. Lett. 79 (1997) 4437.

[26] J.K. Jain, Phys. Rev. Lett. 63 (1989) 199; Phys. Rev. B 40 (1989) 8079; 41 (1990) 7653; Science 266 (1994) 1199.

[27] R.B. Laughlin, Phys. Rev. 60 (1988) 2677.

[28] A. Lopez, E. Fradkin, Phys. Rev. 46 (1991) 5246.

[29] F.D.M. Haldane, Phys. Rev. Lett. 51 (1983) 605.

[30] B.I. Halperin, Phys. Rev. Lett. 52 (1984) 1583; 52 (1984) 2390(E).

[31] X.G. Wu, G. Dev, J.K. Jain, Phys. Rev. Lett. 71 (1993) 153.

[32] A.S. Goldhaber, J.J. Jain, Phys. Lett. A 199 (1995) 267

[33] N.E. Bonesteel, Phys. Rev. B 51 (1995) 9917.

[34] R.R. Du, H.L. Stormer, D.C. Tsui, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 70 (1993) 2944.

[35] D.R. Leadley, R.J. Nicholas, C.T. Foxon, J.J. Harris, Phys. Rev. Lett. 72 (1994) 1906.

[36] H.C. Manoharan, M. Shayegan, S.J. Klepper, Phys. Rev. Lett. 73 (1994) 3270.

[37] R.R. Du, A.S. Yeh, H.L. Stormer, D.C. Tsui, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 75 (1995) 3926.

[38] P.J. Gee, F.M. Peeters, J. Singleton, S. Uji, H. Aoki, C.T.B. Foxon, J.J. Harris, Phys. Rev B 55, (1996) R14 313.

[39] P.T. Coleridge, Z.W. Wasilewski, P. Zawadzki, A.S. Sachrajda, H.A. Carmona, Phys. Rev. B 52 (1995) 11603.

[40] A.S. Yeh, H.L. Stormer, D.C. Tsui, L.N. Pfeiffer, K.W. Baldwin, K.W. West, unpublished.

[41] W. Pan, A.S. Yeh, H.L. Stormer, D.C. Tsui, L.N. Pfeiffer, K.W. Baldwin, K.W. West, unpublished.

[42] R.L. Willett, J.P. Eisenstein, H.L. Stormer, D.C. Tsui, A.C. Gossard, J.H. English, Phys. Rev. Lett. 59 (1987) 1776.

[43] T. Sajoto, Y.W. Suen, L.W. Engel, M.B. Santos, M. Shayegan, Phys. Rev. B 40 (1990) 12 031.

[44] M. Greiter, X.-G. Wen, F. Wilczek, Phys. Rev. Lett. 66 (1991) 3205.

[45] G. Moore and N. Reed, Nuclear Physics B 360 (1991) 362.

[46] R. Morf, Phys. Rev. Lett. 80 (1998) 1505.

[47] N. Read, R. Morf, private communication.

[48] R.R. Du, H.L. Stormer, D.C. Tsui, L.N. Pfeiffer, K.W. Baldwin, K.W. West, Bull. Amer. Phys. Soc. 38 (1993) 235.

Direct observation of a fractional charge [1]

R. de-Picciotto, M. Reznikov, M. Heiblum *, V. Umansky, G. Bunin, D. Mahalu

Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

Abstract

Ever since Millikan's [1] famous experiment it is well known that the electrical charge is quantized in units of the electronic charge $- e$. For that reason, Laughlin's [2,3] theoretical prediction of the existence of fractionally charged quasi-particles, put forward in order to explain the fractional quantum Hall (FQH) effect, is very counter intuitive. The FQH effect is a phenomenon that occurs in a two-dimensional electron gas (2DEG) subjected to a strong perpendicular magnetic field. This effect results from the strong interaction among the electrons and consequently current is carried by the above-mentioned quasi-particles. We directly observed this elusive fractional charge by utilizing a measurement of quantum shot noise. Quantum shot noise results from the discreteness of the current carrying charges and thus is proportional to their charge, Q, and to the average current I, namely, $S_i = 2QI$ Our quantum shot noise measurements unambiguously show that current in a 2DEG in the FQH regime, at a fractional filling factor $v = \frac{1}{3}$, is carried by fractional charge portions $e/3$; in agreement with Laughlin's prediction. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Fractional charge; Quantum Hall effect; Two-dimensional electron gas

The energy spectrum of a two-dimensional electron gas (2DEG) subjected to a strong perpendicular magnetic field, B, consists of highly degenerate Landau levels with a degeneracy per unit area $p = B/\phi_0$, with $\phi_0 = h/e$ the flux quantum (h being Plank's constant). Whenever the magnetic field is such that an integer number v (the filling factor) of Landau levels are occupied, that is $v = n_s/p$ equals an integer (n_s being the 2DEG areal density), the longitudinal conductivity of the 2DEG vanishes while the Hall conductance equals ve^2/h with very high accuracy. This phenomenon is known as the integer quantum hall (IQH) effect [4].

A similar phenomenon occurs at fractional filling factors, namely, when the filling factor equals a rational fraction with an odd denominator q and is known as the fractional quantum hall (FQH) effect [5]. In contrast to the IQH effect, which is well understood in terms of non-interacting electrons, the FQH effect cannot be explained in such terms and is believed to result from interactions among the electrons, brought about by the strong magnetic field.

Laughlin [2,3] had argued that the FQH effect could be explained in terms of quasi-particles of a fractional charge $- Q = e/q$. Although his theory is consistent with a considerable amount of the experimental data, no experiment directly showing the existence of the fractional charge exists. The early Aharonov–Bohm measurements [6] were proved to be in principle in-

---

* Corresponding author. Fax: +97-415-236-011; e-mail: heiblum.weizmann@wis.weizmann.ac.il.
[1] With permission of Nature.

adequate to reveal the fractional charge [7,8]. More recently, in a beautiful experiment based on resonant tunneling, Goldman and Su [9] claimed to have measured the fractional charge. However, in a similar experiment, Franklin et al. [10], interpreted the results differently. The difficulty in such experiments is that the results provide only an average charge per state and not the charge of individual particles, unless Coulomb blockade arguments are invoked. Quantum shot noise, on the other hand, probes the temporal behavior of the current and thus offers a direct way to measure the charge. Indeed, as early as in 1987, Tsui [11] suggested that the quasi-particle's charge could in principle be determined by measuring quantum shot noise in the FQH regime. However, no theory was available until Wen [12] recognized that transport in the FQH regime could be treated within a framework of one-dimensional (1D) interacting electrons, propagating along the edge of the two-dimensional plane, making use of the so-called Luttinger liquid model. Based on this model subsequent theoretical works [13–15] predicted that quantum shot noise, generated due to weak backscattering of the current, at fractional filling factors $v = 1/q$ and at zero temperature, should be proportional to the quasi-particle's charge $Q = e/q$ and to the backscattered current I_B

$$S_i = 2QI_B. \tag{1}$$

In order to realize such a measurement we utilized a quantum point contact (QPC) – a constriction in the plane of a 2DEG – that partially reflects the current. The high-quality 2DEG, embedded in a GaAs–AlGaAs heterostructure, some 100 nm beneath the surface, has a carrier density $n_s = 1 \times 10^{11}$ cm^{-2} and a mobility $\mu = 4.2 \times 10^6$ cm^2/Vs at 1.5 K in the dark. The QPC is formed by two metallic gates evaporated on the surface of the structure, separated by an opening of some 300 nm which is a few Fermi wavelengths wide (see inset in Fig. 1). By applying negative voltage to the gates with respect to the 2DEG, thus imposing a local repulsive potential in the plane of the 2DEG, one may controllably reflect the incoming current. The sample was inserted into a dilution refrigerator with a base temperature of about 50 mK. Noise measurements were done by employing an extremely low-noise home made preamplifier, placed in a 4.2 K reservoir. The preamplifier was manufactured from GaAs transistors, grown in our molecular beam
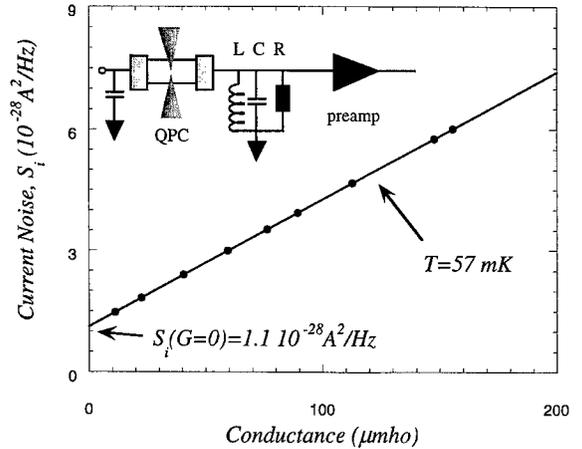
Fig. 1. The total current noise inferred to the input of the preamplifier as a function of the input conductance at equilibrium (full circles). The measured noise is a sum of thermal noise, $4K_BTG$ (leading to a straight line) and the constant noise of the amplifier. This measurement allows the determination of the temperature of the 2DEG. Inset: the quantum point contact in the two-dimensional electron gas is shown to be connected to an LCR circuit at the input of a cryogenic preamplifier.

epitaxy system. The preamplifier has a voltage noise as low as 2.5×10^{-19} V^2/Hz and a current noise of 1.1×10^{-28} A^2/Hz at 4 MHz.

Current fluctuations, generated in the QPC, were fed into an LCR resonant circuit, with most of the capacitance contributed by the coaxial cable which connects the sample at 50 mK to the preamplifier at 4.2 K. Outside the cryostat the amplified signal was fed into an additional amplifier and from there to a spectrum analyzer which measured the current fluctuations within a band of ~100 KHz about a central frequency of ~4 MHz. Since the absolute magnitude of the noise signal is of utmost importance, a careful calibration of the total gain, from QPC to the spectrum analyzer, was done by utilizing a calibrated current noise source. This allows the translation of the spectrum analyzer output into spectral density of current fluctuations (current noise). Although our amplifier has excellent characteristics it still introduces current fluctuations into the circuit. This unwanted current noise must be subtracted from the total measured noise in order to extract the shot noise associated solely with the QPC. By measuring the total current noise while varying the conductance, G, of the unbiased sample

(see Fig. 1), we deduce both the electron temperature, $T = (\delta S_i / \delta G) / 4k_B$ and the contribution of our amplifier to the total noise (extracted from the extrapolated total noise to zero conductance). Note that the temperature we find, 57 mK, is very close to that of the sample holder.

Since the temperature, T, and the applied voltage, V, across the QPC during our measurement are both finite, the results must be compared with a more elaborate theory than the one leading to Eq. (1). Such general calculations were indeed performed numerically [16]. An analytical general expression for the zero frequency spectral density of the total noise is available for a non-interacting single 1D channel and is given by [17–19]

$$S_i = 2g_0 t (1 - t) \left[QV \, \mathrm{Coth} \left(\frac{QV}{2k_B T} \right) - 2k_B T \right] + 4k_B T g_0 t, \tag{2}$$

where the transmission of the QPC, t, is given by the ratio between the conductance, G, and the quantum conductance, $g_0 = e^2/h$. This dependence was experimentally verified [20,21] in the absence of a magnetic field where electron–electron interactions are believed to be non-crucial, with $Q = e$. The same expression, with $Q = e/3$ and $g_0 = e^2/3h$, also does not deviate significantly from the numerical calculations [16] in the limit of weak backscattering of quasi-particles in the FQH regime at $v = \frac{1}{3}$ and in addition reduces to Eq. (1) in the zero temperature limit ($Vg_0 t(1 - t) = I_B t \approx I_B$). Comparing our data with Eq. (2) will thus suffice in order to deduce the quasi-particles charge.

Quantum shot noise measurements as a function of the current through a partially pinched QPC were performed first in the absence of a magnetic field. The results, after calibration and subtraction of amplifier noise, are shown in Fig. 2. The transmission of the lower laying quasi-1D channel in the QPC is simply deduced from the measured conductance normalized by $2e^2/h$ (the factor 2 accounts for spin degeneracy). Our data fits almost perfectly the expected noise of Eq. (2) using the measured electron temperature without any fitting parameters.

The magnetic field was then swept from zero to 14 T. The two-terminal conductance exhibits Hall plateaus, expected in the IQH and in the FQH regimes ($v = \frac{2}{5}, \frac{3}{5}, \frac{2}{3}$ and $\frac{1}{3}$ are clearly visible with a plateau
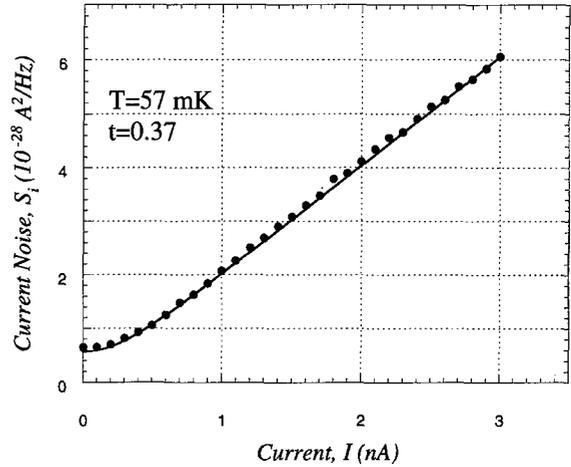
Fig. 2. Quantum shot noise as a function of DC current, I, through the quantum point contact without an applied magnetic field (full circles). The solid line is Eq. (5) with the temperature deduced from Fig. 1.

width of ∼1 T around $v = \frac{1}{3}$). At a filling factor $v = \frac{1}{3}$ and full transmission (zero gate voltage) no excess noise above the thermal one is observed upon driving a current through the sample, thus ruling out noise which is related to overheating. The noise measured upon partially reflecting the current is drastically suppressed compared to the noise measured in the absence of a magnetic field as shown in Fig. 3.

Our data fit very well the expected noise of a current carried by quasi-particles of charge $Q = e/3$. The backscattered current is calculated using the transmission, t, deduced from the ratio of the conductance to $g_0 = e^2/3h$. The slope of the noise versus backscattered current curve increases with applied voltage approaching the expected slope of $2te/3$ at voltages larger than $2k_B T/Q$ as expected. For comparison, the expected noise for $Q = e$ and the same g_0 is also shown.

The noise tends to saturate at even larger backscattered currents (note the deviation of the data points from the solid line). This additional noise suppression is accompanied by an onset of non-linearity in the I–V characteristics (not shown). The non-linearity in the FQH regime may result from the interaction among the electrons, an energy dependence of the bare transmission coefficient and from a finite excitation gap (a gap [11], $\Delta \approx 250\,\mu eV$, is expected at ∼13 T). These
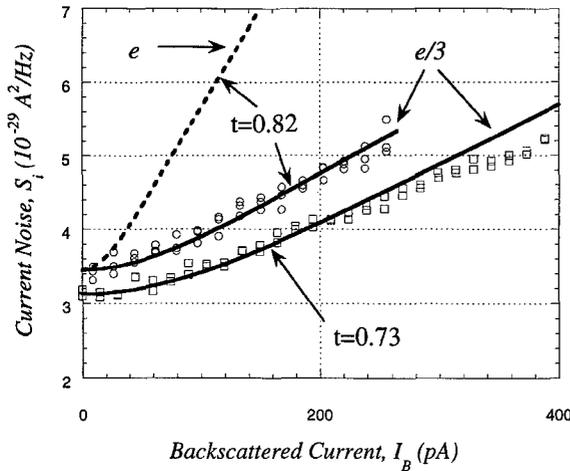
Fig. 3. Quantum shot noise as a function of the backscattered current, I_B, in the fractional quantum Hall regime at $v = \frac{1}{3}$ for two different transmission coefficients through the quantum point contact (open circles and squares). The solid lines correspond to Eq. (5) with a charge $Q = e/3$ and the appropriate t. For comparison the expected behavior of the noise for $Q = e$ and $t = 0.82$ is shown by the dotted line.

Fig. 4. Quantum shot noise as a function of backscattered current, I_B, in the fractional quantum Hall regime at $v = \frac{1}{3}$, for three different temperatures and a constant transmission coefficient, $t = 0.8$, through the quantum point contact.

three sources are practically indistinguishable. Non-linearity complicates the otherwise straightforward interpretation of our results and we thus choose to show data in a smaller voltage range and for moderate reflection coefficients where the I–V is linear.

In order to further investigate the behavior of quantum shot noise at the FQH regime, we measured the noise versus backscattered current for three different temperatures and a fixed transmission through the QPC (shown in Fig. 4). The data fit the curves expected from Eq. (2) with $Q = e/3$. Note that Eq. (2) with a charge $Q = e/3$ suggests not only that the amplitude of the noise is proportional to Q but also that shot noise is observed above the thermal noise at a characteristic voltage $V = 6k_B T/e$, three fold larger than the value for non-interacting electrons. This is because the potential energy of the quasi-particles is $eV/3$. The agreement between the data and the detailed shape of Eq. (2) at small backscattered currents, gives thus an additional indication for the existence of a smaller charge $e/3$.

In conclusion, our noise measurements unambiguously show that the current in the FQH regime, at filling factor $\frac{1}{3}$, is carried by quasi-particles with charge $e/3$. In contrast to conductance measurements, that measure an averaged charge over quantum states or over time, our quantum shot noise measurement is sensitive to the charge itself. The "magic" of an apparent smaller charge due to electron–electron interactions is a beautiful manifestation of the strength of the theoretical methods [2,3] used to predict such a counter intuitive behavior.

During the writing of this manuscript we became aware of similar work [22] in which the authors measure the same charge at a filling factor $\frac{2}{3}$ in the bulk and $\frac{1}{3}$ near the constriction, also using shot noise measurements.

Acknowledgements

References

[1] R.A. Millikan, Univ. of Chicago Press, Chicago, 1917.
[2] R.B. Laughlin, Phys. Rev. Lett. 50 (1982) 1395.
[3] R.B. Laughlin, Int. J. Mod. Phys. B 5 (1991) 1507.
[4] K. von Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. 45 (1980) 494.

[5] D.C. Tsui, H.L. Stormer, A.C. Gossard, Phys. Rev. Lett. 48 (1982) 1559.

[6] J.A. Simmons et al., Phys. Rev. B 44 (1991) 12 933.

[7] N. Byers, C.N. Yang, Phys. Rev. Lett. 7 (1961) 46.

[8] Y. Gefen, D. Thouless, Phys. Rev. B 47 (1993) 10 423.

[9] V.J. Goldman, B. Su, Science 267 (1995) 1010.

[10] J.D.F. Franklin et al., Surf. Sci. 361 (1996) 17.

[11] The Quantum Hall Effect, R.E. Prange, S.M. Girvin (Eds.), Springer, New York, 1987.

[12] X.G. Wen, Phys. Rev. B 41 (1990) 12 838.

[13] C.L. Kane, M.P.A. Fisher, Phys. Rev. Lett. 72 (1994) 724.

[14] C. de Chamon, D.E. Freed, X.G. Wen, Phys. Rev. B 51 (1995) 2363.

[15] A. Fendley, W.W. Ludwig, H. Saleur, Phys. Rev. Lett. 75 (1995) 2196.

[16] P. Fendley, H. Saleur, Phys. Rev. B 54 (1996) 10 845.

[17] G.B. Lesovik, JETP Lett. 49 (1989) 592.

[18] Th. Martin, R. Landauer, Phys. Rev. B 45, (1992) 1742.

[19] M. Buttiker, Phys. Rev. B 46 (1992) 12485.

[20] M. Reznikov, M. Heiblum, H. Srtikman, D. Mahalu, Phys. Rev. Lett. 75 (1995) 3340.

[21] A. Kumar, et al., Phys. Rev. Lett. 76 (1996) 2778.

[22] L. Saminadayar, D.C. Glattli, Y. Jin, B. Etienne, con-mat/9706307, Phys. Rev. Lett. 79 (1997) 2526.

PHYSICA E

The investigation of 1D and 2D phenomena using double-layer electron systems

J.T. Nicholls *, N.P.R. Hill, B. Kardynal, N. Turner, E.H. Linfield, D.A. Ritchie, C.H.W. Barnes, G.A.C. Jones, M. Pepper

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge, CB3 0HE, UK

Abstract

The physics of low-dimensional electron systems continues to be an active field of research; using two examples we show how double-layer electron gases can extend these investigations. First, after reviewing the properties of 2D–2D tunnelling, we present equilibrium tunnelling measurements between a two-dimensional electron gas (2DEG) and an array of independent and identical 1D wires. Using a parallel magnetic field full information about the wires can be obtained, their 1D energies and wave functions, and hence the functional form of the confining potential. In the second case, the Coulomb drag (where a voltage is induced in one layer when a current is passed through the other) is measured between two separately contacted and electrically isolated 2DEGs. At temperatures comparable to the Fermi temperature, the measured transresistivity shows a plasmon enhancement. A comparison between theory and experiment shows the importance of many-body corrections to the coupled plasmon modes. © 1998 Elsevier Science B.V. All rights reserved..

PACS: 73.20.Dx; 71.45.Gm

Keywords: Double-layer electron gases; Tunnelling; Coulomb drag

1. Introduction

The two-dimensional electron gas (2DEG) formed in a quantum well or at a heterojunction is an important building block for studies in low-dimensional semiconductor physics. The low-temperature properties of the 2DEG itself continues to be an active area of research, with experiments in the integer and fractional quantum Hall effect. Studies into lower dimensions can also be based on the 2DEG. Thus, 1D wires and 0D dots are often fabricated using a 2DEG as the starting point. In this case the electrons are further confined using structured gates or etched mesas, both of which are defined to submicron accuracy using electron-beam lithography.

Using the precision of molecular beam epitaxy (MBE) growth to control the width and height of an AlGaAs barrier, double-layer systems have more recently been fabricated and new phenomena have been observed. One example is the new many-body state measured at filling factor $v = 1$ in the quantum Hall effect of a strongly coupled bilayer system. Double-layer systems also provide new methods for probing

---

* Corresponding author. Fax: +44 1223 33 7271; e-mail: jtn11@cus.cam.ac.uk.

low-dimensional systems, such as with compressibility [1], tunnelling, and Coulomb drag measurements. In Sections 2 and 3 we present measurements describing the latter two techniques.

The double-well samples used in the tunnelling and drag studies presented here were grown by MBE. The GaAs wells are 180–200 Å wide and are separated by an $Al_{0.33}Ga_{0.67}As$ barrier of thickness t. An in situ focussed ion-beam system is used to define back gates in an n^+ epilayer prior to growth of the double wells. [2] Using a combination of front and back gates, separate electrical contacts can be made to the individual electron gases. Full details concerning wafer growth and the subsequent processing into working devices can be found in Refs. [3–5].

2. Tunnelling measurements

Double 2DEG tunnelling devices consist of two modulation-doped quantum wells separated by a 125 Å AlGaAs barrier; this gives a distance $d = 325$ Å between the centres of the quantum wells, which allows a small amount of wave function overlap. As shown in the schematic diagram of a 2D–2D tunnelling device in Fig. 1a, front and back gates, with applied voltages V_a and V_b, are used to deplete out the upper and lower 2DEGs and to put the device into a tunnelling configuration. From left to right, the current I flows from the source in the top layer, through the barrier in the tunnelling region, and then out through the lower layer into the drain on the right-hand side. The carrier densities n_1 and n_2 in the tunnelling region are controlled by front and back gate voltages, V_{fg} and V_{bg}. The 2D–2D equilibrium tunnelling conductance, $G = dI/dV$, measured with an AC voltage of 0.1 mV is shown marked with circles in Fig. 1b as a function of the carrier density n_1 of the top layer, with the carrier density of the bottom layer being fixed at $n_2 = 3.27 \times 10^{11}$ cm^{-2}. Satisfying the conditions of energy and momentum conservation [6] gives a tunnelling peak with a maximum at $n_1 = n_2$, and the equilibrium tunnelling line shape is a Lorentzian of width $\Gamma = 0.384$ meV. The line shape gives direct information about the spectral function $A(k, E)$, the probability that an electron with momentum $\hbar k$ has energy E. At low temperatures the spectral function of the ith 2D electron gas $A_i(k, E)$ is
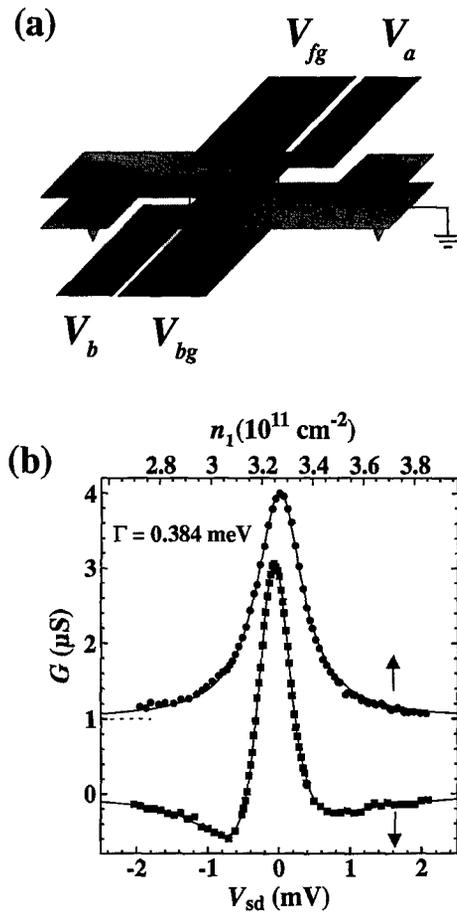
Fig. 1. (a) Schematic diagram of a 2D–2D tunnelling device. Voltages V_a and V_b on the front gate and back gate put the device in a tunnelling configuration, and the carrier densities n_1 and n_2 in the tunnelling region are controlled by V_{fg} and V_{bg}, respectively. (b) A comparison of the equilibrium (circles) and non-equilibrium (squares) tunnelling characteristics at $T = 0.1$ K. The solid lines show fits using $\Gamma = 0.384$ meV. There is a 0.7% mismatch in the carrier densities in the non-equilibrium measurement, which causes $G(V_{sd})$ to be slightly asymmetric.

a Lorentzian of width Γ_i centred at the Fermi energy $E_{i,F} = \hbar^2 k_{i,F}^2 / 2m^*$, where Γ_i is the energy broadening due to scattering. The probability of an electron being available to tunnel is peaked at Fermi energy, and when viewed in k-space the spectral density $A_i(k, E)$ is a ring of width Γ_i with an average radius $k_{i,F}$. The 2D–2D tunnelling current is determined by the overlap integral [7] of $A_1(k, E)$ and $A_2(k, E)$, and the
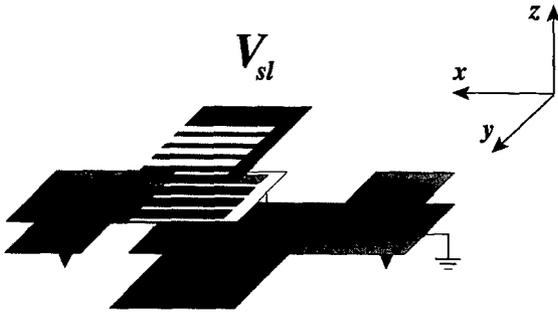
Fig. 2. Schematic diagram of a 1D–2D device in a tunnelling configuration.

resulting Lorentzian line shape of width $\Gamma = \Gamma_1 + \Gamma_2$ has contributions from both 2DEGs.

The sample can also be investigated in a non-equilibrium measurement, where a DC source-drain voltage V_{sd} is added to the 0.1 mV AC signal. The non-equilibrium conductance $G(V_{sd})$ when the carrier densities were matched at $n_1 = n_2 = 3.27 \times 10^{11}$ cm^{-2} is shown as squares in Fig. 1b. Due to capacitive effects a factor $\beta = 0.15$ is introduced (see Ref. [3] for details) to account for charge transfer between the two 2DEGs induced by the applied voltage V_{sd}; taking this into account the measured line widths are enhanced by a factor of $1/(1 - \beta)$ and the resulting fit to $G(V_{sd})$ gives a value of Γ that agrees with equilibrium measurements. The line width Γ has contributions from electron–electron and electron-impurity scattering. The latter term is temperature independent, whereas the electron–electron scattering rate exhibits the $T^2 \log T$ temperature dependence expected theoretically, albeit with a different prefactor [3,8].

Fig. 2 shows a schematic diagram of a 1D–2D tunnelling device, where the upper electron gas is squeezed into 50 parallel 1D wires using a superlattice gate of period 350 nm defined by electron-beam lithography. When the voltage applied to the gate is $V_{sl} = -0.42$ V, the electron gas below the gates in the upper layer is depleted, leaving 1D wires that are approximately 180 nm wide and 50 μm long. As the applied gate voltage V_{sl} is made more negative, both the width and carrier density of the wires are reduced, and the 1D subbands are depopulated. Some of the electric field from the superlattice gate penetrates into the lower electron gas. However, the modulation of the carrier density in the bottom layer is very weak, and

the magnetoconductance can be successfully modelled assuming 1D–2D tunnelling [4].

The wave function in the top electron gas is $\psi_{k_x,n} \sim \exp(ik_x x)\phi_n(y)$, which is a plane wave in the x-direction that is confined in the y-direction; this gives a 1D spectral density that for each subband is separable into k_x and k_y components. The calculated 1D spectral density is shown in Fig. 3a for parabolic confinement in the y-direction and six occupied 1D subbands. Also included is some broadening, $\Gamma = 0.5$ meV, comparable to that seen in 2D–2D tunnelling.

If the 1D–2D equilibrium tunnelling measurement is carried out with a magnetic field B_x applied parallel to the layers, the electron experiences a Lorentz force as it tunnels from one layer to another acquiring some momentum $\hbar \Delta k_y$ in the y-direction. This causes the spectral density in one layer to be translated in k-space by an amount $\Delta k_y = eB_x d/\hbar$ with respect to the other [6]. Fig. 3a shows that as B_x is increased the 2D spectral density ring of the bottom 2DEG is shifted relative to the 1D spectral density; the overlap of the two spectral densities again determines the 1D–2D tunnelling current. The 1D spectral density has much structure. At $k_x = 0$ the dark structures in the $\pm k_y$ direction are the antinodes in $|\phi_n(k)|^2$, the intensity of the Fourier transform of the wave function of the highest occupied 1D subband (in this case $n = 6$). As k_x becomes larger the kinetic energy of the electron's motion along the wire increases, and the highest occupied subband becomes consecutively $n = 5, 4, \ldots$. Although the 2D spectral density overlaps the 1D spectral density over a large area, the dominant contribution to the tunnelling integral comes from the Fourier spectrum of the highest occupied 1D subband. As B_x is increased the trailing edge of the 2D Fermi circle overlaps the structure along $k_x = 0$, and hence the tunnelling current is a direct measure of the intensity of the Fourier spectrum $|\phi(k)|^2$ of the highest occupied 1D subband.

Fig. 3b shows magnetoconductance sweeps as V_{sl} is decreased. As V_{sl} becomes more negative the number of magnetoconductance oscillations decreases reflecting the depopulation of the 1D subbands. At $B_x = 3$ T, the edge of the 2D spectral density intersects the centre of the 1D spectral density function, probing the intensity of the Fourier transform of the wave function of the highest occupied subband $|\phi(k)|^2$ at $k = 0$. The measured magnetoconductance at this magnetic field

(a)

(b)
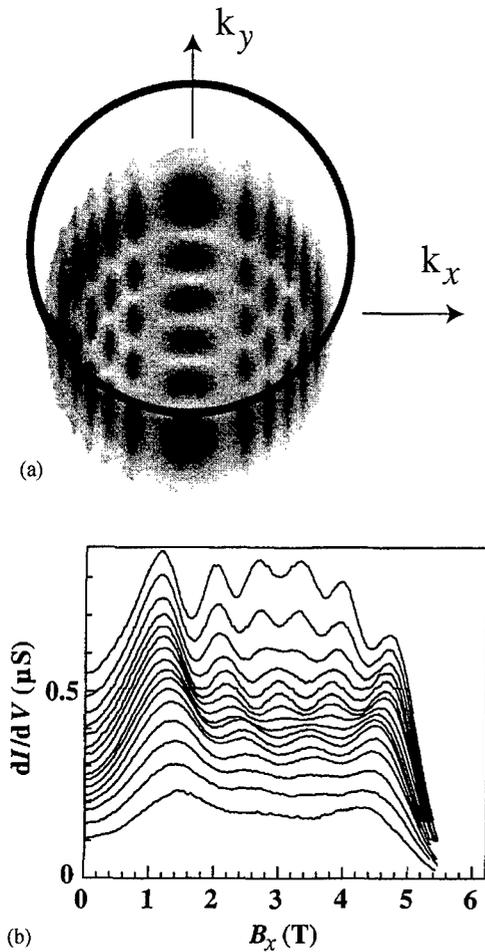
Fig. 3. (a) The calculated 1D spectral density of a wire with parabolic confinement potential, with six 1D subbands occupied and $\Gamma = 0.5$ meV. The 2D spectral density is shifted by an amount $eB_x d/\hbar$ in the k_y direction. The overlap of the two spectral densities determines the magnetoconductance. (b) The magnetoconductance measured at 60 mK as the front gate voltage is decreased from $V_{sl} = -0.47$ V (top trace) to $V_{sl} = -0.62$ V (bottom trace), in steps of 0.01 V. The peaks in dI/dV are the maxima in $|\phi(k)|^2$ of the highest occupied 1D subband.

shows the symmetry of $|\phi(k)|^2$ changing as the subband index n changes from 6 to 5 to 4.

The depopulation of the 1D subbands shown in Fig. 3b can be further studied using a magnetic field B_y applied in the y-direction, which can be used to determine the energy E_n of the 1D subbands. Therefore, a complete picture (energy and wavefunction) of the 1D wires can be obtained, and hence the confine-

ment potential can be determined. Modelling the results shows that a parabolic approximation provides a good description of the confinement when a few subbands are occupied, whereas when the wires are wider the potential becomes flatter at the centre [4].

In contrast to earlier results [9–11], the wave functions $|\phi(k)|^2$ shown in Fig. 3b are obtained from an equilibrium measurement of an array of wires. In the tunnelling formalism each of the 1D wires is considered to be independent, and the measured conductance is the sum of 1D–2D tunnelling contributions from all the wires. The contributions from each of the wires are approximately equal and there is no smearing of the subband structure. Indeed in much wider wires [4] the tunnelling characteristics for the 14th subband can be clearly distinguished from that of the 13th subband. Recent studies [12] of very clean single split-gates show an unusual and reproducible structure in the conductance at $G = dI/dV = 0.7(2e^2/h)$, just after depopulation of the last subband. This is thought to be a many-body effect. Whether the uniformity in the width of the wires continues down to low carrier densities and into the last subband remains to be seen, but 1D–2D tunnelling may prove to be a useful technique to investigate this and similar phenomena in a single-mode quantum wire.

3. Coulomb drag measurements

In order to prevent any interlayer leakage the double-layer samples used in drag experiments have much thicker barriers than tunnelling samples; for the drag measurements presented here, $t = 300$ Å. Using a combination of front and back gates separately contacted Hall bars are formed in each 2DEG. In a Coulomb drag measurement, a drag voltage V_{drag} is induced in one Hall bar when a current I_{drive} is passed through the other. The transresistivity is

$$\rho_t = \frac{V_{drag}}{I_{drive}} \frac{W}{L}, \qquad (3.1)$$

where L is the separation of the voltage probes, and W is the width of the Hall bar. The transresistivity ρ_t is a direct measure of the interlayer electron–electron scattering rate, which is determined by interlayer interaction and the excitations within each layer.

The transresistivity is calculated [13–15] to be

$$\rho_t = -\frac{\hbar^2}{8\pi^2 e^2 n_{\text{drag}} n_{\text{drive}} k_B T} \int_0^\infty dq \int_0^\infty d\omega\, q^3$$

$$\times \left|\frac{V(q)}{\varepsilon(q,\omega)}\right|^2 \frac{\text{Im}\,\chi(q,\omega)_{\text{drive}}\text{Im}\,\chi(q,\omega)_{\text{drag}}}{\sinh^2(\hbar\omega/2k_B T)}, \quad (3.2)$$

where the static interlayer Coulomb interaction is $V(q)$. $\varepsilon(q,\omega)$ is the dielectric function, and the charge density fluctuations in a given layer are characterized by the polarizability $\text{Im}\,\chi(q,\omega)$. At low temperatures and zero magnetic field, Eq. (3.2) predicts a $\rho_t \sim T^2$ temperature dependence [13], IN approximate agreement with the experiment [16]. Deviations from T^2 behavior at low T have been attributed to the exchange of virtual phonons [17]. Further, in a perpendicular magnetic field B the effects of Landau quantisation on the drag have been investigated below 4 K. The measured transresistivity ρ_t shows quantum oscillations periodic in $1/B$, with a magnitude that is enhanced over the zero field value by many orders of magnitude [18,19].

The plasmon dispersion curves $\omega_p(q)$ for a double 2DEG system consists of two branches. The lower (upper) energy branch is the acoustic (optic) plasmon, where the charge density oscillations in the two layers are in antiphase (phase) – these dispersion curves have recently been measured [20] with Raman scattering in double-layer samples similar to our own. At $\omega_p(q)$ the interlayer dielectric constant $\varepsilon(\omega_p, q)$ is at a minimum causing an enhanced interlayer interaction, $W_{12} = V(q)/\varepsilon(\omega_p, q)$. At zero temperature there are no single-particle excitations (SPEs) at $\omega_p(q)$ and the plasmons do not influence the interlayer scattering. At temperatures comparable to the Fermi temperature, T_F, the SPE spectrum is sufficiently broadened to allow the SPEs to sample the poles at $\omega_p(q)$, and the interlayer scattering rate is dominated by these SPEs due to the large interaction, W_{12}. Detailed calculations [15,21] predict a plasmon enhancement of the scaled transresistivity $\rho_t T^{-2}$ starting at $0.2\,T_F$, and peaking close to $0.5\,T_F$. For $T > 0.5\,T_F$ the strong coupling between the plasmons and the SPEs causes Landau damping of the two modes, reducing the plasmon enhancement.

Fig. 4 shows the scaled transresistivity $\rho_t T^{-2}$ versus T/T_F at matched carrier densities of $n = 1.37$
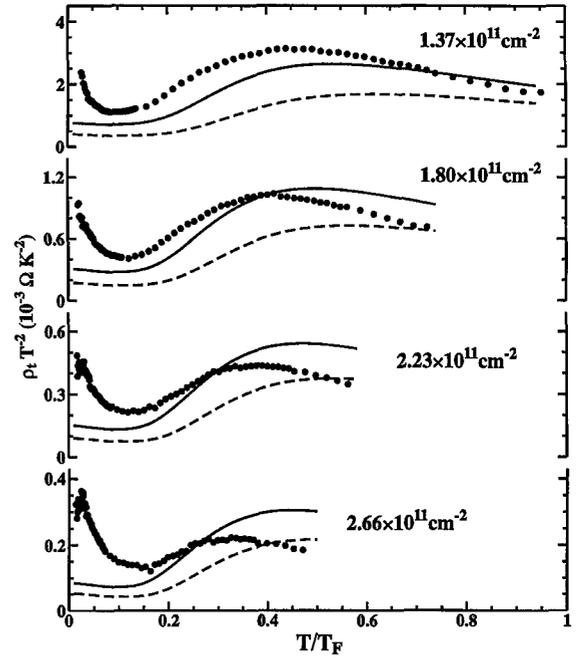
Fig. 4. The scaled transresistivity $\rho_t T^{-2}$ versus the reduced temperature T/T_F, for matched carrier densities $n = 1.37, 1.80, 2.23$, and $2.66 \times 10^{11}\,\text{cm}^{-2}$. The dashed (solid) lines are RPA (Hubbard) calculations. The Hubbard approximation with zero temperature local field corrections improves the agreement between theory and experiment in the low temperature region, but discrepancies remain at elevated temperatures.

$- 2.66 \times 10^{11}\,\text{cm}^{-2}$. Qualitatively the experimental results agree with theoretical predictions. The dashed lines in Fig. 4 show $\rho_t T^{-2}$ based on random phase approximation (RPA) calculations [15] of the Coulomb coupling between the layers, and do not include the phonon exchange that is measured below $0.1\,T_F$. For all densities the transresistivity shows an upturn near $0.2\,T_F$, with a maximum close to $0.5\,T_F$, in good agreement with the RPA calculations. However, the temperature required to excite the plasmons is lower in the experimental traces, suggesting that the RPA overestimates the plasmon energies. The solid lines in Fig. 4 show calculations where the intralayer exchange interactions are included in the Hubbard approximation. The inclusion of many-body correlations lowers the plasmon energy, thereby lowering the temperature required to excite the acoustic plasmon, and improving the fit to the experimental data. However, differences between the theory and experi-

ment remain. The position of the maximum in ρ_t/T^2 lies at a lower temperature than the model, whereas the observed decay at higher temperatures is more pronounced. Recent work [22] on the coupling of the 2DEG to optic phonons shows that the phonon interaction softens the plasmon modes, bringing them closer to the SPE boundary. The resulting traces show a maximum shifted to a lower temperature, with an increased magnitude and a sharper decay. The inclusion of both effects may be required to model the drag measurements.

4. Conclusions

Double-layer electron systems give new insight into the properties of low-dimensional electron systems. 2D–2D tunnelling experiments allow the measurement of the temperature dependent electron–electron scattering rate, and 1D–2D tunnelling gives the energy and wave functions of the 1D system. Drag measurements of double-electron systems show results that depend on the dispersion of the coupled plasmons – properties that are normally measured in an inelastic light scattering experiment and do not usually affect a DC transport measurement.

Acknowledgements

References

[1] J.P. Eisenstein, L.N. Pfeiffer, K.W. West, Phys. Rev. B 50 (1994) 1760.
[2] K.M. Brown, E.H. Linfield, D.A. Ritchie, G.A.C. Jones, M.P. Grimshaw, M. Pepper, Appl. Phys. Lett. 64 (1994) 1827.
[3] N. Turner, J.T. Nicholls, E.H. Linfield, K.M. Brown, G.A.C. Jones, D.A. Ritchie, Phys. Rev. B 54 (1996) 10 614.
[4] B. Kardynal, C.H.W. Barnes, E.H. Linfield, D.A. Ritchie, J.T. Nicholls, K.M. Brown, G.A.C. Jones, M. Pepper, Phys. Rev. B 55 (1997) R1966.
[5] N.P.R. Hill, J.T. Nicholls, E.H. Linfield, M. Pepper, D.A. Ritchie, G.A.C. Jones, B.Y.-K. Hu, K. Flensberg, Phys. Rev. Lett. 78 (1997) 2204.
[6] J. Smoliner, W. Demmerle, G. Berthold, E. Gornik, G. Weimann, W. Schlapp, Phys. Rev. Lett. 63 (1989) 2116.
[7] L. Zheng, A.H. MacDonald, Phys. Rev. B 47 (1993) 10 619.
[8] S.Q. Murphy, J.P. Eisenstein, L.N. Pfeiffer, K.W. West, Phys. Rev. B 52 (1995) 14 825.
[9] W. Demmerle, J. Smoliner, G. Berthold, E. Gornik, G. Weimann, W. Schlapp, Phys. Rev. B 44 (1991) 3090.
[10] N. Mori, P.H. Beton, J. Wang, L. Eaves, Phys. Rev. B 51 (1995) 1735.
[11] P.H. Beton, J. Wang, N. Mori, L. Eaves, P.C. Main, T.J. Foster, M. Henini, Phys. Rev. Lett. 75 (1995) 1996.
[12] K.J. Thomas, J.T. Nicholls, M.Y. Simmons, M. Pepper, D.R. Mace, D.A. Ritchie, Phys. Rev. Lett. 77 (1996) 135.
[13] A.P. Jauho, H. Smith, Phys. Rev. B 47 (1993) 4420.
[14] L. Zheng, A.H. MacDonald, Phys. Rev. B 48 (1993) 8203.
[15] K. Flensberg, B.Y.-K. Hu, Phys. Rev. Lett. 73 (1994) 3572.
[16] T.J. Gramila, J.P. Eisenstein, A.H. MacDonald, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 66 (1991) 1216.
[17] T.J. Gramila, J.P. Eisenstein, A.H. MacDonald, L.N. Pfeiffer, K.W. West, Phys. Rev. B 47 (1993) 12 957.
[18] N.P.R. Hill, J.T. Nicholls, E.H. Linfield, M. Pepper, D.A. Ritchie, A.R. Hamilton, G.A.C. Jones, J. Phys. Condens. Matter 8 (1996) L557.
[19] H. Rubel, A. Fischer, W. Dietsche, K. von Klitzing, K. Eberl, Phys. Rev. Lett. 78 (1997) 1763.
[20] D.S. Kainth, D.R. Richards, H.P. Hughes, M.Y. Simmons, D.A. Ritchie, Phys. Rev. B 57 (1998) R2065.
[21] K. Flensberg, B.Y.-K. Hu, A.-P. Jauho, J.M. Kinaret, Phys. Rev. B 52 (1995) 14796.
[22] K. Güven, B. Tanatar, Phys. Rev. B 56 (1997) 7535.

Phase relaxation of electrons in disordered conductors

B.L. Altshuler [a,b,*], M.E. Gershenson [c], I.L. Aleiner [d]

[a] *NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA*
[b] *Physics Department, Princeton University, Princeton, NJ 08544, USA*
[c] *Serin Physics Laboratory, Rutgers University, Piscataway, NJ 08855, USA*
[d] *Physics and Astronomy Department, SUNY at Stony Brook, Stony Brook, NY 11794, USA*

Abstract

Conduction electrons in disordered metals and heavily doped semiconductors at low temperatures preserve their phase coherence for a long time: phase relaxation time τ_φ can be orders of magnitude longer than the momentum relaxation time. The large difference in these time scales gives rise to well-known effects of weak localization, such as anomalous magnetoresistance. Among other interesting characteristics, study of these effects provide quantitative information on the dephasing rate $1/\tau_\varphi$. This parameter is of fundamental interest: the relation between $\hbar/\tau_\varphi$ and the temperature T (a typical energy scale of an electron) determines how well a single electron state is defined. We will discuss the basic physical meaning of $1/\tau_\varphi$ in different situations and its difference from the energy relaxation rate. At low temperatures, the phase relaxation rate is governed by collisions between electrons. We will review existing theories of dephasing by these collisions or (which is the same) by electric noise inside the sample. We also discuss recent experiments on the magnetoresistance of $1D$ systems: some of them show saturation of $1/\tau_\varphi$ at low temperatures, the other do not. To resolve this contradiction we discuss dephasing by an external microwave field and by nonequilibrium electric noise. © 1998 Published by Elsevier Science B.V. All rights reserved.

Keywords: Phase relaxation of electrons; Disordered conductors

1. Introduction

Anomalous magnetoresistance in disordered conductors (doped semiconductors and metals) has been recognized for almost 50 yr [1]. For a long time this phenomenon has remained a puzzle. The theoretical understanding of the anomalous magnetoresistance emerged as a spinoff of the theory of Anderson localization. It turned out that the correction to the conductivity, which is due to quantum interference at large length scales, is very sensitive to weak magnetic fields. The quantum correction itself may be much smaller than the classical conductivity. Nevertheless, the weak field magnetoresistance is dominated by this correction and its basic features (its amplitude, dependence on both magnitude and direction of the magnetic field, etc.) are very different from that of the classical magnetoresistance. Since the quantum correction can eventually drive the system to the Anderson insulator, the regime when this correction is small, is called the weak localization (WL) regime,

* Corresponding author. Fax: +1-609-951-2496; e-mail: bla@research.nj.nec.com.

and the theory of the anomalous magnetoresistance is now a part of the theory of weak localization.

A qualitative physical interpretation of WL is usually based on estimation of the probability $P(t)$ to find a particle at given time t at the same point where it was at time $t = 0$. The quantum correction to this return back probability are due to the interference between two amplitudes to return along the same classical path in the opposite directions. The quantum correction is sensitive to the magnetic field that through the loop formed by the classical self-returning path [2]. High sensitivity is associated with the fact that the typical area S of this loop is large. For large loops, the area S is proportional to the length of the path, i.e., to the time t it takes for an electron to circle the loop. Indeed, the motion is diffusive provided t exceeds the elastic mean free time τ. This means that the typical size of the loop is of the order of $(Dt)^{1/2}$ and $S(t) \simeq Dt$, where $D = v_F^2 \tau / d$ is the diffusion constant (d is the number of dimensions, v_F is the Fermi velocity of electrons).

Let us consider DC resistance of an "infinite" system. In this case, there are two length scales which determine the relevant size of the return path. The first one is associated with the magnetic field, while the second is determined by inelastic collisions of electrons.

The effect of the magnetic field on the conductance is a manifestation of the Aharonov–Bohm effect [3]. Let us consider the amplitude $A(t) = \sum_j A_j(t)$ of the return probability $P(t) = |A(t)|^2$, index j here labels the classical path. In the absence of the magnetic field, the two paths j and $\bar{j}$ which differ only by direction are characterized by the same classical action. Hence, in the semiclassical approximation their contribution to $A(t)$ have identical phases. This leads to constructive interference, to an enhancement of $P(t)$ as compared with its classical value $(Dt)^{d/2}$. As a result, this interference reduces diffusion constant D and increases resistance R.

The equality of these two phases is a direct consequence of the T – invariance of the system, and is therefore violated by the magnetic field. The difference between the two phases is determined by the magnetic flux encompassed by the returning trajectories $\Phi_{(j)}(t) = H S_j(t)$

$$\phi_j - \phi_{\bar{j}} = 2\pi \frac{\Phi_{(j)}}{\Phi_0}, \quad \Phi_0 = hc/2e, \tag{1}$$

where S_j is the directed area swept by the jth trajectory. Thus, the interference contribution associated with jth path acquires the oscillatory factor $\cos(2\pi\Phi_{(j)}/\Phi_0)$. Since S_j are random, all the contributions from trajectories sweeping typical area larger than $\simeq \Phi_0/H$ are diminished. As a result, the first characteristic length scale is the magnetic length $L_H = \sqrt{\hbar c / eH}$.

Magnetoresistance will be determined by the classical trajectories sweeping the area of the order of L_H^2, provided electrons preserve the phase coherence during the corresponding time L_H^2/D. Elastic collisions do not effect this phase coherence, while inelastic interactions of the electron with environment (other electrons, phonons, etc.) tend to destroy it. One can introduce a length scale L_φ which corresponds to substantial suppression of the coherence. This scale and corresponding time $\tau_\varphi = L_\varphi^2/D$ depends on temperature T, and plays an important role in the description of weak localization phenomena. Since the magnetoresistance depends on the ratio of L_φ and L_H, and L_H can be tuned by changing the magnetic field, magnetoresistance can be used for the study of inelastic collisions and dephasing of electrons at low temperatures.

A conductor can be considered as infinite for our purposes as long as the distance between the leads is much larger than at least $L_{\min}$ which is the smallest of the two characteristic scales L_φ and L_H. It does not mean however, that all of the dimensions of the system should be large. The sample has a dimension $d = 1, 2, 3$ depending on the relation between its transverse dimensions and $L_{\min}$. We can speak about $1d$ wires ($2d$ films) when both thickness of the wire a and its width W (thickness of the film a) is smaller than $L_{\min}$, despite the fact that both a and W are far in excess of the electronic wavelength k_F^{-1}. Since in this lecture we discuss recent experiments on magnetoresistance of $1d$ metallic wires and semiconductor structures, we will consider here only the $1d$ case, though some of the statements are of a more general validity.

The paper is organized as follows. In Section 2, we discuss qualitatively the main theoretical ideas on the dephasing in disordered systems. Readers interested in rigorous derivations are urged to consult Ref. [6]. Section 3 reviews recent observations of the

saturation of τ_φ at low temperatures Ref. [17]. Authors of Ref. [17] have suggested an explanation based on the dephasing by the zero point motion of electrons. We think that this explanation is erroneous and elaborate on this issue more in Section 4. Section 5 describes new experiments in which no saturation was observed, though the values of τ_φ were much longer than the universal "cut-off" suggested. As an explanation of the saturation in Ref. [17], proposed in Section 6 the dephasing by the external microwave radiation, and we show in Section 7 that the microwave radiation can efficiently dephase electron without significant heating.

2. Inelastic e–e collisions and dephasing rate

In early papers (see e.g. Ref. [4]) on the theory of localization, the dephasing rate $1/\tau_\varphi$ was considered to be of the same order as the inelastic collision rate in perfectly clean conductors. The latter can be expressed as the sum of the electron – phonon $1/\tau_{e-ph} \simeq T^3/\Theta_D^2$ and electron – electron $1/\tau_{e-e} \simeq T^2/E_F$ contributions, where E_F and Θ_D are the Fermi and Debye energies correspondingly (here and almost everywhere below we put $\hbar = 1$ and $k_B = 1$). Even under this assumption, the e–e contribution dominates at low enough temperatures. It became clear later that static disorder strongly enhances the e–e contribution to the inelastic scattering rate [5,6], while $1/\tau_{e-ph}$ is less affected [7]. As a result, both dephasing and energy relaxation rates at low temperatures are governed by collisions between electrons.

To recall the main results on the e–e dephasing rate, let us start with a single electron excitation, assuming that $T = 0$, and the rest of the electron gas occupies states below the Fermi level. Dependence of $1/\tau_{e-e}$ on the energy of the excitation (energy of an electron counted from the Fermi level) ε can be determined in a perturbative calculation [5,6,8]. The result (Eq. (4.4) of Ref. [6]) can be rewritten through the dimensionless conductance $g(L)$ (conductance measured in units of $e^2/h \simeq 1/25.8\,\mathrm{k\Omega}$) of a d-dimensional cube of the size L

$$\frac{1}{\tau_{e-e}(\varepsilon)} = C_d \frac{\varepsilon}{g(L_\varepsilon)}, \quad L_\varepsilon \equiv \sqrt{D/\varepsilon}, \tag{2}$$

where C_d is the dimension-and-coupling-constant-dependent coefficient. For a weakly interacting $1d$ electron gas $C_d = \sqrt{2}$. Eq. (2) can also be rewritten as

$$\frac{1}{\tau_{e-e}(\varepsilon)} = C_d \delta_1(L_\varepsilon), \tag{3}$$

where $\delta_1(L) = (L^d v)^{-1}$ is a one-particle mean level spacing in a d-dimensional cube of the size $L \cdot \delta_1$ is determined by one-particle density of states v.

There are several interpretations of this result. One of them [6] is based on the concept of the interaction time which becomes much longer in the disordered case due to diffusive rather than ballistic motion of electrons. It is also possible to appeal to statistical properties of exact one-electron wave functions [9–11], and we outline this interpretation below.

Inelastic rate $1/\tau_{e-e}(\varepsilon)$ is determined by a pair of collisions between electrons with all four energies – two initial ($\varepsilon > 0$ and $\varepsilon' < 0$) and two final ($\varepsilon - \omega > 0$ and $\varepsilon' + \omega > 0$) – belonging to the energy strip with the width 2ε centered at the Fermi level, all the energies here are counted from the Fermi level.

Given the typical absolute value $M(L, \omega, \varepsilon, \varepsilon')$ of the matrix element for such a collision in a sample with a size L, the inelastic rate can be estimated [12,10] with the help of the Fermi Golden Rule

$$\frac{1}{\tau_{e-e}} \propto \sum_{0 < \omega < \varepsilon} \sum_{-\omega < \varepsilon' < 0} \frac{M(L, \omega, \varepsilon, \varepsilon')^2}{\delta_1(L)}. \tag{4}$$

The matrix elements can be represented as integrals of products of four exact one particle wave functions. In a disordered system, these wave functions oscillate randomly in space, and are only weakly correlated with each other. As a result, the matrix elements are random and for L smaller than L_ε (0D case), their typical absolute value $M(L)$ turns out to be of the order of $\delta_1(L)/g(L)$, where the small factor g^{-1} reflects the weakness of the correlation between the wave functions. (In the limit $g \to \infty$ Random Matrix Theory is valid; according to this theory, there is no correlation at all between different eigenvectors and the non-diagonal matrix elements vanish.) Each sum in Eq. (4) leads to the factor $\sim \varepsilon/\delta_1(L)$. As a result, in 0D case $1/\tau_{e-e}$ can be estimated as [11]

$$\frac{1}{\tau_{e-e}} \simeq \frac{\varepsilon^2}{g^2 \delta_1(L)}. \tag{5}$$

It increases with ε, and at $L = L_\varepsilon$, the rate $1/\tau_{e-e}$ becomes of the order of $\delta_1(L_\varepsilon)$. This estimate corresponds exactly to Eqs. (2) and (3), and it remains valid even for large samples, $L > L_\varepsilon$, since $1/\tau_{e-e}$ cannot depend on L in this limit.

When making estimate (5) we assumed that $1/\tau_{e-e}$ in Eq. (2) is determined by the energy transfer ω of the order of ε. To take into account quasielastic processes with small energy transfer, let us find the dependence of the matrix element on the transmitted energy ω. From comparison of Eqs. (3) and (4) this dependence in d-dimensional sample reads

$$M^2 \sim \frac{\delta_1(L)^3 \delta_1(L_\omega)}{\omega^2} = \frac{\delta_1(L)^4 L^d}{\omega^2 L_\omega^d} \propto \omega^{-2+d/2}. \quad (6)$$

This energy dependence of the matrix element reflects the properties of noninteracting disordered system and is not sensitive to the distribution function.

We see from Eq. (6) that the matrix elements diverge when $\omega \to 0$ for $d < 4$. At $T = 0$ this divergence is not dangerous because of two summations in Eq. (5). However, the situation changes when the temperature is finite. In this case $|\varepsilon'|$ in Eq. (5) is determined by $\max\{T, \omega\}$, and at $\omega < T$ summation over ε' can be substituted by the factor $T/\delta_1(L)$. According to Eq. (6), it means divergence of the sum over ω in the lower limit for $d = 1, 2$. Therefore, $1/\tau_{e-e}$ is ill-defined at finite temperatures and in low dimensions [13].

This is not a catastrophe, though: $1/\tau_{e-e}$ itself has no physical meaning. When the energy relaxation rate $1/\tau_\varepsilon$, (i.e. the inverse time of thermalization of an excitation with energy ε much larger than temperature T) is considered, the quasielastic processes are not important. Therefore, Eqs. (2) and (3) give a good estimate of $1/\tau_\varepsilon$. The phase relaxation rate $1/\tau_\varphi$ is more delicate and requires additional consideration [14,6], since it involves the electron with typical energy T, which quasielastic scattering rate is divergent.

An additional phase caused by an inelastic collision is just a product of the energy transfer ω and the time t that passed after the collision. It means that collisions with arbitrary small ω, which give negligible contribution to $1/\tau_\varepsilon$, can cause dephasing, provided the phase is detected over a sufficiently long time. It is clear that the typical observation time t just cannot be larger than the dephasing time itself τ_φ. Therefore

in cases when $1/\tau_{(e-e)}$ diverges ($d = 1, 2$), the divergence should be cut off by $\omega \sim 1/\tau_\varphi$. As a result, instead of Eqs. (2) and (3), we obtain a self-consistent equation for τ_φ and L_φ:

$$\frac{1}{\tau_\varphi(T)} = C_d \frac{T}{g(L_\varphi)} = C_d \delta_1(L_\varphi) T \tau_\varphi, \quad L_\varphi \equiv \sqrt{D \tau_\varphi}. \quad (7)$$

Solving Eq. (7) we find [6,14]

$$\frac{1}{\tau_\varphi} = (T^2 \Delta_\xi)^{1/3}, \quad L_\varphi = \xi \left(\frac{\Delta_\xi}{T}\right)^{1/3} \quad (8)$$

in 1D case. Here energy $\Delta_\xi = D/\xi^2 = 2\pi^2 \delta_1(L)/g(L)$, has the meaning of the level spacing on the localization length ξ (we assume $g(L) \gg 1$). Corresponding result in two dimensions is

$$\frac{1}{\tau_\varphi} = \frac{T}{g} \ln g.$$

Therefore, $\tau_\varphi \to \infty$, when $T \to 0$. It should be noted that Eq. (7) for τ_φ, as well as the WL theory as a whole, is valid when and only when $g(L_\varphi)$ is large. The dephasing length L_φ increases when $T \to 0$. As soon as it reaches the localization length ξ, Eq. (7) cannot be used any more: $g(\xi) \sim 1$ by definition. On the other hand, as long as $g(L_\varphi) \gg 1$, the system behaves as a Fermi liquid, since $T\tau_\varphi \gg 1$.

It is also important to emphasize that the condition of the validity of the WL approach, $g(L_\varphi) \gg 1$ *does not impose any restriction on the total conductance of the wire*, $g(L) = g(L_\varphi)(L/L_\varphi)$. E.g., the samples described in Section 5, $g(L) \simeq 3 \times 10^{-3}$, whereas $g(L_\varphi) > 3$, so that WL consideration is still applicable.

We presented here a rather simplified interpretation of old results, which were derived rigorously about 15 years ago and were re-derived later in several ways [15,16].

3. Experiments on gold wires

These old results have been recalled in connection with recent measurement of the WL magnetoresistance of 1D Au wires [17]. From this measurements authors extracted the temperature dependence

of the dephasing rate $1/\tau_\varphi(T)$. They have found that $\tau_\varphi(T)$, increases with cooling, when T is large enough, but at $T \sim 1\,\mathrm{K}$ it saturates at a level of about 1 nanosecond.

In fact, such saturation has been observed by many experimental groups. However, this apparent contradiction with the theory was attributed to one of the two reasons. The saturation was explained either by overheating the electrons (due to applied voltage or to external noise) or by scattering of electrons by localized spins. Authors of Ref. [17] have demonstrated experimentally in a convincing and elegant way that both reasons for the $\tau_\varphi(T)$ saturation are not applicable for their samples.

First of all, they observed temperature dependence of the resistivity – probably, due to the effects of the interaction between electrons – at T as low as 40 mK. The very fact that such dependence does exist is a strong evidence for the electron gas to have the same temperature as the bath. The effect of paramagnetic impurities was ruled out by *adding* certain concentration of Fe into gold and observing how effect of these additional localized spins disappears with cooling due to the Kondo effect (inelastic cross-section on one-channel Kondo impurities $\propto T^2$ at T smaller than the Kondo temperature).

These arguments convinced the authors of Ref. [17] that the finite dephasing at $T = 0$ is a fundamental and unavoidable consequence of the interaction between electrons. In light of this experimental data the saturation of $\tau_\varphi(T)$ at $T \to 0$ appears to be a problem, so serious, that it inspired several attempts to reconsider the foundations of the theory of disordered conductors in the weak localization regime.

The main puzzle in experiments Ref. [17] is that $\tau_\varphi(T)$ saturates when conductance is still very large: $g(L_\varphi) \sim 10^3$. It means that corresponding zero-temperature dephasing length $L_\varphi(T = 0)$ is much smaller than the localization length ξ. Therefore, *assuming that this relation between L_φ and ξ always holds, one should conclude that any interaction between electrons rules out localization of quantum states in a weakly disordered wire with a finite cross-section.*

The authors of Ref. [17] came up with an analytical estimate of $L_\varphi(T = 0) = L_{\mathrm{MJW}}$. They argued that their experimental data, as well as all other data available are consistent with the estimate, $L_\varphi(T = 0) \sim L_{\mathrm{MJW}}$.

For a wire with a thickness a and a width W their expression for L_{MJW} can be rewritten in the form

$$L_{\mathrm{MJW}} = Wak_{\mathrm{F}} = N/k_{\mathrm{F}} = \xi/(k_{\mathrm{F}}l), \qquad (9)$$

where k_{F} is Fermi wavenumber, $N = Wa/k_{\mathrm{F}}^2$ is the number of channels in the wire, and l is the mean free path of the electrons. Note that in the WL regime $k_{\mathrm{F}}l \gg 1$, and the localization length ξ can be estimated as $\xi \sim lN$. Assuming that Eq. (9) gives correct upper limit for L_φ, one concludes that interactions between electrons prevent localization, provided $N > 1$ and $k_{\mathrm{F}}l > 1$.

In fact, one can discuss a possibility that for some reason interaction between electron, in addition to inelastic dephasing, causes static violation of the T-invariance, e.g. orbital ferromagnetism. This violation would saturate the temperature dependence of the magnetoresistance. In this hypothetical case, localization is possible and would be similar to the 1D localization of non-interacting electrons in the presence of magnetic field (unitary ensemble). It would mean, though that the localization length does not depend on the magnetic field at all (in contrast to the experimental evidence [23], see below).

Therefore, assuming that the fact that L_φ saturation is a fundamental law, one ends up with the conclusion that 1D localization either does not exist in WL regime, or it is magnetic-field-independent.

4. Nyquist–Johnson noise and zero-point oscillations

Dephasing caused by the electron–electron quasielastic collisions Eq. (7), which we interpreted through general properties of matrix elements, can be also understood from a slightly different point of view [14]. Instead of thinking about many colliding electrons, we can consider one electron, which while moving around the loop, is subject (in addition to the quenched disordered potential) to a random time- and space-dependent electric field. This field is created by the rest of the electron gas and is nothing but the *equilibrium electric noise* inside the conductor. The advantage of this approach is that correlation functions of this field at large times and distances are determined solely by the conductivity of the system.

It turned out to be possible to take the dephasing effect of this noise into account in a non-perturbative way, and to determine the quantum correction to the conductivity $\sigma(H, T)$ as a function of magnetic field and temperature. In 1D case

$$\delta g(L > L_\varphi) \propto \frac{L_\varphi}{L} \frac{1}{\left[\ln \mathrm{Ai} \left(\frac{(W L_\varphi H)^2 r}{\varphi_0^2} \right) \right]'}, \qquad (10)$$

where $[\ln \mathrm{Ai}(x)]'$ is the logarithmic derivative of the Airy function, L_φ is determined by Eq. (8) and r is geometry-dependent coefficient of order unity.

Since Eq. (8) is in qualitatively contradiction to the experimental results, authors of Ref. [17] made an attempt to explain dephasing at $T = 0$ by zero-point oscillations of the electric field [18], i.e. they have assumed that dephasing is determined by the processes with the *energy transfer ω much larger than temperature*. Their consideration resulted in the length scale L_{MJW}.

We do not believe that zero point oscillations can cause any dephasing. Indeed, one can naively consider any environment as a set of harmonic oscillators. This set is characterized by the distribution of frequencies and couplings with a given quantum particle [19]. A collision between the particle and an oscillator is inelastic, provided the particle either transfers energy to the oscillator and excites it, or receives energy from the oscillator. The energy of the zero-point oscillations cannot be transferred, since this energy is simply a difference between the ground state energy of the oscillator and the bottom of the harmonic potential. At T much smaller than the frequency of the oscillator ω_l, inelastic collisions are impossible: the oscillator is in the ground state, and the particle does not have enough energy to excite the oscillator. Therefore, with probability exponentially close to unity, the collision is elastic and the oscillator has as little chance to cause dephasing as any static impurity.

We have to address the question again: why is the experimentally observed dephasing length is *always* (as it is pointed out in Ref. [17]) smaller than L_{MJW}? The answer is: it is *not always the case*! In the next section we briefly discuss experiments where dephasing lengths much larger than L_{MJW} have been observed.

5. New data on L_{MJW} decoherence in 1D δ-doped GaAs wires

Recently, new data on the temperature dependence of L_φ have been obtained for sub-micron-wide "wires" fabricated from the δ-doped GaAs structures [23]. In these samples, a single δ-doped layer with concentration of Si donors $N_{\mathrm{D}} = 5 \times 10^{12} \, \mathrm{cm}^{-2}$ is $0.1 \, \mu\mathrm{m}$ beneath the surface of an undoped GaAs. The 1D wires were fabricated by electron beam lithography and deep ion etching. A 50 nm thin silver film deposited on top of the structure was used as a "gate" electrode: the electron concentration n and the resistance of the samples can be "tuned" by varying the gate voltage V_{g} (for more details, see [23]). Below we discuss the data obtained for the sample comprising 360 wires connected in parallel; the length L of each wire is 500μm, the effective wire width $W = 0.05 \, \mu\mathrm{m}$. Relatively high concentration of carriers ensures that the number of occupied 1D sub-bands is large (~ 10) in the wires. The mean free path of electrons l increases with n from 17 to 58 nm; ($k_{\mathrm{F}} l \approx 6$–$30$, where k_{F} is the Fermi wave number). The sample is *one-dimensional* with respect to the quantum interference effects at low temperatures: $W < L_\varphi(T) \lesssim \xi$. These 1D conductors demonstrate the Thouless crossover [4] from weak localization (WL) to strong localization (SL) with decreasing the temperature [24,26]; the crossover temperature T_0 can be varied over a broad range by the gate voltage. In strong magnetic fields, the crossover "shifts" toward lower temperatures; this shift is accompanied with *doubling* of the localization length, and *halving* of the hopping activation energy in the SL regime [24,26].

Both the temperature and magnetic field dependences of the resistance of these samples are consistent with the theory of WL and interaction effects on the "metallic" side of the crossover ($T > T_0$) [23]. The phase coherence length L_φ has been estimated from the WL magnetoresistance; the procedure of extraction of L_φ for 1D conductors has been described in detail in Refs. [23,25]. The dependences $L_\varphi(T)$ are shown in Fig. 1 for different values of the gate voltage V_{g}. The experimental values of L_φ are well described Eq. (8). over the whole temperature range that corresponds to the WL regime. The theoretical dependences Eq. (8) are extended in Fig. 1 down to the crossover temperature. The dependence $L_\varphi(T)$ *do*
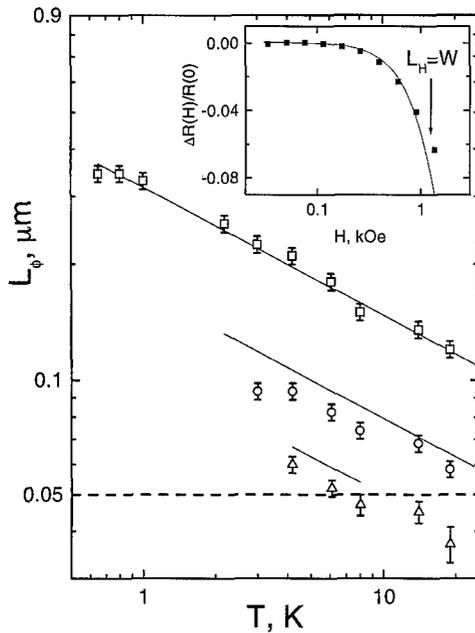
Fig. 1. The phase coherence length versus temperature for a δ-doped GaAs wire (for parameters, see text) at different V_g: □ – 0.7 V; ○ – 0 V; △ – 0.35 V. Solid lines – Eq. (8), the dashed line – Eq. (9). The inset shows the magnetoresistance at $T = 8$ K, $V_g = 0$ V, the solid line – the WL theory fit.

not saturate down to the crossover temperature and the quasiparticle description holds over the *whole* WL temperature range.

The observed dependences $L_\varphi(T)$ *argues against* the idea of the decoherence due to zero-point fluctuations of the electric field [17]. Indeed, Eq. (9) implies that (a) narrow channels fabricated from 2DEG *cannot* demonstrate 1D quantum corrections to the conductivity, since for a strip of 2DEG, L_{MJW} simply equals to the trip width W, and (b) the localization-induced crossover *should not* be observable in such channels (L_φ, being limited by $L_0 = W$, is always much smaller than $\xi \gg W$ in this case). The existence of the "cut-off" time L_{MJW}^2/D would also preclude observation of the *interaction-driven* 1D crossover with decreasing the temperature. Indeed, as soon as τ_φ approaches τ_0, the broadening of the electron energy levels, $\hbar/\tau_\varphi$, becomes temperature-independent. It is worth noting that at $D\hbar/L_{MJW}^2 > k_B T$ the Fermi-liquid description should also break down.

Both consequences of Eq. (9) contradict available experimental data. First, the 1D corrections to the

resistance of narrow channels fabricated from 2DEG have been studied for more than a decade. Second, the Thouless crossover has been observed in δ-doped wires with the ratio ξ/W as large as 16 [23]. For the data shown in Fig. 1, L_φ near the crossover exceeds the estimate L_0 by a factor of ~ 7 for $V_g = 0.7$ V; hence, the maximum experimental values of τ_φ exceed L_{MJW}^2/D by a factor of 50. The experimental values of τ_φ an order of magnitude greater than the estimate Eq. (9) have been also observed in Ref. [27]. Thus, there is a strong experimental evidence that saturation of the $L_\varphi(T)$ dependence is not intrinsic and the mechanism of this saturation is not related to the zero-temperature fluctuations of the electric field.

6. Dephasing by high-frequency radiation and external noise

We believe that saturation of the dependence $L_\varphi(T)$ observed in the experiments Ref. [17] is due to phase breaking by the *external* microwave electromagnetic noise. In order to explain our viewpoint, let us recall another old story about weak localization. Suppose we have applied an AC electric field to our sample (microwave radiation with some frequency Ω). The question we want to address is: how will this radiation affect DC conductance and its dependence on magnetic field? Obviously, the radiation can heat the sample, and the temperature dependence of the conductance will transform into its dependence on the amplitude of the AC electric field E_{AC}. However, it turns out that the dephasing effect of radiation can be much more important than this heating [20].

The dephasing effect of the microwave radiation depends on both its amplitude E_{AC} and frequency Ω. According to Ref. [20] this mechanism of dephasing is not effective when frequency is too low or too high. At $\Omega \to \infty$ one can take the radiation into account using perturbation theory, which gives $1/\tau_\varphi \propto \Omega^{-2}$. The fact that very high-frequency radiation is not effective in dephasing is easy to understand since such field is averaged out in a course of diffusive motion of electron during time τ_φ.

In the opposite limit of low frequencies one can consider dephasing by the electric field, which is linear in time $E(t) \sim E_{AC}\Omega t$, since the phases, which cor-

respond to two directions of circling the loop remain equal even if a DC field is present. Consider a closed trajectory with the return time t. Let us divide this trajectory into small segments j; each segment has the length of the order of the elastic mean free path l. Let the electron circle around the loop in, say, clockwise direction pass segment j at the time moment t_j; then, the electron traveling along the same trajectory but in counterclockwise direction will pass the same segment at a moment $t-t_j$. Therefore, when passing the same segment in different directions, the electron will acquire the different energies; the difference in this energies $\delta\varepsilon_j$ can be estimated as

$$\delta\varepsilon_j = \alpha_j le[E(t_j) - E(t - t_j)] \simeq \alpha_j eE_{AC}\, l\Omega(2t_j - t).$$

$$(11)$$

Here α_j is a random number which depends on the angle that the electric field makes with the direction of the electron path over the region j $\langle\alpha_j\rangle = 0$; $\langle\alpha_i\alpha_j\rangle \simeq \delta_{ij}$. In order to find the total energy difference acquired by the electron during the time interval $[0, t_j]$, we have to add up contributions from different segments $\varepsilon_j = \sum_{0<t_i<t_j} \delta\varepsilon_i$ and in order to find the total accumulated phase, we have to integrate the energy difference over time

$$\delta\varphi(t) = \tau \sum_{0<t_j<t} \varepsilon_j = \tau \sum_{0<t_j<t} \sum_{0<t_i<t_j} \delta\varepsilon_i. \quad (12)$$

Accumulated phase (12) is the random quantity, and, therefore, it is represented by $\langle\delta\varphi^2\rangle$. Substituting Eq. (11) into Eq. (12) and averaging its square with the help of $\langle\alpha_i\alpha_j\rangle \simeq \delta_{ij}$, we obtain

$$\langle\delta\varphi(t)^2\rangle \simeq (eE_{AC}l\Omega t)^2\tau^2 \left(\frac{t}{\tau}\right)^3 = D(\Omega eE_{AC})^2 t^5.$$

$$(13)$$

Determining τ_φ from the equation $\langle\delta\varphi(\tau_\varphi)^2\rangle \sim 1$, we find that at small frequencies $1/\tau_\varphi$ increases with Ω:

$$\frac{1}{\tau_\varphi} \simeq D^{1/5}(\Omega eE_{AC})^{2/5} \quad (14)$$

All this consideration holds provided $\Omega\tau_\varphi \lesssim 1$. At larger frequencies, as we have already mentioned, the dephasing rate decreases very fast with frequency. Rate $1/\tau_\varphi$ reaches its maximum at a certain frequency

Ω_E, which is determined by E_{AC}

$$\left[\frac{1}{\tau_\varphi(\Omega)}\right]_{max} = \frac{1}{\tau_\varphi(\Omega_E)} \sim \Omega_E, \quad \Omega_E \approx (e^2 E_{AC}^2 D)^{1/3}.$$

$$(15)$$

These theoretical predictions have been verified experimentally on inversion channels at the Si surface [21] and on Mg films [22]. In these papers it was found that conductance depends substantially on the power of microwave radiation. The frequencies used were $\Omega = 9.1$ MHz [21] and $\Omega = 0.66, 3.61$ GHz [22]. This dependence had nothing to do with heating (in Ref. [21] the effect was opposite to that caused by heating) and was in a reasonable agreement with the theory [20].

It is clear that any external noise should have dephasing effect similar to one of the microwave radiation. *We believe that it is external nonequilibrium noise that causes saturation of the dephasing rate in the experiments* Ref. [17], while at $T > 1$ K the equilibrium (and, thus, temperature-dependent) Nyquist–Johnson noise determines τ_φ. In the next section we compare dephasing and heating effects of the external noise in the GHz frequency range for a particular case of the Au wires Ref. [17].

7. Dephasing and heating in Au wires

Let us first estimate power of the microwave radiation sufficient for phase breaking at the time scale τ_φ in a 1D conductor. We assume that the radiation is in the optimal frequency range $\Omega\tau_\varphi \sim 1$. As we have seen, it also means that $\Omega_E\tau_\varphi \sim 1$ or, according to Eq. (15), $e^2E_{AC}^2D\tau_\varphi^3 \sim 1$. The radiation dominates dephasing as soon as $1/\tau_\varphi$ given by this estimate exceeds the rate provided by the equilibrium noise $\sim (T^2D/\xi^2)^{1/3}$, see Eq. (8). For a wire with a length L and a large conductance $g(L) \gg 1$ this happens when $eE > T/[Lg(L)]$, since the localization length can be written as $\xi \sim Lg(L)$. In terms of the power $P_{AC} = (LE_{AC})^2/2R$, where $R = h[e^2g(L)]^{-1}$ is the resistance of the wire, this inequality takes the form

$$P_{AC} > P_\varphi = R\left(\frac{ek_BT}{\hbar}\right)^2 = \frac{2\pi}{\hbar g(L)}(k_BT)^2. \quad (16)$$

Here we restore the Planck and Boltzmann constants. Note, that this power is proportional to the total resistance of a wire. Because of a very large resistance of the δ-doped wires studied in [24,26,23] ($R(4\,\mathrm{K}) \simeq 9\,\mathrm{M\Omega}$ for a single wire at $V_\mathrm{g} = 0.7\,\mathrm{V}$ in Fig. 1), the microwave power required for decoherence in such samples is rather large: $P_\varphi \simeq 4 \times 10^{-9}\,\mathrm{W}$ at $T = 1\,\mathrm{K}$. However, for 1D Au wires with a small $R \simeq 0.3$–$1.8\,\mathrm{k\Omega}$ [17], P_φ should be smaller by four orders of magnitude. Note that the numerical coefficeint in Eq. (16) is sensitive to the spectrum of the microwave radiation, however, it suffices for our estimates.

Does this microwave power heat the wire? The answer depends on how efficiently the extra energy is removed from the sample. One can propose two mechanisms of cooling: (a) phonon emission and (b) heat flow along the sample into "cold" leads. Let us start with the second mechanism, which is more important at low temperatures even for rather long wires. The expression for the power removed from the wire due to the hot-electron out-diffusion can be obtained using the Wiedemann–Franz law [28]. Let T be the temperature of the leads, and T_e the temperature of electrons in the wire. (The electron–electron interaction is sufficiently strong in thin metal films at low temperatures to ensure thermalization of electrons and to justify the approximation of local electron temperature [31]). Calculations of the temperature rise for different power levels and sample parameters can be found in Ref. [29]: for a small absorbed power, the difference $\Delta T = T_\mathrm{e} - T$ has a parabolic profile along the wire, with the peak equal to 1.5 of ΔT averaged along the wire. The estimate of the heat flow out of both ends of the wire into the "cold" leads for small $\Delta T \ll T$ has been done by Prober [28]:

$$P_\mathrm{es} \simeq \left(\frac{2\pi k_\mathrm{B}}{e}\right)^2 \frac{T\Delta T}{R} = 2\pi g(L) \frac{k_\mathrm{B}^2 T\Delta T}{\hbar}. \tag{17}$$

Comparison of Eq. (16) with Eq. (17) shows that at overheating is very small at $P_\mathrm{AC} \sim P_\varphi$ even in the absence of other cooling mechanisms:

$$\frac{\Delta T}{T} = g(L)^{-2}. \tag{18}$$

Thus, for any sample with $g(L) \gg 1$, *the microwave radiation can efficiently destroy the phase coherence*

of the electron wavefunction without heating the electron gas.

In fact, formula (18) even overestimates ΔT for long wires. In this case, the contribution of the phonon emission becomes dominant. For a sufficiently thin film, the "bottleneck" for the energy transfer from electrons to the thermal bath at low temperatures is formed by the electron–phonon interaction: the nonequilibrium phonons escape ballistically into the substrate [30,31]. In this case, the rate at which energy flows out of the electron gas by phonon emission, $P_\mathrm{e-ph}$, is given by the expression:

$$P_\mathrm{e-ph} = \frac{C_\mathrm{e}}{\tau_\mathrm{eph}(T)} \Delta T. \tag{19}$$

Here $C_\mathrm{e} = (aWL)\gamma T_\mathrm{e}$ is the heat capacity of the electron gas in a wire of volume aWL, γ is the Sommerfield parameter ($\simeq 70\,\mathrm{J/m^3 K^2}$ for Au), and τ_eph is the inelastic electron–phonon scattering time. Equation (19) is valid when $\Delta T = T_\mathrm{e} - T$, where T is the temperature of equilibrium phonons, is much smaller than both T_e and T. For an estimate of τ_eph, we can use recent results for the electron–phonon scattering time in thin Au films: $\tau_\mathrm{eph}(T) \simeq 1\,\mathrm{ns} \times (1\,\mathrm{K}/T)^2$ [32].

We believe that the experimental results reported in Ref. [17] can be explained by the nonequilibrium external noise. To be specific, let us estimate typical values of P_φ, P_es, and $P_\mathrm{e-ph}$ for sample Au-2 [17]: a gold film (thickness $a = 60\,\mathrm{nm}$, width $W = 110\,\mathrm{nm}$, length $L = 207\,\mathrm{\mu m}$) with the resistance $R = 302\,\Omega$ and diffusion constant $D = 612\,\mathrm{cm^2/s}$. The temperature dependences of P_φ as well as $P_\mathrm{e-ph}$ and P_es calculated for $\Delta T = 0.3 T_\mathrm{ph}$ are shown in Fig. 2; note that P_φ is in the sub-picowatt range at $T \leqslant 0.1\,\mathrm{K}$. At low-temperatures, electron diffusion is the process controlling energy flow out of the electron gas, while at higher temperatures the phonon emission dominates the electron gas cooling. For conditions of the experiment [17], balancing of the noise power P_φ, sufficient for phase breaking, by the outcoming power due to phonon emission and hot-electron out diffusion, $P_\mathrm{e-ph} + P_\mathrm{es}$, corresponds to a *negligible* increase of the electron temperature. In this situation, the rf noise can efficiently *destroy* the phase coherence of the electron wave function *without* heating the electron gas. This ex-

Table 1
Parameters of the samples studied in Ref. [17], (Au−1–Au−4) and the samples described in Section 5 (GaAs). R is the resistance of the sample, L is its length, $g = 25.8\,\mathrm{k}\Omega/R$ is the dimensionless conductance, τ_φ^{sat} is the experimental saturation value of the dephasing time claimed in Ref. [17], V_{DC} is the measuring DC bias applied to the sample, P_{DC} is the corresponding DC power, V_φ and P_φ is, respectively, the AC voltage and power at optimal frequency $\Omega \simeq 1/\tau_\varphi^{sat}$ needed to produce τ_φ^{sat}, see Eq. (14). P_φ and V_φ for GaAs sample not showing saturation is the estimate of the AC power required to affect the observed dependence

|	Au−1	Au−2	Au−3	Au−4	GaAs
R, Ω	1687	302	1443	1812	9×10^6
L, μm	57.9	207	155	57.9	500
$g(L)$	15.1	84.1	17.6	14.0	2.9×10^{-3}
$g(L_\varphi)$	160	1090	525	225	3–10
τ_φ^{sat}, ns	3.41	4.19	2.24	1.56	No saturation
V_{DC}, μV	8.37	14.57	14.68	8.64	50
P_{DC}, W	4.2×10^{-14}	7.0×10^{-13}	1.5×10^{-13}	4.0×10^{-14}	2.8×10^{-16}
V_φ, μV	2.0	2.0	8.8	6.4	2×10^4
P_φ, W	2.4×10^{-15}	1.4×10^{-14}	5.4×10^{-14}	2.2×10^{-14}	$>4 \times 10^{-9}$
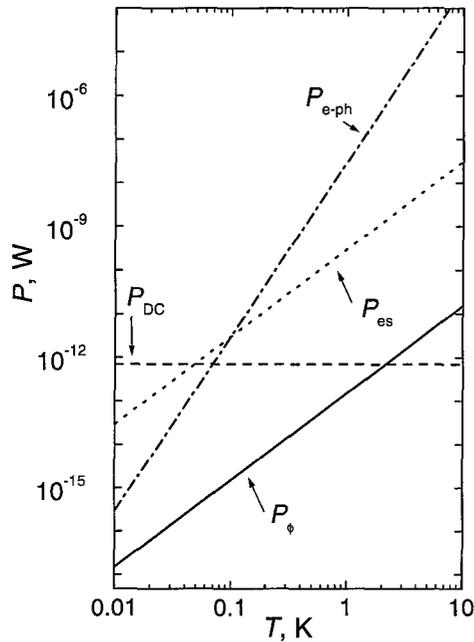
Fig. 2. The temperature dependences of P_φ, P_{e-ph}, and P_{es} for sample Au-2 [16]. The horizontal dashed line is the power of the DC current that has been used in measurements.

plains why a well-pronounced temperature dependence of the resistance (due to the interaction effects) has been observed at temperatures where the dependence $L_\varphi(T)$ was already completely saturated [17].

In fact, the dephasing power P_φ is even much smaller than the power dissipated by the DC voltage which was applied for the measurements (see

Fig. 2). Table 1 presents some parameters of the samples [17] and the noise power in the frequency range $f = 4 \times 10^7$–2×10^8 Hz, which would provide the experimentally observed dephasing rate $1/\tau_\varphi$. One can see that P_φ for most samples is about an order of magnitude smaller than P_{DC}, which is experimentally proven not to heat the wire.

8. Conclusion

In this lecture we argue that there is no need to modify the existing theory of dephasing in disordered conductors due to the electron–electron interactions. We have demonstrated that in δ-doped GaAs structures the theory describes both the temperature and magnetic field dependences of the conductivity in the weak localization regime up to the crossover to the strong localization.

We have proposed a mechanism that may be responsible for saturation of the dephasing rate at $T \to 0$ in many experiments. This is dephasing by the external non-equilibrium noise. It turns out that the dephasing effect of this noise is much stronger than its heating effect. Therefore, the temperature dependence of the conductivity observed in Ref. [17] *does not prove* the absence of this noise in the system.

In order to preserve the phase coherence for a long time, one has to reduce the noise amplitude in the frequency range $\Omega \sim 1/\tau_\varphi$ below a very low level. Such reduction could be a difficult technical problem at ultra-low temperatures.

Acknowledgements

References

[1] R.A. Chentsov, Zh. Exp. Teor. Fiz. 18 (1948) 374.
[2] B.L. Altshuler, P.A. Lee, Physics Today 41 (1988) 36.
[3] Y. Aharonov, D. Bohm, Phys. Rev. 115 (1959) 485.
[4] D.J. Thouless, Phys. Rev. Lett. 39 (1977) 1167.
[5] A. Schmid, Z. Phys. 271 (1973) 251.
[6] B.L. Altshuler, A.G. Aronov, in: A.L. Efros, M. Pollak (Eds.), Electron–Electron Interaction in Disordered Systems, North-Holland, Amsterdam, 1985.
[7] M. Reyzer, A.V. Sergeev, Sov. Phys. JETP 65 (1986) 616.
[8] B.L. Altshuler, A.G. Aronov, JETP Lett. 30 (1979) 482.
[9] Ya.M. Blanter, Phys. Rev. B 54 (1996) 12 807.
[10] B.L. Altshuler, Y. Gefen, A. Kamenev, L.S. Levitov, Phys. Rev. Lett. 78 (1997) 2803.
[11] V. Prigodin, B.L. Altshuler, preprint cond-mat/9703071.
[12] U. Sivan, Y. Imry, A. Aronov, Europhys. Lett. 28 (1994) 115.
[13] E. Abrahams, P.W. Anderson, P.A. Lee, T.V. Ramakrishnan, Phys. Rev. B 24 (1981) 6783.
[14] B.L. Altshuler, A.G. Aronov, D.E. Khmelnitskii, J. Phys. C 15 (1982) 7367.
[15] A. Stern, Y. Aharonov, Y. Imry, Phys. Rev. A 41 (1990) 3436.
[16] S. Chakravarty, A. Schmid, Phys. Rep. 140 (1986) 193.
[17] P. Mohanty, E.M.Q. Jarivala, R.A. Webb, Phys. Rev. Lett. 78 (1997) 3366.
[18] P. Mohanty, R.A. Webb, Phys. Rev. B 55 (1997) 13 452.
[19] A.O. Caldeira, A.J. Leggett, Phys. Rev. Lett. 46 (1981) 211.
[20] B.L. Altshuler, A.G. Aronov, D.E. Khmelnitskii, Solid State Commun. 39 (1981) 619.
[21] S.A. Vitkalov, G.M. Gusev, Z.D. Kwon, G.I. Leviev, JETP Lett. 43 (1986) 185; S.A. Vitkalov, G.M. Gusev, Z.D. Kwon, G.I. Leviev, V.I. Falko, JETP 67 (1988) 1080.
[22] Shiguang Wang, P.E. Lindelof, Phys. Rev. Lett. 59 (1987) 1156.
[23] Yu.B. Khavin, M.E. Gershenson, A.L. Bogdanov, to be published.
[24] M.E. Gershenson, Yu.B. Khavin, A.G. Mikhalckuk, H.M. Bozler, A.L. Bogdanov, Phys. Rev. Lett. 79 (1997) 725.
[25] P.M. Echternach, M.E. Gershenson, H.M. Bozler, A.L. Bogdanov, B. Nilsson, Phys. Rev. B 48 (1993) 11 516.
[26] M.E. Gershenson, Yu.B. Khavin, A.L. Bogdanov, Sov. Phys. – Uspekhi, February 1998.
[27] H. Linke, P. Omling, H. Xu, P.E. Lindelof, Phys. Rev. B 55 (1997) 4061.
[28] D.E. Prober, Appl. Phys. Lett. 62 (1993) 2119; D.E. Prober, Phys. Rev. Lett. 75 (1995) 3964.
[29] A. Mittal et al., Surf. Sci. 361/362 (1996) 537.
[30] T. J. Thornton et al., Phys. Rev. Lett. 56 (1986) 1198; S. Wind et al., Phys. Rev. Lett. 57 (1986) 633.
[31] E.M. Gershenzon et al., Sov. Phys.- JETP 59 (1984) 442.
[32] B. Belevtsev et al., Sov. Phys. Low Temp. Phys. 21 (1995) 646; N.G. Ptitsina et al., Phys. Rev. B 56 (1997) 10 089.

Scaling at the metal–insulator transition in two dimensions

Elihu Abrahams *

Serin Physics Laboratory, Rutgers University, 136 Freylinghuysen Road, Piscataway, NJ 08854-8019, USA

Abstract

The recently discovered two-dimensional metal–insulator transition will be reviewed and discussed in the light of the scaling theory of localization. It will be shown that an observed symmetry relating conductivity and resistivity follows directly from the critical behavior associated with such a transition. The status of both Fermi-liquid and non-Fermi-liquid approaches will be discussed. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Metal–insulator transition; Two-dimensional electron gas

1. Introduction

A number of recent experiments [1,2] on the two-dimensional electron gas in various semiconductor devices have revealed the existence of a continuous low-temperature metal-insulator transition as the electron density is varied. In the 1980s, similar experiments appeared to confirm a widely discussed prediction that the two-dimensional disordered electron system would have no metallic states [3]. In that same decade, the combined effects of disorder and electron–electron interaction were studied by renormalization group (RG) methods [4–7] and there were some indications that one effect of the interaction might be to stabilize a metallic state in two dimensions. However, the nature of the metallic state was unknown and in any case, the conclusion was not very definite since the RG flows went to a strong coupling regime beyond

the range of validity of the RG equations . Therefore, the possibility of a two-dimensional metallic state remained an open issue from the theoretical point of view [8].

The new experiments have stimulated renewed interest in the problem of the effects of disorder and electron–electron interaction in two dimensions. In this paper, I review the early theoretical work and the present experimental situation. This will be followed by a discussion of the status of scaling ideas and Fermi-liquid and non-Fermi liquid approaches.

2. Metals versus insulators

Let us be precise about the difference between a metal and an insulator. The distinction is properly made only at the absolute zero of temperature since thermal excitations permit an insulator to carry a current at non-zero temperature. If the resistance R is finite at $T = 0$ then the system is metallic; otherwise,

* Fax: +1-732-445-4400; e-mail: abrahams@physics.rutgers.edu.

Conduction band   Conduction band   Conduction band

E_F ----- — — — — ----- E_F -----

Valence band   Valence band   Valence band

a). $x = 0$ b). *small x*   c). *large x*

Fig. 1. Silicon valence and conduction bands.

it is insulating. Thus, it is the nature of the *ground state* which determines the behavior. Whether at some temperature the temperature coefficient of resistance dR/dT is positive ("metallic-like") or negative ("insulating-like") is not necessarily relevant. The metal–insulator transition is then a transition between two distinct ground states which is driven by varying some parameter. It is an example of a *quantum phase transition* – that is, a transition caused by a fundamental change of the ground state of a system as a parameter is varied.

3. Metal–insulator transition in a doped semiconductor

A simple example which illustrates these ideas is that of phosphorus-doped silicon. Undoped silicon is a *band insulator*. Its valence band is full, its conduction band empty as shown in Fig. 1a. If one adds a few phosphorus impurities, each one of which carries an extra electron which will not be accommodated in the valence band, one finds insulating behavior for low doping. Each extra electron remains bound to its phosphorus donor ion and the Fermi level E_F moves up into the region of these occupied localized states as shown in Fig. 1b. The donor concentration is denoted by x. At higher concentration, Fig. 1c, the previously localized electron wave functions begin to overlap and become extended. The Fermi level now lies within this

"impurity band" and the system is metallic. The combination of disorder and electron–electron interaction provides the physics underlying this metal–insulator transition as a function of x, the donor concentration.

4. Two mechanisms for electron localization

There are two routes towards electron localization. In 1949, Mott demonstrated [9] that the repulsion between electrons can produce a metal–insulator transition even in a system without disorder when the interactions are large enough compared to the kinetic energy bandwidth, the latter being determined by the overlap of the electron wave functions. Later, Anderson discovered [10] that disorder in the electrostatic potential, such as that caused by random impurities, can drive a metal–insulator transition even in a system of non-interacting electrons. The theoretical description of what happens when both effects are present is a central unsolved problem, especially when the electron–electron interactions are strong, as is the case in the new experiments.

5. Scaling at the metal–insulator transition

In a solid, it is the character of the electronic wave functions at the Fermi level which determines the electrical behavior. If the wave functions are exponen-

tially localized, the behavior will be insulating, while if they are extended to the edges of the sample, metallic behavior will obtain. In either case, the disorder will cause the wave functions to have characteristic bumps which have some characteristic size. If we examine the system on a length scale smaller than this size, it will be impossible to distinguish a metal from an insulator. Just as in the case of a thermodynamic phase transition, it is necessary to probe the system at scales larger than the characteristic length of wave function (or order parameter) fluctuations to distinguish the localized insulating (disordered) phase from the extended metallic (long-range order) phase.

6. Scaling for Anderson localization

In 1979, simple scaling arguments [3] were put forward which have shaped much of our thinking about metal–insulator transitions in disordered systems. Considering only the case of non-interacting electrons with disorder, the authors discussed what would happen if one examined wave functions at the Fermi level at larger length scales. They argued that a measure of localization is given directly by the conductance G ($= 1/R$) since that quantity is determined by the amplitude of the Fermi-level wave functions at the boundary of the sample, where the external leads are attached. Thus, for very small sample sizes, as discussed above, it is not possible to distinguish a metal from an insulator, but at large length scales, G is either large (metal) or exponentially small (insulator). In the former case, one expects $G = L^{d-2}\sigma$, where L is the linear size of a hypercubic sample in dimension d and σ is the conductivity, which is independent of length scale. For the insulator, one expects the wave functions typically to be exponentially small at the edge of the sample so that $G \sim \exp(-L/\xi)$, where ξ represents the characteristic size (localization length) of the typical Fermi-level wave function.

These ideas are expressed in terms of the so-called beta-function, defined as the logarithmic derivative of the conductance with respect to the length scale: $\beta(G) = \mathrm{d}\ln G/\mathrm{d}\ln L$. $\beta(G)$ may be evaluated in the two asymptotic limits described above. Then, at large G, $\beta(G) = d - 2$ and at small G, $\beta(G) = \ln(G)$. An interpolation between these limits gives the curves shown in Fig. 2 for different dimensionalities. It

was also shown that including the leading correction to $d - 2$ at large G in a perturbative treatment of the conductivity [11,12] leads to the expression $\beta(G) = d - 2 - a/G$, where a is a universal number. This form is consistent with the interpolated behavior. The physics behind this correction to ordinary Ohmic behavior is a suppression of the conductivity due to singular coherent backscattering of electrons in the disordered system. In particular, in two dimensions, where the conductance and the conductivity are identical, the result reads

$$G = \sigma = \sigma_0 - (e^2/h)(2/\pi)\ln(L/l), \qquad (1)$$

where σ_0 is the usual Drude conductivity associated with a mean free path l, L is the linear size of a square sample and h is Planck's constant.

The unexpected result of this analysis was that in two dimensions, no metallic states are expected since the beta-function is always negative. Thus, as scale size is increased toward the thermodynamic limit, G always decreases, eventually becoming exponentially small [$\beta(G) \sim \ln(G)$, insulating behavior]. In contrast, the beta-function in three dimensions has a zero so that depending upon the initial conditions at small length scale, the beta function may be positive (flow to metal) or negative (flow to insulator). Thus a metal–insulator transition at $d = 3$, but not at $d = 2$.

By means of computer experiments, beta-functions in one two and three dimensions were determined numerically for non-interacting electrons on finite lattices with disorder. The results agreed very well with the conclusions of the scaling picture [13].

7. Temperature sets the length scale

In experiments on real materials, it is not very convenient to continuously change the sample size (length scale). However, the temperature actually sets a scale. For example, in the scaling picture for Anderson localization it was realized that inelastic scattering would destroy the coherence of the back-scattering which is responsible for the enhancement of resistance. This would be equivalent to an effective length which increases as the inelastic scattering disappears as the temperature decreases. This behavior is generally true in the neighborhood of a quantum critical point. That
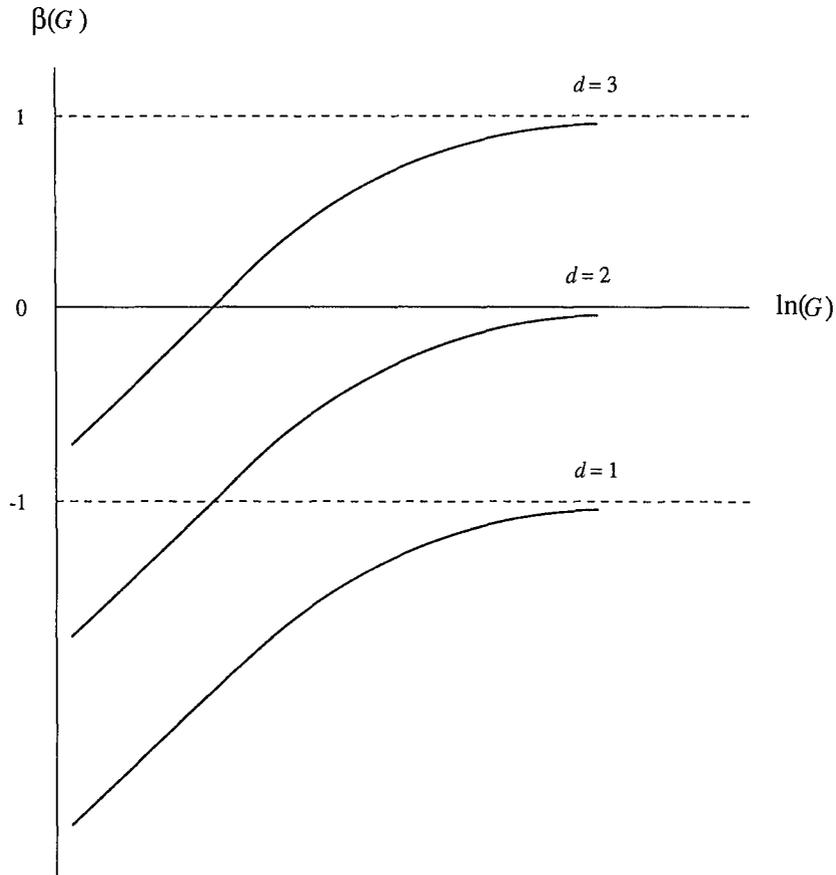
Fig. 2. Beta-functions.

is, there is an effective length determined by the temperature according to

$$L_T \propto T^{-1/z}, \tag{2}$$

where z is the "dynamical exponent". In what follows from Eq. (1), the temperature-dependent conductivity should be of the form

$$G(T) = \sigma(T) = \sigma_0 + (e^2/h)(2/\pi z) \ln[T/T_0(l)], \tag{3}$$

where $T_0(l) \propto l^{-z}$.

8. Electron–electron interaction

In the 1970s, the interplay of disorder and electron-electron interaction was already a subject of investigation notably by Altshuler and Aronov and by Schmid [19]. The methods they developed, in which the Coulomb interaction between electrons was treated perturbatively, were applied to calculate the conductivity in the presence of both disorder and interactions. In two dimensions, the result was an extra contribution to the conductivity of precisely the same form as that due to disorder alone, Eq. (3). The total expression for the conductance (now and henceforth expressed in units e^2/h) became

$$G = \sigma_0 + \frac{2}{\pi z} \ln(T/T_0) + \frac{1}{\pi}(1 - P) \ln(T\tau). \tag{4}$$

Here, τ is the elastic mean free time and P is coefficient which comes from disorder (diffusive) corrections to the Hartree interaction energy. In the perturbative treatment, P is always less than unity so that the interaction correction has the same sign as the

localization term and they both lead to insulating be-havior in two dimensions at low temperatures. There are experiments which attempt to separate the effect of the two terms [14]. Later, it was realized that the interaction effects are also scale dependent and in a se-ries of works pioneered by Finkel'stein, the Coulomb interaction and disorder were treated together in a renormalization group (RG) treatment [4–7]. We shall discuss this further below.

9. Experimental situation

Early experiments on semiconductor devices [14–16] and on amorphous metallic films [17,18] confirmed the predictions of the scaling analysis. The low-temperature behavior showed an *increase* in the resistance which was logarithmic in the temperature, consistent with Eq. (3). No evidence of a resistance which *decreased* with decreasing temperature was found. However, we now have evidence of the ap-pearance of metallic states in the low-density two-dimensional electron gas [1,2] and this has forced us to revisit the problem from the theoretical side.

The new experiments on silicon MOSFET [1,2] devices give the most dramatic evidence for metallic states at low temperature, but metallic behavior has been seen also in GaAs–AlGaAs hole gases [20,21], in p-type SiGe/Si quantum wells [22,23] and in n-type Si/SiGe structures [24]. In these systems, the carrier concentration x is varied continuously by adjusting a gate voltage. As the density increases through some critical density x_c the temperature co-efficient of resistance dR/dT changes from negative (insulating behavior) to positive (metallic behavior).

Attempts are made to present the data in a scaling form appropriate for a quantum critical point. This proves to be quite successful for the Si MOSFET sam-ples, somewhat less so for the others. The scaling form for the resistivity ρ as a function of concentra-tion $\delta_x = (x - x_c)/x_c$ and temperature is

$$\rho(\delta_x, T) = \rho_c f[T/T_0(\delta_x)], \tag{5}$$

where ρ_c is the value of the resistance right at the crit-ical concentration. Thus for each x, a T_0 is found such that the rescaled data falls on two common curves, one for the insulating examples $\delta_x < 0$ and another for the conducting ones $\delta_x > 0$.

When this was done a remarkable symmetry was found: The rescaling parameter was found to be inde-pendent of the *sign* of δ_x, with the form $T_0(\delta_x) \propto |\delta_x|^{\nu z}$, where ν is the localization length exponent and z is the dynamical exponent defined earlier. Furthermore, at not too low a temperature, the universal insulating curve is inversion symmetric with the universal metal-lic one, that is to say $\rho(\delta_x, T)/\rho_c = \rho_c/\rho(-\delta_x, T)$.

10. Phenomenological scaling analysis

The symmetry of the experimental results led Dobrosavljević et al. [25] to examine to what extent a simple scaling analysis might be applicable. They ar-gued that the presence of a metal–insulator transition implies that the effect of electron–electron interaction is to modify the beta-function for two dimensions so as to produce a zero at some value of the conduc-tance (as in three dimensions). Then the beta-function would be positive at G large, so that a metallic state could be achieved in two dimensions. Possible behaviors are shown in Fig. 3. Even if the beta-function approaches zero at large G, the conductance would continuously increase with scale size, that is, as the temperature is lowered. Furthermore, since at $d = 2$ the conductance and the conductivity are the same, the implication is that the conductivity would be scale dependent and that the resistivity would ac-tually scale to zero as $T \rightarrow 0$, a behavior which is not characteristic of a Fermi liquid. This point about non-Fermi liquid behavior is further emphasized by the observation that if the interactions were turned off, the metallic fixed point would disappear and the elec-tronic states would change completely: they would localize.

Given that the beta-function has a zero at a finite value of G, one can analyze the behavior in the *quan-tum critical region* associated with the metal–insulator transition. The transition occurs when the density is tuned to the critical value x_c at which point the con-ductance is G_c. Near to G_c, we may linearize the beta-function in the variable $t = \ln(G/G_c)$; the slope turns out to be the inverse of the localization length expo-nent

$$\beta(t) = \frac{dt}{d(\ln L)} \approx \frac{1}{\nu} t. \tag{6}$$
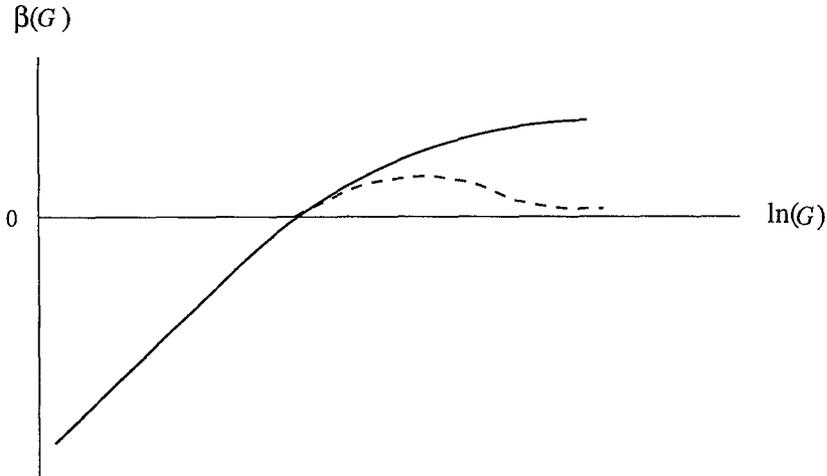
Fig. 3. Beta-function.

To find the behavior of the conductance as a function of length scale (temperature), we integrate Eq. (6) from l to L. We find

$$\ln[G(L)/G_c] = \ln[G_0/G_c] \left(\frac{L}{\ell}\right)^{1/v}, \tag{7}$$

where G_0 is the starting value of the conductance at small length scale (l), or high temperature. It is tuned by varying x near x_c, that is δ_x near 0. In the critical region, we start with a G_0 very close to G_c, so that $\ln[G_0/G_c] \approx (G_0 - G_c)/G_c \propto \delta_x$. Then from Eq. (7), we have for the conductance

$$G(L) = G_c \exp[A\delta_x(L/\ell)^{1/v}]. \tag{8}$$

We pass to non-zero temperature by the replacement $L \to T^{-1/z}$:

$$G(\delta_x, T) = G_c \exp \operatorname{sgn}(\delta_x)A[T_0(\delta_x)/T]^{1/vz}, \tag{9}$$

where we have defined a crossover temperature T_0

$$T_0(\delta_x) \sim |\delta_x|^{vz}. \tag{10}$$

Here, A is an unknown dimensionless constant of order one. Let us define the scaled conductance as $g^*(\delta_x, T) = G(\delta_x, T)/G_c$. From Eq. (9), we then immediately find the striking symmetry relating the conductance on the metallic side ($\delta_x > 0$) to the resistance on the insulating side ($\delta_x < 0$) of the transition

at $g^* = 1$ (i.e. at $\delta_x = 0$)

$$g^*(\delta_x, T) = 1/g^*(-\delta_x, T). \tag{11}$$

The behavior in Eqs. (10) and (11) is exactly that reported in the experiments [1,2].

Thus the critical behavior deduced from a simple beta-function gives an excellent description of the experiments. The scaling symmetries should disappear at low temperature, where one has already scaled out of the region in which the beta-function is linear. This is observed. As the slope of the beta-function decreases on the metallic side, one expects that the rate at which the resistivity increases in the insulator should be greater than the rate at which it decreases in the metal. This is also observed. Unlike the case of Anderson localization in the absence of disorder, one does not expect the critical conductance G_c to be universal since the interactions must be playing an important role. However, we may conclude that in the quantum critical region, simple scaling gives an excellent description of the data in Si MOSFET samples.

11. Less is different

It is necessary to understand what distinguishes the new samples which exhibit the $d = 2$ metal-insulator

Table 1

Type	Sample	x_c	ρ_c	r_s	$k_F l$	E_F (K)
n	Old MOSFET	2×10^{12}	0.04	4	13	140
n	Old GaAs	3×10^{11}	0.1	0.6	10	205
n	New MOSFET	9×10^{10}	3	19	0.2	6
p	Cambridge I	5×10^{10}	0.5	11	2	4.6
p	Cambridge II	1×10^{10}	0.4	21	2.5	1.2
p	Weizmann–Princeton	1×10^{10}	0.8	23	1.3	0.87

transition from the older ones which did not. The quantities of interest in the problem are the Coulomb energy, the Fermi energy and the disorder. A measure of the Coulomb energy is $V_C = e^2/(\kappa r_0)$, where κ is the dielectric constant and r_0 is the average interparticle spacing related to the particle density by $r_0 = 1/\sqrt{\pi x}$. The Fermi energy is given by $E_F = \hbar^2 k_F^2/2m^*$, where k_F is the Fermi wave number and m^* is the effective mass. In $d = 2$, the resistivity is proportional to the disorder in the sense that for weak disorder it is given (in units of h/e^2) by $1/(k_F l)$. The ratio V_C/E_F can then be expressed as

$$\frac{V_C}{E_F} = \frac{s}{a^*\sqrt{\pi x}} \equiv r_s. \tag{12}$$

Here, s is the number of valleys and a^* is the effective Bohr radius, given by $a^* = \hbar^2 \kappa/m^* e^2$. Up to the factor s, r_s is seen to be the interparticle spacing in units of the effective Bohr radius. Table 1 which follows shows representative values for these quantities for the different samples.

As has been pointed out [1,2], what distinguishes the new samples from the old is that the carrier density is *much* lower so that the Coulomb energy is by far the largest energy in the problem. It is therefore no surprise that the results obtained by treating the electron–electron as weak are not relevant in the new situation.

12. Scaling with interactions and disorder

For the "generic" situation (zero magnetic field, no spin or spin–orbit scattering), the RG treatment of disordered interacting electrons [4–7] indicates the pos-

sibility of a stable metallic state in two dimensions. This work has been reviewed recently in the present context by Castellani et al. [26]. The simplified discussion which follows is sufficient for our purposes.

The essential scaling variables are the disorder-averaged conductance or resistivity (as in the Anderson transition) and the disorder-averaged scattering amplitude in the electron–hole spin-triplet channel. What is found in the perturbative RG is that this scattering amplitude scales out of the perturbative region toward infinity with a consequent divergence in the disorder-averaged uniform spin susceptibility. This suggests the formation of local moments, or perhaps a tendency toward ferromagnetism, which might be expected in the low-density electron gas [27]. Unfortunately, as soon as the dimensionless scattering amplitude exceeds unity, the perturbative RG procedure fails. However, as the temperature is lowered, the resistivity first increases and then begins to decrease (just as the scattering amplitude becomes too large) which suggests the possibility that a low-temperature metallic state might be achieved. This behavior is seen in Fig. 4a where a typical scaling trajectory, obtained by integrating the RG equations [4,5] is shown. For different high-temperature (initial) conditions, slightly different curves are found. Temperature decreases in the direction of the arrow. There is evidence for just such behavior in the observed resistivity which first increases a bit as the temperature is lowered and then decreases or increases more rapidly, depending upon the sign of δ_x.

Because of the failure of the RG procedure, this possibility was never taken seriously until recently [26]. Furthermore, the expectation is that should local moments or ferromagnetism occur, there would be a crossover to a universality class in which all states are insulating.

Let us suppose, as suggested in Refs. [6,7], that a metal–insulator fixed point actually occurs in the space of triplet amplitude and resistance. Then there might be different behaviors separated by crossover regions as shown in Fig. 4b. A "metallic" and an "insulating" scaling trajectory are shown. It is clear that depending on the temperature regime investigated, there would be different values of the apparent critical resistivity and apparent critical concentration. This is in fact observed in experiments and is a realization of the remark made at the end of Section 10.
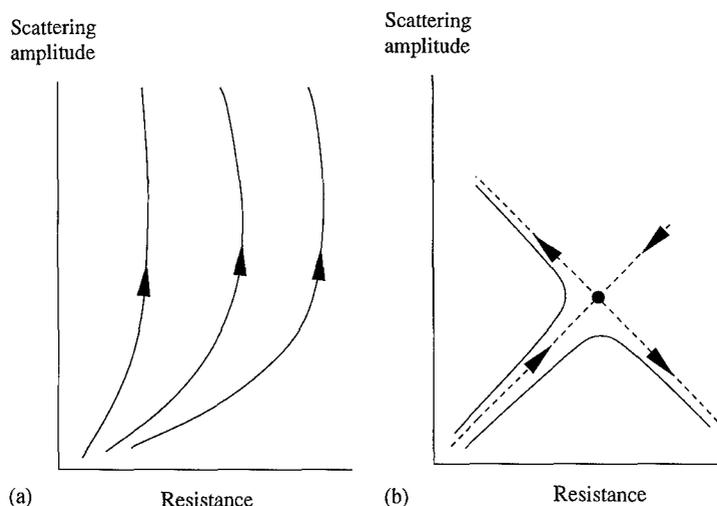
Fig. 4. Resistance–scattering amplitude curves.

13. Effect of magnetic field

There is little effect of the magnetic field on samples which are in the strongly insulating regime, that is those whose resistivity is increasing essentially exponentially with decreasing temperature. In contrast, in the metallic range of density, a magnetic field parallel to the plane of the sample gives a dramatic positive magnetoresistance so that at fields of a few Tesla, the previously metallic density situation becomes insulating [1,2,28,29]. In perpendicular field, for densities well on the metallic side, there is an initial negative magnetoresistance below 0.1 T followed by a large positive magnetoresistance which again leads to insulating behavior.

A parallel magnetic field will not substantially effect the in-plane orbital motion of the carriers so that it is believed that the fact that insulating behavior is restored in parallel fields of order $E_F/g\mu_B$ indicates that spin polarization of the carriers is an essential feature of the destruction of metallicity.

Within the scaling picture of Finkel'stein et al. [4,5], as emphasized in Ref. [26], the spin splitting in magnetic field destroys the metallic behavior which is suggested by the scaling trajectories of Fig. 4 and the system goes over to an insulating regime. No detailed comparison of experiment with this theory has yet been made.

14. A non-Fermi liquid?

As explained in Section 10, it was suggested in Ref. [25] that a metallic state in the two-dimensional disordered electron system might not be that of a Fermi liquid since the conductivity would depend on length scale. Non-Fermi liquid behavior is unambiguously known for the one-dimensional interacting electron gas. In that case, the conductivity *is* scale-dependent. For repulsive electron–electron interactions, disorder is relevant and drives the system to an insulating state [30–32]. The important contributor to this behavior is the singular behavior at $2k_F$ of the density–density correlation function. This effectively enhances the impurity scattering potential via a vertex correction.

It has been argued [33,34] that *if* one had a non-Fermi liquid in two dimensions, then since the $2k_F$ singular behavior would be absent (for a circular Fermi surface, say), the scaling behavior of the conductivity might be simply determined from that of the single-particle Green's function of the interacting system. Without considering the microscopic origins, the authors of Ref. [34] considered the role of disorder for a class of non-Fermi liquid in which the Green's function is characterized by an exponent α which vanishes for a Fermi liquid. Briefly, they investigated the lowest order of the scaling formulation [35,36] which was adopted [4,5] for the interacting problem. For large

enough α, that is strong enough interaction, they found that the disorder was irrelevant and that if the pure system were metallic, that behavior would survive the presence of disorder. The difference between this approach and previous work [4,5] is that the interactions are not being treated perturbatively; their effects are included in the anomalous behavior which leads to the exponent α.

15. Conclusions and questions

The observation of metallic states in the two-dimensional disordered electron gas raises an number of puzzling questions. Most of these are determined by the fact that in the samples in which the metal–insulator transition is observed, the Coulomb interaction energy is by far the largest energy in the problem. Thus, we have a unambiguous example of the *strong-coupling many-body problem* for which our theoretical machinery is not very powerful.

In those cases where a metal–insulator transition is observed [1,20–24] the metallic phase occurs as the carrier density increases. However, we know that at still higher densities, smaller r_s, when the ratio of Coulomb energy to Fermi energy is much smaller, no metallic states are observed [14,17,18]. How is this re-entrant behavior to be understood? And is it related to a crossover from strong to weak coupling?

As discussed in Section 12, and more completely in Ref. [26], the interactions plus disorder RG analysis [4–7] gives an indication of the possible stability of a metallic state. However, the necessary and appropriate strong-coupling analysis has not been given. The weak-coupling regime may be accessible in the experiments and the detailed magnetic field and temperature dependences of the resistivity need to be analyzed in terms of the existing theory.

As a guide to the theory, it is important to know whether at even lower temperatures than those so far investigated the resistivity continues to fall or approaches a constant. That is, is the metallic state that of a perfect conductor (non-Fermi liquid?) or an ordinary one with a residual resistance?

Is the observed metallic state that of a non-Fermi liquid in which ordinary quasiparticles are not to be expected? If so, how does it arise and what are its properties? This question is not approached in the discussion of Ref. [34], where the non-Fermi liquid state is assumed at the outset.

References

[1] S.V. Kravchenko et al., Phys. Rev. Lett. 77 (1996) 4938 and references therein.
[2] D. Popović, A.B. Fowler, S. Washburn, Phys. Rev. Lett. 79 (1997) 1543.
[3] E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan, Phys. Rev. Lett. 42 (1979) 673.
[4] A.M. Finkel'stein, Z. Phys. B 56 (1984) 189.
[5] C. Castellani et al., Phys. Rev. B 30 (1984) 1596.
[6] For reviews, see D. Belitz, T.R. Kirkpatrick, Rev. Mod. Phys. 66 (1994) 261.
[7] A.M. Finkel'stein, in: I.M. Khalatnikov (Ed.), Sov. Sci. Reviews/Section A-Physics Reviews, vol. 14 (1990) 3.
[8] For a review of the situation in the mid-1980s, see P.A. Lee, T.V. Ramakrishnan, Rev. Mod. Phys. 57 (1985) 287.
[9] N.F. Mott, Proc. Phys. Soc. London, Ser. A 62 (1949) 416.
[10] P.W. Anderson, Phys. Rev. 109 (1958) 1492.
[11] P.W. Anderson, E. Abrahams, T.V. Ramakrishnan, Phys. Rev. Lett. 43 (1979) 718.
[12] L.P. Gor'kov, A.I. Larkin, D.E. Khmel'nitskii, Zh. Eksp. Teor. Fiz. Pis'ma Red. 30 (1979) 248 [JETP Lett. 30 (1979) 248].
[13] A. MacKinnon, B. Kramer, Phys. Rev. Lett. 47 (1981) 1546.
[14] Van Keuls et al., Phys. Rev. B 56 (1997) 13263.
[15] D.J. Bishop, D.C. Tsui, R.C. Dynes, Phys. Rev. Lett. 44 (1980) 1153.
[16] M.J. Uren, R.A. Davies, M. Pepper, J. Phys. C 13 (1980) L985.
[17] G.J. Dolan, D.D. Osheroff, Phys. Rev. Lett. 43 (1979) 721.
[18] See also S.-Y. Hsu, J.M. Valles, Phys. Rev. Lett. 74 (1995) 2331.
[19] For a review, see B.L. Altshuler, A.G. Aronov, in: M. Pollak, A.L. Efros (Eds.), Electron–Electron Interactions in Disordered Systems, North-Holland, Amsterdam, 1984, p. 1.
[20] Y. Hanein et al., Phys. Rev. Lett. 80 (1998) 1288.
[21] M.Y. Simmons et al., Phys. Rev. Lett. 80 (1998) 1292.
[22] J. Lam et al., Phys. Rev. B 56 (1997) R12741.
[23] P.T. Coleridge et al., Phys. Rev. B 56 (1997) R12764.

[24] K. Ismail et al., cond-mat/9707061.

[25] V. Dobrosavljević, E. Abrahams, E. Miranda, S. Chakravarty, Phys. Rev. Lett. 79 (1997) 455.

[26] C. Castellani, C. Di Castro, P.A. Lee, cond-mat/9801006.

[27] For example: N.W. Ashcroft, N.D. Mermin, Solid State Physics, Holt, Rinehart and Winston, Philadelphia 1976, p. 682.

[28] D. Simonian et al., Phys. Rev. Lett. 79 (1997) 2304.

[29] V.M. Pudalov et al., Pis'ma Zh. Eksp. Teor. Fiz. 65 (1997) 887. [JETP Lett. 65 (1997) 932].

[30] L.P. Gor'kov, I.E. Dzyaloshinskii, ZhETF Pis. Red. 18 (1973) 686.

[31] JETP Lett. 18 (1973) 401.

[32] T. Giamarchi, H.J. Schulz, Phys. Rev. B 37 (1988) 325.

[33] M. Ogata, P.W. Anderson, Phys. Rev. Lett. 70 (1993) 3091.

[34] S. Chakravarty, L. Yin, E. Abrahams, cond-mat/9712217.

[35] K. Efetov, A.I. Larkin, D.E. Khmel'nitskii, Zh. Eksp. Teor. Fiz. 79 (1980) 1120.

[36] Sov. Phys. JETP 52 (1980) 568.

Metal–insulator transition in two dimensions

V.M. Pudalov [a,*], G. Brunthaler [b], A. Prinz [b], G. Bauer [b]

[a] *Russian Academy of Sciences, Troisk, Moscow District 142092, Russian Federation*
[b] *Institut für Halbleiter Physik, Johannes Kepler Universität, Linz, A-4040, Austria*

Abstract

Studies of the temperature dependence of the conductivity on large number of Si–MOS samples with different disorder give evidence for the lack of a universal one-parameter scaling. We demonstrate that the 2D metallic state survives over a wide range of temperatures from 16 mK to 8 K, carrier densities from $0.7–35 \times 10^{11} \, cm^{-2}$ and conductances, G, from $G \simeq 0.3$–$120 \, e^2/h$. We have revealed experimentally a weak logarithmic correction to the conductivity, which is of the delocalizing sign. We discuss a novel relationship between the partial scaling functions and the conductivity, which explains both, the exponential temperature dependence and the weak logarithmic corrections to the conductivity of the 2D strongly interacting system. © 1998 Published by Elsevier Science B.V. All rights reserved.

PACS: 71.30.+h; 73.40.Qv; 73.20.Fz

Keywords: Metal–insulator transition; Two-dimensional electron gas

The one parameter scaling theory (OPST) of localization [1] is considered as one of the milestones of contemporary condensed matter physics and gives a clear physical picture of localization. The theory assumes the existence of a universal scaling function for the conductance G:

$$\beta = \frac{\partial \ln G}{\partial \ln L}, \tag{1}$$

which is a function of the system size, L, only. Particularly, according to the conventional interpretation of the OPST, there is no true metallic state and no metal–insulator transition in two dimensions (2D). All the states of the electrons in 2D are expected to be localized, strongly or, at least, weakly. However,

* Corresponding author. Fax: +7 095 334 0012; e-mail: pudalov@sci.lebedev.ru.

numerous attempts to prove rigorously the one parameter scaling meet so far with major difficulties [2]. On the experimental side, a large body of data on quasi-2D systems support mainly the concept of "quantum corrections" to the classical diffusion, rather than the OPST.

Convincing evidence for the existence of a 2D metallic state Si–MOS structures at zero magnetic field has been obtained in studies of the quantum Hall effect to insulator transitions [3,4]. The extended states, which in high magnetic field H are centered in the corresponding Landau bands, were found to coalesce and remain in a finite energy range as H approaches 0 [3]. This behavior could not be expected in the framework of the OPST, where the extended states are anticipated to "float up" in energy as $H \to 0$ [5]. The experimental findings thus suggest the

existence of a metal–insulator (MI) transition, whereas the predicted floating would evidently correspond to complete localization.

In studies at zero magnetic field [6,7], the conductivity in high-mobility Si–MOS structures was found to scale with temperature and with electric field [8], and the scaling parameter demonstrated a critical behavior appropriate for a MI transition. Moreover, in high-mobility Si–MOS structures, the temperature dependence of the resistivity shows exponentially strong drop at temperatures below about 2 K, which indicates a metallization rather than localization [6]. A similar behavior was found up to date in different material systems: 2DEG system in Si–MOS structures [6,7,9], 2DEG and 2DHG system in Si/SiGe heterostructures [10–12], and 2DHG system in GaAs/AlGaAs structures [13,14]. It is worth to note earlier studies of the conductivity in low mobility Si–MOS structures [15,16], where a weak positive temperature derivative, $\mathrm{d}\rho/\mathrm{d}T > 0$, was reported.

The experimental finding of the metallic state and of the MI transition in two dimensions has fueled a debate over the nature of the 2D metal [17–23].

1. Brief overview of the experimental data

The phenomena manifest a number of common generic features, pronounced stronger or weaker in different materials:

(i) The resistivity drops exponentially fast as temperature decreases below $T_0 \sim 2\,\mathrm{K}$ [6,7,21]:

$$\rho(T) = \rho_0 + \rho_1 \exp(-(T_0/T)^p), \qquad (2)$$

where $p \sim 1$. The resistivity drop is the most pronounced in high-mobility structures [7,21,13]. The corresponding temperature, T_0 is sample dependent and raises with density [21,13]: $T_0 \propto |\delta n|^q$, where $\delta n = (n - n_c)$ and $q \approx 1$.

(ii) The resistivity for each particular sample (at not too low temperatures) may be scaled into the two branches, using a single scaling parameter T/T_0. The scaling parameter, T_0, thus demonstrates a critical behavior around a critical density, n_c, [6,7,9,10,14,21], whereas the exponent q obtains a meaning of the product, $q = z\nu$, of the dynamical exponent, z, and of the correlation length exponent ν [8].

(iii) A regular negative magnetoresistance was found in a weak perpendicular field [21,9], which indicates the quantum interference contribution in the conductivity of the 2D metallic state.

(iv) The in-plane magnetic field $H_\parallel$ causes the resistivity to rise exponentially strong [24–26,9] $\rho(H_\parallel) \propto \rho_1 \exp(-((T_0 - bH)/T)^p)$, where $b \sim (g^* \mu_B^*)/k_B$ [28]. Thus, the parallel field, roughly, restores the "high-temperature" resistivity value, $\rho(T > T_0)$.

(v) The in-plane magnetic field component applied in addition to the perpendicular field causes the extended states to float up [29] and restores their conventional behavior [5], appropriate for the noninteracting 2DEG system in GaAs/AlGaAs [30].

It was suggested recently [31] that the OPST may be modified phenomenologically to incorporate the interacting system case where the scaling β-function in 2D changes the sign at a critical value of the conductivity, G_c. On the other hand, the theory developed earlier by Finkel'stein [32], predicted failure of the one-parameter scaling picture for the strongly Coulomb-interacting system and predicted the occurrence of the metallic state in the latter case. In the latter theory, spin, and valleys beyond the Coulomb interaction, favor the formation of the metallic state. It was unclear so far, which of the interaction mechanisms drives the 2D system into a metallic state, as well as it was unclear if the conductivity of the 2D metallic system obeys one- or two-parameter scaling.

In the Si–MOS systems under study, the interactions are, no doubt, extremely strong. At the critical density, the ratio of the Coulomb energy to Fermi energy, E_{ee}/E_F is of the order of 10 [33]. Besides, due to the broken reflection symmetry in the triangular potential well, the energy spectrum includes a linear term $\pm \alpha k_F$. The latter one is large in Si–MOS, so that the energy gap at the Fermi energy between the two branches of the spectrum with different chirality, $\alpha(k_F^+ - k_F^-) \sim 2\,\mathrm{K}$, is of the order of the Fermi energy, $E_F \gtrsim 6\,\mathrm{K}$, [21,34]. The electron system is thus spin-polarized in the momentum space so that the population ratio of the spin-subbands with different ($\pm$) chirality, $(k_F^-/k_F^+)^2$, reaches 2 in the vicinity of n_c. The spin polarization of the system affects the exchange energy, in addition to that it favors the forward scattering in the quantum corrections to the conductivity [22,23]). The

zero-field spin-gap was experimentally observed also in p-Si–MOS systems [35], and in p-GaAs/AlGaAs heterojunctions with asymmetric potential well [36] (with a similar magnitude in the former system, and 10 times smaller magnitude in the latter one). In p-SiGe samples, the corresponding energy levels originate from the bulk $|M_j| = \frac{3}{2}$ states, and thus the spin effects are essential and may contribute in a regular way, via the g-factor.

In the current paper we focus on general questions of the dimensionality of the parameter space describing the system, on the microscopic origin of the relevant interaction and on the range of existence of the 2D metallic state. We address experimentally the following questions: Does the conductivity of the 2D metal obey indeed the one-parameter scaling? What is the asymptotic behavior of the conductivity as $T \to 0$? To what state the system is driven as the temperature decreases, or, correspondingly, as the size of the system increases?

We consider here the high conductance range, $G \gg 1$ (the conductance, throughout the paper, is in units of e^2/h). We have revealed the weak logarithmic correction to the conductivity, and found that its sign remains negative over the range of G from $G = G_c \sim 0.3$–2 up to $G = 120$. Our data confirm that the 2D metallic state, once formed by increasing carrier density n above n_c, remains further stable at $n > n_c$ upon decreasing the temperature (increasing the system size) down to the lowest temperature we studied, 16 mK. We present experimental data demonstrating lack of the one-parameter scaling approach and suggest a novel scaling behavior, which fits the experimental data.

In order to test the role of disorder and of interaction, we have performed measurements on a large number of Si–MOS samples with different disorder (the latter may be characterized by the peak mobility value, μ, which varied by a factor of 8.5[1]). Seven samples have been studied most intensively,

---

[1] We characterise the samples with peak mobility values taken at $T = 20$–300 mK, because the high mobility samples exhibit an exponentially strong drop in resistivity at $T < 2$ K. This is in contrast to the low mobility samples in which ρ does not change substantially at $T < {\sim}4$ K, and which are usually characterised by mobility measured at 4 K. Typical temperature dependence of mobility for these samples in the range 0.1–10 K is presented in Ref. [6].

Si-15 ($\mu = 51\,000\,\text{cm}^2/\text{Vs}$ at $T = 0.3$ K), Si-5 ($\mu = 43\,000$), Si-2 ($\mu = 39\,000$), Si-11 ($\mu = 36\,300$), Si-22 ($\mu = 29\,000$), Si-43 ($\mu = 19\,600$), and Si-39 ($\mu = 4800$). All the samples exhibited the characteristic temperature dependence with a sample-dependent critical density, n_c, which separates the metallic ($\text{d}\rho/\text{d}T > 0$) and the insulating ranges ($\text{d}\rho/\text{d}T < 0$). Fig. 1a and c shows the corresponding set of the temperature dependencies of the resistivity, $\rho(T)$ taken at different electron densities n^j for two extreme examples: the least disordered sample Si-15 and the most disordered one, Si-39.

The dashed line in Fig. 1 separates, conditionally, the metallic and insulating regions. As the sample mobility decreases, both, the characteristic magnitude and steepness of the drop in $\rho(T)$ decrease, whereas the drop shifts to higher densities. For example, in Si-15 the drop occurs in exponential fashion (see Eq. (2)) [21] and amounts to $\Delta\rho(T \gg T_0)/\rho(T \ll T_0)$ ~ 7, whereas for Si-39 the drop is almost linear and is within a few percents. The weak and almost linear temperature dependence of the conductivity in the latter sample is reminiscent of the data reported in 1980s [16,15]. This was attributed to the temperature variations of the screening length [15,37,38], or of the electron–electron scattering [40]. In both models, the temperature dependence was described by a series expansion to powers of T/E_F. We reproduced such a linear $\rho(T)$ dependence in our worst sample Si-39. At the same time, the new data taken on a variety of samples [34] reveal a continuous transformation of the linear raise to the steep exponential $\rho(T)$-dependence, as sample mobility increases (compare Fig. 1a and c). Whereas the linear $\rho(T)$-dependence can be described in different frameworks [38,39], the steep exponential dependence cannot, evidently, be fitted by the series expansion to T/E_F.

2. β-function: one-parameter vs. two-parameter scaling

The occurrence of the MI transition unavoidably means that the scaling function, $\beta = \text{d}\ln G/\text{d}\ln L$, if it exists, should change the sign at a certain critical value of conductance G_c. The anticipated behavior of the β-function, which, at first sight, corresponds to the existence of the 2D metallic state, is shown
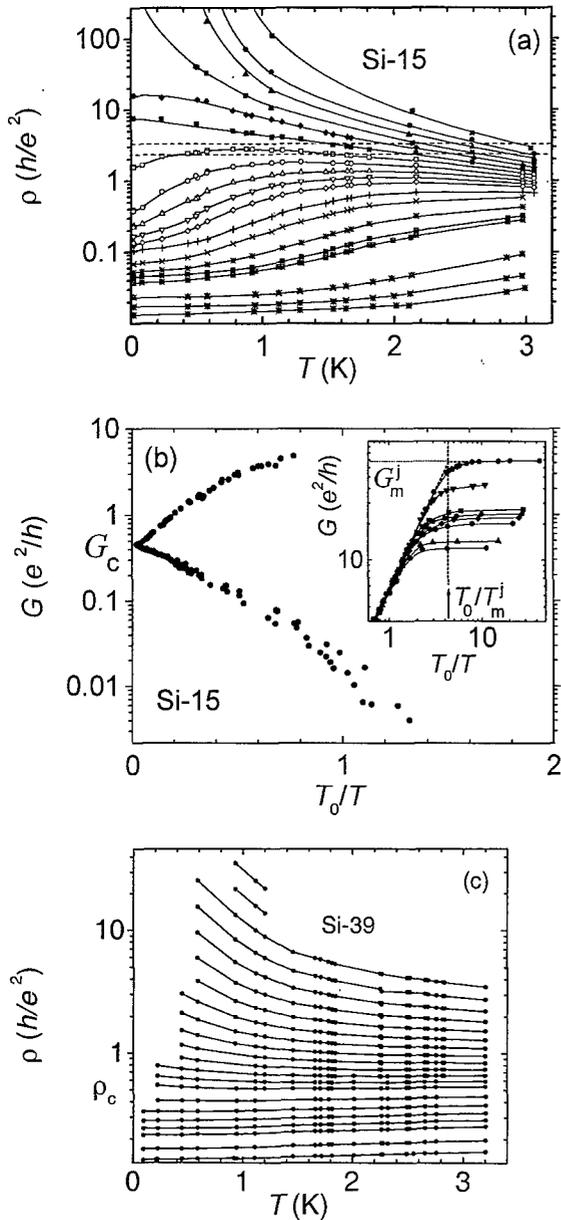
Fig. 1. (a) and (c) Resistivity vs. temperature measured on the samples Si-15 and Si-39. Dashed lines indicate the critical boundary for the temperature ranges $T < 0.1E_F/k$ and $0.15E_F/k$, correspondingly. (b) Conductivity vs. temperature for Si-15 scaled with a single parameter (T_0/T) in the range $T \leqslant 0.1T_F$. Different symbols in Fig. 1a correspond to the carrier densities (in units of 10^{10} cm^{-2}): from 4.49 to 9.89 (in steps of 0.54), 10.97, 12.05, 14.21, 16.4, 17.45, 28.2, 39.0, 49.85; whereas in Fig. 1c correspond to n from 19.5 to 33.88 (in steps of 1.025), 35.93, 37.98, 40.03, 42.08, 44.13, 49.25, 54.4. The inset shows the expanded low-temperature part of the scaling plot in the region of the saturation.
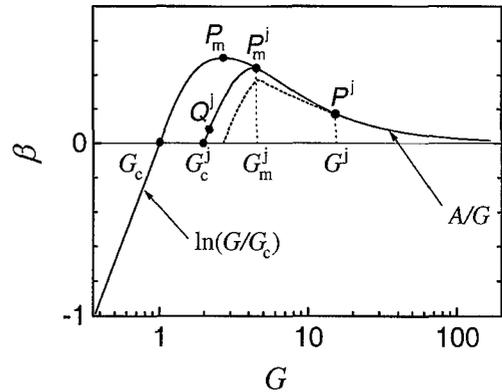
Fig. 2. Schematic view of the scaling β-function for a two-dimensional interacting system.

schematically in Fig. 2 by the bold line. The negative part, $\beta(G) < 0$, at $G < G_c$ corresponds to the strongly localized state, studied in Ref. [33].

It was pointed out by Dobrosavljević et al. [31] that nearly exponential drop of the resistivity vs. temperature in the metallic phase, in Eq. (2), reflects a linear rise of the scaling function on the logarithmic conductance scale, $\beta(G) \propto \ln(G/G_c)$. However, it remained unclear, which microscopic mechanism drives the β-function linearly through zero, surprisingly far away from the critical point: for example, the linear rise extends from $G_c \approx 0.4$ to $G \approx 10$ in the least disordered samples Si-15.

There were a few attempts to plot the scaling function for 2D metallic state, based on the data scaled with the temperature [9,40]. This can be easily plotted in the range of the low conductance $G \gtrsim G_c \sim 1$, i.e. in the range of the exponential drop of $\rho(T)$. However, these attempts failed at higher conductance, $G \gg 1$, because (i) the relation of the length scale to the temperature was unknown, (ii) the anticipated weak temperature dependence of the conductivity which should correspond to the smooth tail of $\beta(G)$ has not been revealed until now, and (iii) it was even unclear which part of $\rho(T)$, should correspond to $\beta(G)$ at high G. As an illustration, from the data presented in Ref. [40] it seemed that $\beta(G)$ may become negative at high G again, introducing an interesting attractive critical point.

We adopt as an assumption, that the metallic phase in the range of $\beta > 0$ originates from interactions.

The experiments in parallel magnetic fields [25,24,9] clearly demonstrated the crucial role of the electron spin. We presume, for the sake of definiteness, the relevant interaction is related to spin. As for the origin of the interaction mechanism, there were a number of suggestions: spin–orbit coupling due to the breaking of the symmetry and changing the universality class [21–23], *p*-wave superconductivity [17], spin–triplet pairing [19] etc. For our analysis it will be important only that the interaction introduces the corresponding energy Δ_s [41], and the characteristic diffusive parameters: spin-breaking time τ_s, and spin-breaking length, $l_s = (D\tau_s)^{1/2}$.

2.1. On universality of the scaling picture

The first issue we wish to explore is as follows: *is the critical G_c value a universal constant?* A priori there might be a number of different possibilities:

1. complete universality of G_c, independent on material, on the strength of interaction, and on disorder (this would clearly demonstrate a success of the OPST).
2. G_c may represent an asymptotic ideal value, whereas samples with different disorder may obey a scaling with their particular scaling functions which only approach to an ideal $\beta(G)$ as disorder decreases.
3. G_c may be nonuniversal at all, being disorder and interaction dependent. The latter scenario would mean the lack of one-parameter scaling in the interacting system, as predicted by Finkel'stein [32].

In order to answer this question, we have performed the scaling analysis of the conductivity on a number of Si–MOS samples. This procedure is straight forward for moderate mobility samples for which the boundary between the insulating ($d\rho/dT < 0$) from metallic ($d\rho/dT > 0$) regions on the $\rho - T$ plane was temperature independent, similar to that reported in Refs. [6,7]. However, for the highest (like Si-15) and lowest (Si-39) mobility samples, the boundary appeared to be strongly temperature dependent. In this cases we performed the scaling analysis in the range of temperatures limited by $T \leqslant 0.1E_F/k = 0.75$ K and by $0.15E_F/k = 1.13$ K. Since in the scaling procedure the high-temperature data obtain more weight, the resulting G_c values correspond mainly to the data in the
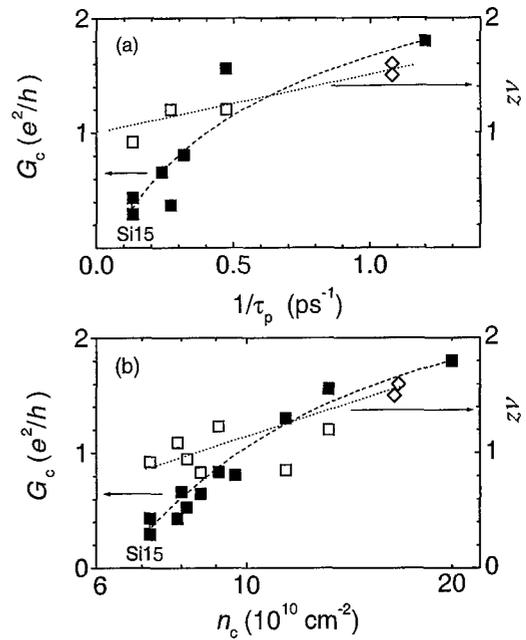

Fig. 3. Critical conductivity G_c (filled symbols, left Y-axis) and the critical index zv (opened symbols, right Y-axis): (a) vs. elastic scattering rate, for six samples. (b) vs. critical electron density n_c for 10 samples. The squares correspond to Si–MOS samples under current study, diamonds – to two Si–MOS samples studied in Ref. [9].

vicinity of the highest temperature. The corresponding scaling plot for Si-15 is shown in Fig. 1b.

As a result of scaling, we obtained the critical conductance, G_c, the critical density, n_c, and the scaling temperature exponent zv for each sample. Fig. 3a summarizes G_c-values for different samples plotted as a function of the elastic collision rate at the peak mobility, which we have chosen as a measure of disorder. G_c is strongly sample dependent, increasing from ~ 0.3 for the best sample Si-15, to 1.75–2.4 for the most disordered one. In the samples with temperature-dependent boundary, the scaling results depend on the considered temperature range. The double data point ($G_c = 0.25$–0.33) in Fig. 3a and b indicates the uncertainty in the scaling procedure, and was obtained for the range of temperatures, $T \leqslant 0.1E_F/k$ and $T \leqslant 0.15E_F/k$. We note that the temperature dependence of the critical boundary indicates already lack of the universal scaling.

In order to exclude a certain arbitrariness in the choice of the disorder parameter, we plotted in Fig. 3b the same data on the critical conductance, but as a function of the logarithm of the critical density, n_c. The latter is roughly proportional to the number of scatterers [33,38], and thus may serve as a different measure of the disorder. Both dependencies, Fig. 3a and b does not exhibit a saturation at a universal conductivity value as disorder decreases, and instead, clearly shows the absence of such asymptotic value. Thus, we arrive at the latest possibility, namely, lack of universality, at least in the range of G_c from 0.4 to 2. In addition to the data on our *n*-Si–MOS, Popović et al. [9] reported $G_c = 0.5$–0.65 for two other *n*-Si–MOS samples with mobility increased by backside biasing. Two results, obtained on 2D hole system in GaAs/AlGaAs are worth noting: $G_c = 0.65$, $n_c = 1.2 \times 10^{10}$ cm^{-2} reported in Ref. [13], and $G_c = 2.1$, $n_c = 5.1 \times 10^{10}$ cm^{-2} [14]. One more data point, $G_c = 100$, $n_c = 1.8 \times 10^{11}$ cm^{-2}, was measured on *n*−Si/SiGe [12]. The data taken on different materials do not lie on the dashed dependence for our Si–MOS samples and are outside of the plotted region in Fig. 3, nevertheless, they support our conclusion on the lack of universality of the G_c-value.

It is reasonable to assume that the onset of the metallic state occurs when the disordering potential is overwhelmed by an interaction energy. Fig. 3 demonstrates that *higher disorder corresponds to higher n_c and G_c values*. The correlation between the critical density, G_c and disorder holds also for the *p*-GaAs/AlGaAs system, as one can figure out from the above cited data. Since the ratio of the Coulomb interaction energy E_{ee} to kinetic energy E_F decreases with density $\propto n^{-1/2}$, we conclude further that there must exist another interaction, beside purely Coulomb, whose *strength increases with density*. The latter one, we presumed above, is related to spin.

Fig. 3b shows also the exponent, zv, obtained from the scaling analysis on different samples. These data are less sample dependent than G_c, however, a trend is seen for zv to increase from $zv \sim 0.8$–1 appropriate for the least disordered samples, to $zv \approx 1.6$ for moderate mobility samples. This trend is in a good agreement with data taken on a 2D hole system in GaAs/AlGaAs. Indeed, in Ref. [13] the exponent $zv = 1$ has been reported for the M–I transition at

a low density, $n_c = 1.2 \times 10^{10}$ cm^{-2}, in the sample exhibiting an exponentially strong drop, whereas in Ref. [14] the exponent $zv \approx 4$ was found for the transition at 4.3 times higher density, $n_c = 5.1 \times 10^{10}$ cm^{-2} and with almost linear temperature dependence of the resistivity. For the former sample $\rho(T)$-data [13] look analogous to those shown in Fig. 1a for our least-disordered sample Si-15, and the latter data [14] are similar to those shown in Fig. 1c for Si-39. The critical values reported for *p*-SiGe in Ref. [10], $zv = 1.6$–2 and $G_c = 2$, also do not contradict the trend discussed above.

It is instructive to compare further the result of Ref. [14] with our data on Si–MOS. For our least-disordered sample Si-15, the transition occurs at a density $n_c = 7.5 \times 10^{10}$ cm^{-2}, which is only 1.5 times higher than that reported in Ref. [14]. Thus, MI transition in this Si–MOS sample takes place at the same Coulomb energy $E_{ee} = (\pi n_s)^{1/2} e^2 / \kappa \approx 70$ K, at the similar Fermi energy ($E_F = 5.5$ K in Si-15 and 4.6 K, in Ref. [14]), but at a mobility $\sim$25 times lower than in Ref. [14]. Nevertheless, the transition in *p*-GaAs/AlGaAs is much weaker and its features are more similar to those shown in Fig. 1c for our most disordered sample Si-39, where it occurs at seven times higher density. This again confirms the involvement of another interaction different from purely Coulomb one.

2.2. Two-parameter driven behavior and the family of partial one-parameter scaling trajectories

Despite the absence of a universal critical conductance (for different disorder and for the samples with temperature dependent critical boundary), every particular sample did display a scaling behavior. Therefore, one can restore a partial scaling β-function for a particular sample. The steep linear part of the β-function, $\beta \propto \ln(G/G_c)$, was plotted in Refs. [40,9] and is a direct consequence of the exponential temperature dependence of the resistivity in the metallic phase, Eq. (2). However, the attempts to restore the β-function further, at higher G, failed so far [40]. We show below that the major problem in such an analysis is due to the principle two-parameter character of the β-function and due to the two-parameter driven trajectories. We assume in the following that a scaling function (in the above narrow

meaning, for a particular sample and a particular density) does exist and will clarify how it should be related to the experimentally measured features in $\rho(T)$.

2.2.1. Large L, high conductance, $G \gg 1$

The scaling β-function is expected to decay slowly and the quantum corrections, therefore, are small in this range. For the sake of simplicity (and in accordance with experiments [34]), we assume $\beta \propto A/G$ [22,42], as plotted schematically in Fig. 2. According to our assumption, the scaling behavior of the conductivity should be governed by the corresponding spin-breaking length, and hence the spin breaking time τ_s must be smaller than τ_i, the inelastic length. Thus, the following inequality must hold on over this range:

$$\tau_i > \tau_\phi \approx \tau_s \gg \tau_p, \tag{3}$$

here τ_ϕ is the phase breaking time, and τ_p is the elastic time.

We consider a sample of a large macroscopic size $L \gg l$, the elastic length, having a conductance $G^j \gg 1$. The conductance G^j corresponds to an initial point P^j at the scaling β-function, as shown in Fig. 2. As the temperature raises, the effective sample size L decreases $\propto T^{-s/2}$ and τ_i decreases $\propto T^{-s}$, where s varies between 1 and 2 [31,39,43]. On the other hand, τ_s does not depend much on temperature for the known spin-relaxation mechanisms, being mainly determined by the average electric field seen by an electron during its orbital diffusion. Since $\beta = \mathrm{d} \ln G / \mathrm{d} \ln L > 0$, the point representing the conductance of our system will proceed in the left direction along the $\beta(G)$-curve, with the conductance driven by the variation of the sample cutoff length L/L_s [43].

This will take place as long as inequality (3) holds. At a length L_m^j (or, equivalently, at a temperature T_m^j) the inequality (3) breaks. At this point, designated P_m^j in Fig. 2, τ_i is equal to τ_s and $L_i = L_s = L_m^j$.

2.2.2. Small L, small conductance, $G \sim G_c$

Passing the point P_m^j, the system must depart from the gradual trajectory and start following another trajectory being driven further by the orbital inelastic time τ_i which becomes now less then τ_s. Thus, in the

region of conductivities, $G_m^j > G$, another inequality holds

$$\tau_s > \tau_i \approx \tau_\phi > \tau_p. \tag{4}$$

As L decreases further, τ_i decreases, and finally, when τ_i becomes equal to τ_p, the system arrives at a point Q^j corresponding to the Drude conductivity of the sample of the size $L = l$ at a carrier density n^j. This point should be located close to the line of $\beta = 0$.

Both competing parameters τ_i and τ_s are two different functions of the electron density, of the disorder and of the temperature. Therefore, the turning point P_m^j at which $L_i = L_s$, is, in general, density and disorder dependent. Such a situation is shown schematically in Fig. 2, where a partial trajectory (the dashed curve), departs from the main trajectory at a different point P_m^j for a different density. The above consideration explains the nonuniversality of G_c shown in Figs. 1 and 3. The picture given by Fig. 2 corresponds to such a set of $\rho(T)$ curves where the critical boundary $\rho_c(T)$, separating the insulating and metallic ranges, have a negative slope $\mathrm{d}\rho_c/\mathrm{d}T < 0$. In this case, the partial trajectories arrive at G_c^j which is always greater than G_c. The last turning point P_m and the last critical point G_c obtain thus the meaning of limiting parameters (for the particular sample) characteristic for the lowest density at which the system still is in the metallic state. Among the samples studied, we noticed that less-disordered samples have the stronger tilted critical boundary, i.e. with larger $\mathrm{d}\rho_c/\mathrm{d}T$, whereas samples with lower mobility show a less tilted or horizontal boundary with $\mathrm{d}\rho_c/\mathrm{d}T \approx 0$. At what side of the limiting point, G_c, the partial trajectories will arrive, depends on the ratio of $\mathrm{d}\tau_s/\mathrm{d}n$ and $\mathrm{d}\tau_i/\mathrm{d}n$. One cannot, therefore, exclude the opposite situation, where the partial trajectories and G_c^j will occur on the left side of G_c. Nevertheless, among the samples studied, as well as from the data available in the literature, we did not find an example of such behavior.

The steep part of β-function is nonuniversal and density dependent, therefore the gradual part cannot be universal as well. The gradual part of β, right to the turning point P_m, corresponds to a spin-governed delocalized state. This state is realized in the range of $L > L_s$, or equivalently, at $T < T_s$, where $T_s \sim (D\tau_s)^{-1/2}$. The steep part of β, left from the point P_m, corresponds, presumably, to a Coulomb and/or

disorder-dominated delocalized state. It is thus the ratio, τ_s/τ_p, that determines how far the steep exponential part of β will extend from the critical point G_c^j. In the least-disordered sample Si-15 shown in Fig. 1a, $\tau_s/\tau_p \gtrsim 10$ [43], the steep part of β extends to $G/G_c^j = 7$ and the peak in β amounts to ~ 1. In the most-disordered sample, Si-39, the ratio τ_s/τ_p is ~ 1 and the maximum of β amounts to $\sim 10^{-2}$.

At the turning point, P_m^j in Fig. 2, $L_s = L_i$; this point corresponds to the temperature T_m, at which the steep exponential dependence $\rho(T)$ almost saturates at about ρ_0. The inset in Fig. 1b shows the expanded low-temperature region of the $G(T_0/T)$ data (scaled to the single curve over the range of $T > T_m$). The saturation point for one of the curves is indicated by an arrow. As seen, the conductance at the saturation point, G_m, is density dependent. Thus, the position of the turning point on the β-function is density dependent too; this confirms our conclusion on the nonuniversality of β.

2.3. Relationship between the partial β-function and the temperature dependence of the conductivity

In order to calculate the conductivity for the sample of a macroscopic size L from the known partial $\beta(G)$, one has to integrate the differential equation given by the β-function, starting from the point Q^j to the final point P^j. Below, we reproduce the integration, similar to that done in Refs. [1,31], modifying it to the nonuniversal case. The integral may evidently be separated into two parts, the one from the point Q^j to P_m^j and another one from P_m^j to P^j. For simplicity, we choose the β-function in the form

$$\beta^j = \begin{cases} \dfrac{1}{v} \ln\left(\dfrac{G}{G_c^j}\right) & \text{if } L > L_m \\ \dfrac{A}{G} & \text{if } L \leqslant L_m, \end{cases} \tag{5}$$

where, from the continuity at P_m^j, $A = (G_m^j/v)\ln(G_m^j/G_c^j)$ follows.

2.3.1. High temperatures, $T \geqslant T_0$, small $L < L_s$

On the first interval, from Q^j to P_m^j, we follow Ref. [31] and obtain on integrating the exponential

temperature dependence of the conductivity

$$G(T) = G_c^j \exp\left(\ln \frac{G(Q^j)}{G_c^j} \left(\frac{L}{l}\right)^{1/v}\right)$$

$$\approx G_c^j \exp\left(\frac{T_0}{T}\right)^{1/zv} \tag{6}$$

with $T_0 \propto |n - n_c|^{zv}$ [31].

2.3.2. Low temperatures, $T \ll T_0$, large $L \gg L_s$

On integrating over the interval from L_s to L, (or from the point P_m^j to P^j) we obtain, as usual, a weak temperature dependence:

$$G(T) = A \ln(L/L_s) + G_m^j \propto -\frac{A}{z} \ln T, \tag{7}$$

with $A > 0$. In the intermediate range, $T \sim T_0$, the temperature behavior cannot be described by the simple model Eq. (5). On the other hand, the "quasiclassical dependence" Eq. (2) provides a good interpolation for the "quantum dependences", Eqs. (6) and (7), over the entire temperature range. This explains why the former empirical dependence fits the experimental data so well [13,21].

The empirical "zero temperature scattering rate", ρ_0, thus originates from the conductance at the turning point, G_m^j (see Eq. (7)). The success of the previous one-parameter scaling of the conductivity at not too low temperatures [6,7,9,13,10] is an attribute of the samples with a horizontal boundary, where G_c^j is density and is temperature independent. It also indicates that the slope of the steep part of the partial β-trajectories, $1/v$, is almost density independent (see Eq. (6) and the dashed curve in Fig. 2).

3. Asymptotic behavior of the conductivity as $T \to 0$

We address, next, the following question: if the β-function decreases with G, will it, or will it not change sign again at sufficiently high conductance, $G \gg 1$? We note that it is required only that $|\beta|$ approaches 0 asymptotically as G increases, whereas it may have any sign. Such a crossing point, G_{c2}, where $\beta = 0$ would be an attractive point (in contrast to G_c which is repulsive). This point would represent a metal with

a finite conductance as T approaches zero. The negative range of the β-function would correspond to an unusual state with $d\rho/dT < 0$ at finite temperature but vanishing to zero at $T = 0$.

In order to answer this question, we studied the conductance in the $T \to 0$ limit, at different electron densities. It was shown above, that the weak quantum corrections to the conductivity must be revealed only at sufficiently low temperatures $T < T_m$. In other words, such corrections should be most definitely pronounced in the range of the saturation (plateau) of the exponential $\rho(T)$ dependence. The corresponding analysis would be simplified if the exponential dependence is frozen out (if the plateau is wide enough in the $\ln T-$ scale), this can be achieved experimentally in the limit of high T_0, T_m (i.e. high carrier density) and in the limit of low temperatures. The former limit extends the plateau to the high temperature side (since $T_0 \propto \delta n$), whereas the latter one extends the plateau to the low temperature side.

We have analyzed the temperature dependence of the conductivity $G(T)$ in both limits, at $T \to 0$, and at high conductance, $G \gg 1$. Fig. 4 represents the results taken in the two extreme cases, down to the lowest available temperature of 16 mK, and at a high density, 9×10^{11} cm^{-2} where $G \approx 70$. Continuous lines in Fig. 4 show the exponential function, $\rho_0 + \rho_1 \exp(-T_0/T)$ fitting the experimental data with three parameters, ρ_0, ρ_1, and T_0 [34]. This function saturates as $T \to 0$, and on its background one can clearly see a weak temperature dependence which we fitted with $C \ln T$ curves. In both cases, the logarithmic dependence is of delocalizing sign. At relatively low G (Fig. 5a), the prefactor, $C = -0.4 \pm 0.1$ is of the order of the "conventional" value u_v/π [43], where $u_v \sim (2 - 1)$ takes into account two valley contribution to conductivity. At $G = 70$ (Fig. 4b), the measured prefactor $C = -3.6 \times 10^{-2}$ is 10 times smaller than the conventional value. The prefactor remains negative and decreases $\propto n^{-1/2}$ till the highest conductance, $G = 120$, we studied. The strong decrease of the prefactor with density is not consistent with the quantum corrections magnitude calculated in Ref. [22,23].

In conclusion, we presented experimental evidence for the lack of one-parameter scaling of the conductivity of the strongly interacting 2D metallic state in Si–
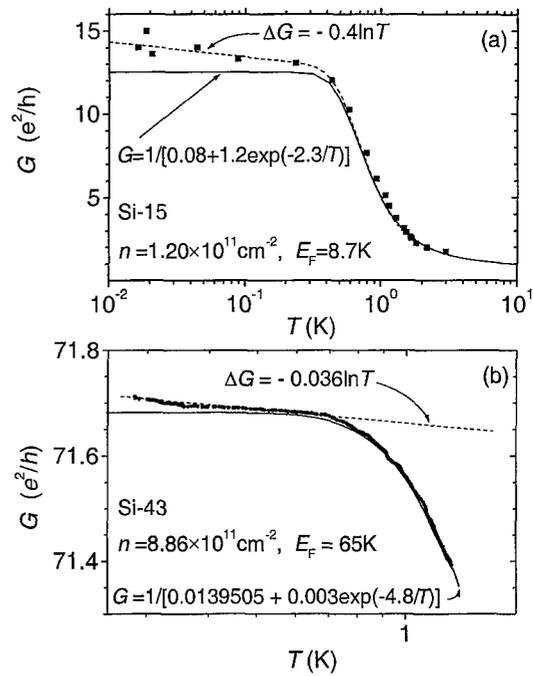

Fig. 4. Temperature dependence of the conductivity in the low-temperature limit, (a) for Si-15 at low electron density, and (b) for Si-43 at high electron density. Thick continuous lines show the best fit by the exponential dependence Eq. (2), dashed lines show the logarithmic temperature dependence.

MOS. We have demonstrated the absence of the second critical point over the studied conductivity range up to $G = 120$, and down to the lowest temperature of 16 mK. We have revealed experimentally the weak logarithmic correction to the conductivity, which is of delocalizing sign. We suggest a novel relationship between the partial scaling functions and the conductivity, which sets the corresponding temperature intervals for the exponential temperature dependence and the weak logarithmic corrections to the conductivity.

Acknowledgements

References

[1] E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan, Phys. Rev. Lett. 42 (1979) 673.

[2] For a review see D. Belitz, T. Kirkpatrick, Rev. Mod. Phys. 66 (1994) 261.

[3] M. D'Iorio, V.M. Pudalov, S.G. Semenchinsky, Phys. Rev. B 46 (1992) 15 992. V.M. Pudalov, M. D'Iorio, J.W. Campbell, Surf. Sci. 305 (1994) 107.

[4] S.V. Kravchenko, W. Mason, J.E. Furneaux, V.M. Pudalov, Phys. Rev. Lett. 75 (1995) 910.

[5] D.E. Khmel'nitskii, Phys. Lett. 106 (1984) 182; JETP Lett. 38 (1983) 556.

[6] S.V. Kravchenko, G.V. Kravchenko, J.E. Furneaux, V.M. Pudalov, M. D'Iorio, Phys. Rev. B 50 (1994) 8039.

[7] S.V. Kravchenko, W.E. Mason, G.E. Bowker, J.E. Furneaux, V.M. Pudalov, M. D'Iorio, Phys. Rev. B 51 (1995) 7038.

[8] S.V. Kravchenko, D. Simonian, M.P. Sarachik, W. Mason, J.E. Furneaux, Phys. Rev. Lett. 77 (1996) 4938.

[9] Dragana Popović, A.B. Fowler, S. Washburn, Phys. Rev. Lett. 79 (1997) 1543. K.P. Li, Dragana Popović, S. Washburn, Proc., EP2DS-12, Tokyo, 1997.

[10] P.T. Coleridge, R.L. Williams, Y. Feng, P. Zawadzki, Cond-mat/9708118.

[11] M. D'Iorio, D. Brown, H. Lafontain, Cond-mat/9708201.

[12] K. Ismail, J.O. Chu, D. Popović, Cond-mat/9707061.

[13] Y. Hanein, U. Meirav, D. Shahar, C.C. Li, D.C. Tsui, H. Shtrikman, Phys. Rev. Lett. 80 (1998) 1288.

[14] M.Y. Simmons, A.R. Hamilton, M. Pepper, E.H. Linfield, P.D. Rose, D.A. Ritchie, Phys. Rev. Lett. 80 (1998) 1292.

[15] E.A. Vyrodov, V.T. Dolgopolov, S.I. Dorozhkin, N.B. Zhitenev, ZhETF 94 (1988) 234. [Sov. Phys. JETP 67 (1988) 998].

[16] K.M. Cham, R.G. Wheeler, Phys. Rev. Lett. 44 (1980) 1472.

[17] P. Phillips, Y. Wan, Cond-mat/97042000.

[18] P. Phillips, Y. Wan, I. Martin, S. Knysh, D. Dalidovich, Cond-mat/9709168.

[19] D. Belitz, T.R. Kirkpatrick, Cond-mat/9705023.

[20] F.C. Zhang, T.M. Rice, preprint.

[21] V.M. Pudalov, Cond-mat/9707053. Cond-mat/9707076. JETP Lett. 66 (1997) 170. [Pis'ma ZhETF 66 (1997) 168].

[22] M.A. Skvortsov, Cond-mat/9 712 135. JETP Lett. 66 (1998). [Pis'ma ZhETF 66 (1998) 118].

[23] Yu. Lyanda-Geller, Cond-mat/9 801 095 Phys. Rev. Lett. 80 (1998) 4273.

[24] V.M. Pudalov, G. Brunthaler, A. Prinz, G. Bauer, JETP Lett. 65 (1997). [Pis'ma ZhETF 65 (1997) 887. Cond-mat/9 707 054.]

[25] D. Simonian, S.V. Kravchenko, M.P. Sarachik, V.M. Pudalov, Phys. Rev. Lett. 79 (1997) 2304.

[26] In Ref. [27], the magnetoresistance was fitted by the parabolic dependence $\rho(H) \propto (H/T)^2$ in the vicinity of the critical point, $G \sim G_c$.

[27] D. Simonian, S.V. Kravchenko, M.P. Sarachik, V.M. Pudalov, Cond-mat/9 712 223.

[28] V.M. Pudalov, G. Brunthaler, A. Prinz, G. Bauer, Proc., EP2DS-12, p. 57, Tokyo, 1997. Physica B (1998), in print.

[29] S.V. Kravchenko, D. Simonian, M.P. Sarachik, A.D. Kent, V.M. Pudalov, Cond-mat/9 709 255.

[30] I. Glozman, C.E. Johnson, J.W. Jiang, Phys. Rev. Lett. 74 (1995) 594.

[31] V. Dobrosavljević, E. Abrahams, E. Miranda, S. Chakravarty, Phys. Rev. Lett. 79 (1997) 455.

[32] A.M. Finkel'stein, Z. Phys. B 56 (1984) 189. Sov. Sci. Reviews/section A- Physics Reviews, Ed. I.M. Khalatnikov, 14 (1990) 3.

[33] V.M. Pudalov, M. D'Iorio, J.W. Campbell, S.V. Kravchenko, Phys. Rev. Lett. 70 (1993) 1866.

[34] V.M. Pudalov, G. Brunthaler, A. Prinz, G. Bauer, Cond-mat/9 801 077. JETP Lett. 68 (1998). [Pis'ma ZhETF 68 (1998) 415].

[35] S.I. Dorozhkin, E.B. Ol'shanetskii, JETP Lett. 46 (1987) 502. S.I. Dorozhkin, Sol. St. Commun. 72 (1989) 211.

[36] H.L. Störmer, Z. Schlesinger, A. Chang, D.C. Tsui, A.C. Gossard, W. Wiegmann, Phys. Rev. Lett. 51 (1983) 126. J.P. Eisenstein, H.L. Störmer, V. Narayanamurtu, A.C. Gossard, W. Wiegmann, Phys. Rev. Lett. 53 (1984) 2579.

[37] F. Stern, Phys. Rev. Lett. 44 (1980) 1469.

[38] A. Gold, V.T. Dolgopolov, Phys. Rev. B 33 (1986) 1076.

[39] K.K. Choi, Phys. Rev. B 28 (1983) 5774.

[40] S.V. Kravchenko, W. Mason, G.E. Bowker, J.E. Furneaux, V.M. Pudalov, M. D'Iorio, in: EP2DS-11 Workbook, Univ. of Nottingham, 463, 1995.

[41] The relevant energy in the spin–orbit coupling mechanism, is $\Delta_s = h/\tau_s$. The spin relaxation time τ_s is $\sim \tau_p/\Delta_0^2$ for Dyakonov–Perel spin-relaxation mechanism, which seems to be dominant in Si–MOS structures with strong gradient of the potential at the interface.

[42] S. Hikami, A.I. Larkin, Y. Nagaoka, Prog. Theoret. Phys. 63 (1980) 707.

[43] V.M. Pudalov, G. Brunthaler, A. Prinz, G. Bauer, E.M. Dizhur, to be published elsewhere.

Self-organized epitaxial growth of low-dimensional structures

J. Tersoff *

IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

Abstract

In heteroepitaxy, misfit strain can lead to a variety of fascinating phenomena during growth. These include self-organization of regular arrays of islands ("quantum dots"), of step bunches (which may serve as a template for growth of "quantum wires"), and of compositionally modulated layers ("quantum wells"). Some recent theoretical treatments of these phenomena are described. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 68.65.+g; 81.15.−z; 85.40.Ux; 68.55.−a

Keywords: Strain; Epitaxy; Self-assembly; Self-organization; Quantum dots; Quantum wires; Superlattices

1. Introduction

Fabrication of nano-scale low-dimensional structures, such as "quantum dots" and "quantum wires", represents a difficult challenge. One intriguing approach uses self-assembly. This approach has obvious limitations – at best one could obtain a regular array of nanostructures, suitable for storage or opto-electronic applications but not for logic devices. In practice, the limited degree of order and uniformity makes even the simplest applications a challenge. Here, I review some recent theoretical studies, which suggest strategies for fabricating relatively uniform low-dimensional structures by self-assembly.

The term "self-assembly" is perhaps too broad, since even crystal growth is a kind of self-assembly. Our interest here is in self-organized growth of uniform nano-scale features. Such growth can be driven

by an underlying instability, such as misfit strain in heteroepitaxial systems, or spinodal decomposition in alloys. Then uniformity may arise from a competition between the thermodynamic instability and kinetic effects.

The work reviewed here focuses on the case of strain as a driving force, as in the case of a misfitting heteroepitaxial layer. I briefly review how strain can lead to self-organization of zero-dimensional "quantum dots", one-dimensional "quantum wires", and two-dimensional "quantum well" superlattices.

2. Quantum dots

In heteroepitaxial growth, misfit strain can drive nucleation of three-dimensional "islands" of a strained material on the surface. Once buried in a matrix with a larger band gap, these islands can serve as "quantum dots". However, the energetics are consistent

* Corresponding author. E-mail: tersoff@us.ibm.com.

with random nucleation and coarsening [1], tending to produce a broad size distribution. (An interesting exception occurs when growth is restricted to two dimensions, as for monolayer films [2].)

Technological applications, in contrast, require a narrow size distribution. Two possible approaches to obtaining narrower size distribution suggest the opportunities for self-assembly.

In multilayer structures, buried dots can influence the nucleation in subsequent layers, leading to self-organization of a more ordered and uniform array [3,4]. The basic mechanism is quite simple. One grows a thin layer of Ge on Si (0 0 1) (or some similar system, such as InAs on GaAs). This layer grows as islands, which are then buried under a "spacer layer" of Si. The buried island creates a strained region on the surface above. In the next layer of Ge, the islands nucleate preferentially above the buried dots.

However, if two buried islands are close together (relative to the spacer layer thickness), there is a single favored site for nucleation, half-way in between. Thus with growth of successive layers, short island–island spacings are progressively filtered out. The final result is a rather uniform, almost ordered array of islands in each layer, with lateral spacing determined by the Si spacer-layer thickness [3,4]. (The poorly ordered initial layers remain buried underneath the later more uniform layers. The inferior uniformity of the early layers may be a drawback in some applications.)

Even for a single layer, a shape transition at a critical island volume can lead to novel coarsening behavior, producing a much narrower size distribution than classic Ostwald ripening [5]. As a Ge island grows, it retains its (1 0 5)-facetted pyramid shape up to a critical volume, where it assumes a multi-facetted shape. The island's chemical potential drops abruptly as a result, so it can extract material from neighboring islands in a coarsening process. Thus, once enough of these larger islands have formed, the remaining pyramidal islands quickly shrink and disappear. Then there is a transitional period with only multi-facetted islands, all having sizes just slightly above the critical volume [5].

In this case there is less spatial self-organization than for the multi-layer approach, and the size uniformity is achieved only at a larger island size. Therefore, this approach may have fewer potential applications, but in any case it illuminates the role of island shape in coarsening.

3. Quantum wires

On a vicinal surfaces, strain creates an attractive interaction between the atomic steps, driving them to assemble into bunches [6,7]. During growth, this energy-driven bunching competes with debunching due to step-flow kinetics. The result (according to theoretical simulations) can be self-organization of regular arrays of uniform step bunches [8].

Such a step-bunch array could serve as templates for growth of quantum wires. This general approach was proposed many years ago by Petroff, using the steps on a vicinal surface as the template. But single steps tend to meander, and are not uniformly spaced. The step bunches are much straighter and more uniform, and so should provide a superior template. Moreover, the height is determined by the number of steps per bunch, which can be controlled via the growth rate or temperature, as well as the surface miscut.

4. Spontaneous superlattice formation

For alloys, the presence of a uniform array of step bunches leads to the prediction of another kind of self-organized growth. The non-uniform stress across the step bunch can lead to preferential incorporation of one species at the top of the bunch, and the other at the bottom. For example, in growth of SiGe on Si, the compressive stress is partially relieved at the top of the bunch, but increased at the bottom. Therefore, Ge incorporates preferentially at the step at the top of the bunch, with higher Si incorporation at the bottom step. The predicted result is spontaneous formation of a compositional superlattice, despite the continuous deposition of a constant alloy composition [9].

5. Other mechanisms of self-organization

There are a variety of other mechanisms which could potentially lead to self-organization of nanoscale structures. Two preliminary analyses suggest that spinodal decomposition during growth could lead to spontaneous superlattice formation [10,11]. And various ways have been suggested in which even in thermal equilibrium a system of ordered structures could form [12–14]. These include formation of stress

domains associated with surface facets [12] or phase boundaries [13,14]. It has even been suggested that, in certain cases, three-dimensional islands on a surface could form an equilibrium ordered structure, rather than coarsening [15]. Such suggestions are quite intriguing, although uncertainties remain regarding the actual observation of these effects.

Acknowledgements

References

[1] J. Tersoff, F.K. LeGoues, Phys. Rev. Lett. 72 (1994) 3570.
[2] J. Tersoff, R.M. Tromp, Phys. Rev. Lett. 70 (1993) 2782.
[3] J. Tersoff, C. Teichert, M.G. Lagally, Phys. Rev. Lett. 76 (1996) 1675.
[4] C. Teichert, L.J. Peticolas, J.C. Bean, J. Tersoff, M.G. Lagally, Phys. Rev. B 53 (1996) 16 334.
[5] F.M. Ross, J. Tersoff, R.M. Tromp, Phys. Rev. Lett. 80 (1998) 984.
[6] J. Tersoff, Y.H. Phang, Z. Zhang, M.G. Lagally, Phys. Rev. Lett. 75 (1995) 2730.
[7] J. Tersoff, Phys. Rev. Lett. 74 (1995) 4962.
[8] F. Liu, J. Tersoff, M.G. Lagally, Phys. Rev. Lett. 80 (1998) 1268.
[9] J. Tersoff, Phys. Rev. Lett. 77 (1996) 2017.
[10] J. Tersoff, Phys. Rev. B 56 (1997) R4394.
[11] A. Barabasi, Appl. Phys. Lett. 70 (1996) 764.
[12] V.I. Marchenko, Sov. Phys. JETP 54 (1981) 605.
[13] V.I. Marchenko, JETP Lett. 33 (1981) 381.
[14] O.L. Alerhand, D. Vanderbilt, R.D. Meade, J.D. Joannopoulos, Phys. Rev. Lett. 61 (1988) 1973.
[15] V.A. Shchukin, N.N. Ledentsov, P.S. Kop'ev, D. Bimberg, Phys. Rev. Lett. 75 (1995) 2968.

Patterned high-index substrates as templates for novel quantum wire and quantum dot arrays in hydrogen-assisted MBE

Klaus H. Ploog *, Richard Nötzel

Paul Drude Institute for Solid State Electronics, Hausvogteiplatz 5/7, D-10117 Berlin, Germany

Abstract

We report new approaches to fulfill the stringent criteria on the geometrical features of GaAs quantum wire and quantum dot structures to be met for useful devices. The clever combination of self-organized growth on high-index substrates with lithographic patterning and atomic-hydrogen assistance allows fabrication of GaAs quantum wires and quantum dots with superior electronic properties by molecular beam epitaxy. In addition, our method allows the precise positioning of the wires and dots on specific wafer areas. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Quantum dot; Quantum wire; Molecular beam epitaxy

1. Introduction

Over the past decade there have been numerous attempts to fabricate semiconductor quantum wire (QWR) and quantum dot (QD) structures for fundamental research as well as for device applications [1]. Although some progress has been made in the various fabrication methods, the QWR and QD structures available today have not justified the original claims for highly improved devices [2], at least not in the way originally proposed, and at present we are not sure whether or not we can ever advance toward useful devices. For a significant step forward we, obviously, need techniques that combine some form of patterning to control the location of the nanostructures and, in addition, self-organization to construct features of

the order of 5–10 nm reliably and repeatedly. In this paper, we demonstrate that some ingenious combination of self-organized growth on high-index substrates with lithographic patterning and atomic-hydrogen assistance allows the fabrication of superior-quality GaAs QWR and QD arrays by molecular beam epitaxy (MBE). Patterning of the substrate not only positions the QWRs and QDs on the wafer, but it also provides a unique means to significantly reduce any fluctuations of the feature size. Step bunching induced by atomic hydrogen increases the feature sizes by an order of magnitude and stabilizes them reliably.

The paper is organized as follows. We start with a brief discussion of the concept to combine self-organized growth with substrate patterning and hydrogen assistance to generate a template for QWR and QD formation during MBE. Then, after a short description of sample growth, we will in detail discuss

* Corresponding author. Tel.: 0049 30 203 77 365; fax: 0049 203 77 201; e-mail: ploog@pdi-berlin.de.

three distinct achievements:

1. Formation of GaAs quantum wires at the fast grow-
   ing sidewalls of [0 1 $\bar{1}$] oriented mesa stripes on
   (3 1 1) substrates [3]. Their smooth convex-surface
   profile without any faceting and the strong carrier
   confinement result in unique electronic properties
   observable even at room temperature.
2. Formation of conductive GaAs quantum wire ar-
   rays by hydrogen-induced step bunching on (3 3 1)
   substrates [4]. The improved uniformity with a step
   height of 12 nm and a lateral periodicity of 250 nm
   extends over more than 10 μm long straight step
   edges. The electron conductivity parallel to the step
   edges is by more than one order of magnitude larger
   than normal to the edges.
3. Formation of GaAs quantum dot arrays by
   hydrogen-induced step bunching perpendicular to
   the [0 1 $\bar{1}$] oriented mesa stripes on (3 1 1) sub-
   strates [5]. These quantum dots exhibit minimized
   size fluctuations as demonstrated by the extremely
   narrow single luminescence emission with 60 μeV
   line width (limited by instrumental spectral reso-
   lution).

2. Concept of template formation and transfer

Condition for the formation of QWR and QD struc-
tures directly by epitaxial growth is the transfer of the
pattern of any template to the generation of vertical
and lateral materials interfaces for carrier confine-
ment. Promising templates are patterned substrates
[6–9], where the surface profile after growth is deter-
mined by the actual geometric pattern, and high-index
substrates [10], which form corrugations composed of
low-index facets to lower the surface energy at
growth conditions. In addition, in lattice-mismatched
layer systems the strain relaxation via the Stranski–
Krastanow growth mode can produce coherently
strained islands [11–14] which act as quantum dots.
However, the size uniformity of such islands is not
yet sufficient for any meaningful device application.

The pattern transfer of patterned substrates relies
on the formation of crystal facets due to the pref-
erential migration of group-III adatoms among the
competing growth planes from the slower to the faster
growing facets [8]. This material transfer from planes
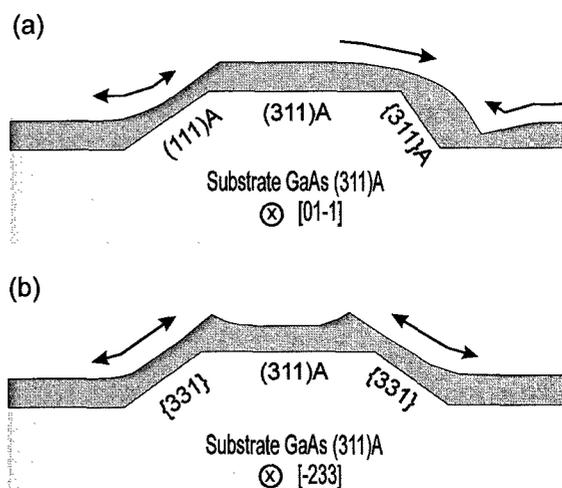
Fig. 1. Schematic illustration of growth mode on patterned
GaAs(311)A substrates with mesa stripes oriented (a) along [0 1 $\bar{1}$]
and (b) along the perpendicular [$\bar{2}$ 3 3] direction. The arrows in-
dicate the preferential migration of Ga adatoms resulting in the
selectivity of growth across the mesa sidewalls.

with larger adatom migration lengths to those with
smaller ones is directly related to the atomic configu-
ration of the respective planes and to the growth con-
ditions. The surface profiles after overgrowth are thus
defined by the nearest facets with the slowest growth
rates (or the highest adatom migration lengths). In
patterned growth on GaAs(1 0 0) and (1 1 1) sub-
strates this growth mechanism produces a variety of
crystal shapes. Depending on growth conditions, ori-
entation and shape of the patterns, various facets can
develop and coexist during growth, most commonly
represented by slow growing {1 1 1}, {1 1 0}, {3 1 1}
or {4 1 1} planes. Combined with the differences in
migration length between Al and Ga adatoms, the
accumulation and depletion of material on distinct
facets was successfully used for the fabrication of
V-groove and ridge-type quantum wires.

Lithographic patterning combined with the concept
of spontaneous patterning (self-organization) on high-
index substrates during epitaxial growth introduces a
new dimension of flexibility and control for the fabri-
cation of QWR and QD structures [3]. The selectivity
of growth on patterned (3 1 1)A substrates differs
qualitatively from that on low-index (1 0 0) and (1 1 1)
substrates, as schematically depicted in Fig. 1. During
MBE growth of $Al_xGa_{1-x}As$ over mesa stripes along
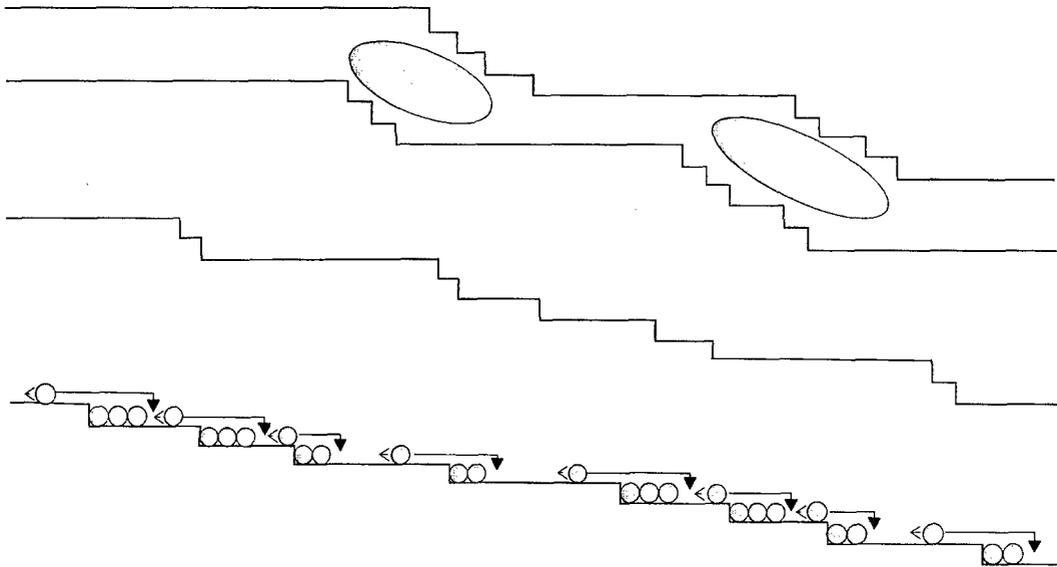
Fig. 2. Schematic illustration of step bunching using a one-dimensional step-flow model. The preferential incorporation of adatoms on step-down sites leads to an attractive interaction of step edges. In the upper part the formation of QWR on the multiatomic step array is indicated.

[0 1 $\bar{1}$], a fast growing sidewall evolves on one side, i.e. in the sector towards the next (1 0 0) plane. This sidewall exhibits a smooth, convex-curved surface profile without any faceting. On the opposite side, i.e. in the sector towards the next (1 1 1) plane, a slow growing (1 1 1) sidewall is formed. The formation of the fast growing, convex-curved sidewall relies on the preferential migration of Ga adatoms from both the mesa top and mesa bottom towards the sidewall. The direction of adatom migration is thus opposite to that on low-index surfaces, where slow growing side facets are formed due to the preferential adatom migration away from the sidewall and a surface profile at the mesa bottom with opposite, i.e. concave, curvature is produced. If, however, the mesa stripes are oriented along [$\bar{2}$ 3 3], two slow growing side facets close to the next {3 3 1} planes develop (Fig. 1b) leading to surface profiles similar to those observed on patterned GaAs(1 0 0) substrates. The growth selectivity on patterned (3 1 1)A substrates can be applied directly to the formation of lateral QWR arrays on the fast growing mesa sidewall along [0 1 $\bar{1}$] when the height of the mesa stripes is adjusted to the quantum-size regime of 10–20 nm [3]. Even for these shallow mesa heights the unique growth mode is conserved

and produces a well developed convex-curved surface profile with a distinct corner at the bottom of the sidewall. We assume that the highly anisotropic microscopic surface corrugation along [$\bar{2}$ 3 3] stabilizes this particular growth front and prevents roughening. In addition to the improved ordering and size uniformity on patterned (3 1 1)A substrates, the precise positioning of the nanostructured features meets one of the most important demands for device applications.

Arrays of multiatomic steps formed by step bunching can provide useful templates for the fabrication of QWR and QD structures. Accumulation of steps of monolayer height in bunches takes place when the lateral growth velocity of large terraces is larger than that of smaller ones, leading to an attractive interaction of step edges [15]. This happens when the incorporation probability of adatoms on step-down sites is larger than that on step-up sites, as schematically shown in Fig. 2 . The evolution of multiatomic step arrays during growth of, e.g., $Al_x Ga_{1-x} As$, heterostructures can thus produce periodic alloy fluctuations and/or thickness modulations [16,17]. However, the kinks existing at monolayer steps and the meandering of terrace widths are to a large extent transferred to the multiatomic step arrays, resulting in

pronounced variations of the height of the bunched steps and in crossings of the step edges. In contrast to step arrays on vicinal low-index planes, the step structure on high-index substrates is more stable against kink formation due to the similar size of the microscopic terraces and steps, which is related to the surface reconstruction of the constituent planes [18]. We have chosen GaAs(3 3 1)A, (2 1 0), and (3 1 1)A substrates as the three distinct representatives composed of equal units of the low-index (1 0 0), (1 1 0), and (1 1 1) faces to study the promotion of step bunching by atomic hydrogen. GaAs(3 3 1) substrates, e.g., develop microscopic step arrays along $[1\bar{1}0]$ at low substrate temperature, which are composed of two unit meshes of (1 1 0) terraces followed by two unit meshes of (1 1 1) steps. At elevated substrate temperature, bunching of these microscopic step arrays leads to straight coherently aligned multiatomic step arrays along $[1\bar{1}0]$. The multiatomic step height of the quasi-periodic arrangement increases with the substrate temperature, whereas the lateral periodicity remains almost unchanged. In the presence of atomic hydrogen the step height increases by more than a factor of two, thus yielding very straight step arrays of up to 10 nm height [4]. Although the distinct mechanism of the hydrogen effect in solid-source MBE is not yet clarified [19,20], we assume that hydrogen enhances the incorporation rate on step-down sites thus promoting step bunching. On GaAs(1 0 0) surfaces atomic hydrogen is known to passivate free bonds which alters the chemical structure. In particular, at step edges this effect is large enough to modify the adatom incorporation.

On GaAs(3 1 1)A the natural multiatomic step arrays generated by atomic hydrogen run along $[\bar{2}33]$ [21]. This direction is perpendicular to the quantum wires formed in patterned growth on the fast growing sidewall of shallow mesa stripes along $[0 1 \bar{1}]$. The superposition of the natural step arrays generated in hydrogen-assisted MBE over the sidewall quantum wires on patterned (3 1 1)A substrates in conventional MBE thus generates an additional periodic thickness modulation of the wires, resulting in an ordered array of very uniform quantum dots along the mesa sidewall [5]. The lateral periodicity of the step arrays in hydrogen-assisted MBE of about 40 nm is comparable in size to the width of the wires in patterned growth.

3. Experimental

The GaAs(3 1 1)A substrate was patterned by using either optical lithography and wet chemical etching in $H_2SO_4 : H_2O_2 : H_2O$ (1 : 8 : 40) solution or holographic lithography and dry etching. The mesa stripes of width ranging from 80 to 0.5 μm and depth ranging from 400 to 15 nm were oriented along the two perpendicular $[0 1 \bar{1}]$ and $[\bar{2}33]$ azimuths. To remove the photoresist and clean the samples, concentrated H_2SO_4 was used prior to sample loading. Additional cleaning and removal of the native oxide was carried out in the preparation chamber by atomic hydrogen supplied by a cracker cell with a hot (1800°C) W filament (CreaTec HLC-35/283) [4]. A second atomic hydrogen source was employed in the growth chamber to promote step bunching during growth. The GaAs/$Al_{0.5}Ga_{0.5}As$ heterostructures on (3 1 1)A substrates were grown at 620°C with a substrate rotation of 6 rpm [3]. The growth rates were 0.5 μm/h for GaAs and 1.0 μm/h for $Al_{0.5}Ga_{0.5}As$ with V–III flux ratio of five. Growth on (3 3 1) substrates was carried out in the temperature range 530–625°C to optimize the uniformity of the multiatomic step arrays [4].

4. Results and discussion

4.1. Quantum wires on patterned high-index substrate

Our systematic investigation of the mode and the selectivity of growth on a number of stripe-patterned GaAs substrates with different [n 1 1]A and B orientation [22] has revealed that the formation of a fast growing sidewall with a smooth convex curved surface profile without facets on one side of the $[0 1 \bar{1}]$ mesa stripes is unique to (3 1 1)A substrates (Fig. 3). While the growth mode was studied on 0.4 μm high mesa stripes, the formation of quantum wires through the fast growing mesa sidewall along $[0 1 \bar{1}]$ on patterned GaAs(3 1 1)A substrates requires a reduction of the height of the mesa stripes to about 15 nm (quantum size regime) [23]. Even for these shallow mesas the distinct growth mode is preserved and a well-defined convex curved surface profile with a distinct corner at the bottom of the sidewall is formed. As shown in Fig. 4 the selectivity of growth is not hampered, and

(a) GaAs well　GaAs wire　GaAs well

GaAs cap

AlGaAs

AlGaAs

GaAs buffer

[311]↑
[01-1] ⊗　Substrate GaAs (311)A
[-233] →

50 nm

(b)

GaAs well　　　GaAs wire　GaAs well
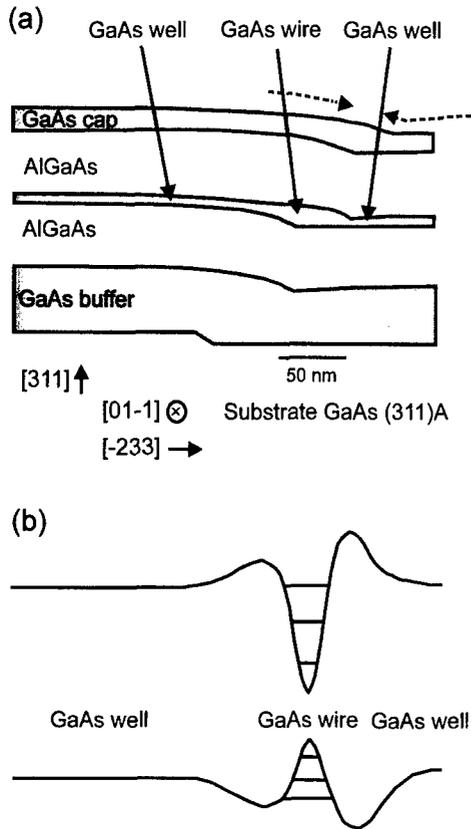
Fig. 3. Schematic cross-section of sidewall GaAs QWR structure along shallow [0 1 $\bar{1}$] mesa on patterned (3 1 1)A substrate (a) and real-space energy band diagram along the lateral direction (b). The 1D confinement arises from the strong variation of the quantum well width along [$\bar{2}$ 3 3].

Fig. 4. AFM image of the convex-curved surface of the GaAs QWR (a) and cross-sectional TEM image taken along [0 1 $\bar{1}$]. The 12 nm thick QWR originates from the preferential migration of Ga from the 6 nm wide quantum well towards the {311}A sidewall which results in a 1D confinement over a lateral QWR width of 50 nm.

QWRs are formed with a width of several 10 nm and a thickness of about twice the quantum well in the flat part of the mesa. The selectivity of growth is comparable to that in V-groove structures. The vertical stacking of the QWRs in growth direction has been achieved with the growth mode exactly reproducing the QWR dimensions [22]. This opens a direct path to the fabrication of 3D QWR arrays with a large active volume. A distinct advantage for further device processing by lithographic techniques is that the shallow mesa height results in quasi-planar structures.

The single GaAs QWR depicted in Fig. 4, which was formed at the 15 nm high fast growing mesa sidewall, has a width of 50 nm and a height of 12 nm and is embedded in 50 nm $Al_{05}Ga_{0.5}As$ lower- and upper-barrier layers. The AFM image of Fig. 4a demonstrates a very smooth convex-curved surface morphology without any roughening of the growth front. As a result, GaAs single QWR and QWR arrays can be fabricated on patterned GaAs(3 1 1)A substrate whose geometric size is solely related to the evolution of the fast growing sidewall of the shallow mesa stripes with the height easily controlled on the nanometer scale by the etching process without any need for sophisticated lithographic techniques. In addition to the ease of adjusting the shape and hence the optical properties, the QWR can be exactly positioned on the wafer.

The strong lateral confinement and the structural perfection of these QWR manifest themselves in narrow luminescence lines and high luminescence efficiency up to room temperature [3], as shown in Fig. 5 . On the flat area of the mesa only the peaks originating from the 6 nm quantum well are observed, while the spectra excited at the sidewall split into two (three)
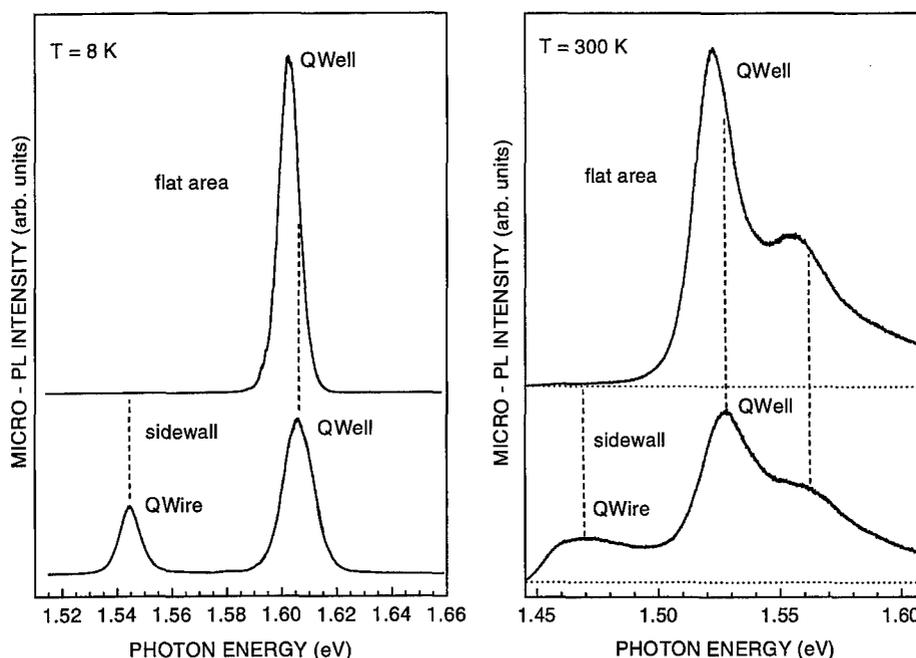
Fig. 5. Five μm-PL spectra taken at 8 K (left) and 300 K (right) from the flat area of the mesa (upper curves) and from the {311} sidewall (lower curves).

lines at lower and higher energy due to the emission from the quantum wire (1.540 eV at 8 K) and from the adjacent quantum well (1.605 eV at 8 K). The PL peak energy position corresponds to a QWR thickness of 12 nm which is in excellent agreement with the TEM data of Fig. 4. The blue shift of the quantum well emission near the sidewall compared to the flat part of the mesa confirms the existence of narrower well regions in the vicinity of the wire which form the lateral energy barriers.

Using scanning near-field optical microscopy (SNOM) with a spatial resolution of 250 nm [24] allowed us to precisely map the lateral potential barriers in a single GaAs QWR due to the thinning of the wells. The confinement energies of electrons and holes with respect to the adjacent 6 nm quantum well turned out to be 60 and 15 meV, resp., as indicated in the band diagram of Fig. 3b. In addition, the real-space transfer and the trapping of photogenerated carriers from the adjacent quantum well as well as the quasi-1D subband structure was studied in detail by temperature-dependent SNOM measurements. The observed 1D subband separation of 12 meV for

electrons is in good agreement with a simple theoretical model taking into account the actual QWR shape derived from TEM images.

The shape and hence the electronic properties of these sidewall QWR can be tuned by adjusting the mesa height through the etching process. Starting from a 3 nm thick GaAs quantum well and $x = 0.5$ in the $Al_xGa_{1-x}As$ barriers, the lateral QWR confinement energy reaches 100 meV for shallow mesas with step heights of 10–30 nm. This value of the confinement energy reproduces the 2 : 1 growth selectivity for the thickness of the QWR to that of the well. With increasing 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. The vertical stacking of the sidewall QWR on patterned GaAs(3 1 1)A as well as the formation of periodic lateral arrays with 4 μm periodicity has been demonstrated successfully [21,22]. If, however, the lateral periodicity of QWR arrays is reduced to the submicron range, it becomes smaller than the migration

length of the group-III adatoms. Therefore, in particular in connection with vertical stacking of QWR, the growth mechanism at work to form the convex-curved sidewall should be questioned. On the other hand, the optical emission characteristics are expected to be dominated by radiative transitions in the QWR region. In submicron-pitch QWR arrays the lateral periodicity is also smaller than the diffusion length of free carriers. Carriers generated in the connecting (thinner) quantum wells are thus transferred efficiently into the active (thicker) QWR region with lower ground-state energy. Hence, stacking of QWR in dense three-dimensional arrays on submicron gratings is the ultimate goal for device application. We have fabricated stacks of GaAs QWR on GaAs(3 1 1)A patterned with 500 nm-pitch periodic gratings by holographic lithography and dry etching to a depth of 15 nm [25]. Using cross-sectional TEM we found that the fast growing sidewall develops the convex-curved surface profile while the opposite slow growing sidewall is entirely smeared out. Now, being smaller than the surface migration length of adatoms, the washed-out concave-shaped sidewalls establish a uniform template for the quantum wells connecting the bottom step corner of the fast growing sidewalls. The reduced growth selectivity of slow growing sidewalls on submicron gratings, which hamper the formation of V-groove structures, results in a greatly improved uniformity for these sidewall QWR arrays on patterned GaAs(3 1 1)A. This selflimiting lateral growth mechanism leads to three-dimensional QWR stacks with virtually the same shape and size of the constituent QWR, as depicted in the cathodoluminescence spectra of Fig. 6 . The spectra are dominated by the QWR emission. At low temperatures a weak high-energy emission appears due to carrier localization in the quantum well region inbetween the individual wires. The narrow line width of the QWR emission and its observation up to room temperature demonstrate the structural perfection and the uniformity in size and shape of the stacked wires. The disappearance of the well luminescence reveals thermally activated delocalization of the carriers in the well allowing perfect transfer into the QWR before radiative recombination occurs.

On the submicron-pitch grating of (3 1 1)A GaAs the selectivity of growth is higher compared to the wider mesa stripes, because the Ga adatoms can migrate directly from the slow growing to the fast grow-
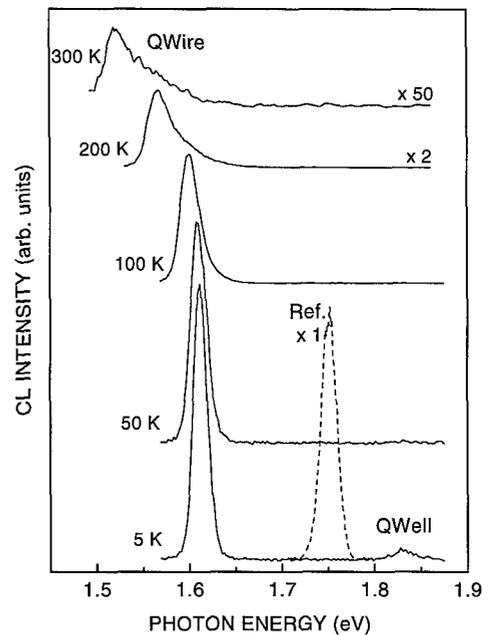
Fig. 6. CL spectra of submicron-pitch QWR array on patterned GaAs(3 1 1)A between 5 K and room temperature (solid lines) and of the reference quantum well on the planar (3 1 1)A substrate (dashed line).

ing sidewall. This increases the thickness of the QWR at the direct expense of the thickness of the entire connecting quantum well. As a result the redshift of the QWR luminescence and, in particular, the blueshift of the adjacent-well luminescence with respect to the reference quantum well are more pronounced as compared to QWR at wider mesa stripes. The stronger thinning of the adjacent well due to the higher growth selectivity gives rise to a lateral confinement energy as high as 220 meV for the present QWR structure. This shifts the subband spacing for electrons into the range of $k_B T$ at room temperature (25 meV). The submicron-pitch QWR array thus exhibits strong luminescence emission up to room temperature, as shown in Fig. 6. The observed overall intensity drop by one order of magnitude from 5 to 300 K due to nonradiative recombination and/or thermal re-emission of carriers into the $Al_{0.5}Ga_{0.5}As$ barriers underlines the high structural perfection of the QWR array [25].

We have also fabricated strained $Ga_x In_{1-x}As$ sidewall quantum wires on patterned GaAs(3 1 1)A [26], which exhibit a strong enhancement of the optical

nonlinearity compared to two-dimensional quantum wells, due to their strong in-plane piezoelectric field. These strained QWR arrays are important for the fabrication of optical modulators operating with high modulation depth at low incident power.

4.2. Conductive quantum wires by hydrogen induced-step bunching

The hydrogen-induced one-dimensional step bunching, which is most pronounced on GaAs(331)A, has been utilized for the fabrication of wire-like modulation-doped GaAs/$Al_xGa_{1-x}As$ heterostructures. In a detailed study we have determined the dependence of the step height on the MBE growth parameter to optimize their uniformity [4]. Very uniform arrays with high steps are produced on GaAs buffer layers grown at high ($>610°C$) temperature without hydrogen which are subsequently overgrown at reduced ($<550°C$) temperature with atomic hydrogen irradiation. The AFM top view in Fig. 7 reveals that the step array with a height of 12.5 nm is straight over more than 10 μm length without any crossing of the step edges.

Inspection of GaAs/$Al_xGa_{1-x}As$ heterostructures by top-view AFM as well as the RHEED diffraction patterns during growth reveal the $Al_xGa_{1-x}As$ layers to reproduce the surface morphology of the underlying GaAs layer. The n-type Si-modulation-doped single heterostructures exhibit a pronounced anisotropy of the conductivity (Fig. 8) which correlates with the height of the step arrays [4]. The ratio of conductivity parallel and perpendicular to the step edges depicted in Fig. 8 shows that in the sample having a 6 nm high step array the anisotropy of the conductivity reaches a value of 14 at low temperature. In this sample the electron mobility parallel to the step edges is 2.2×10^5 cm²/V s with an average two-dimensional sheet carrier density of 3×10^{11} cm^{-2}, as determined by Shubnikov–de Haas and Hall effect measurements at 1.5 K. For a given step height the anisotropy of the conductivity increases with the carrier concentration after illumination. This behavior, together with the dependence of the conductance anisotropy on step height and on temperature can, in principle, be related to anisotropic interface scattering or to lateral miniband formation. The mechanisms governing the electronic properties of these highly anisotropoic het-
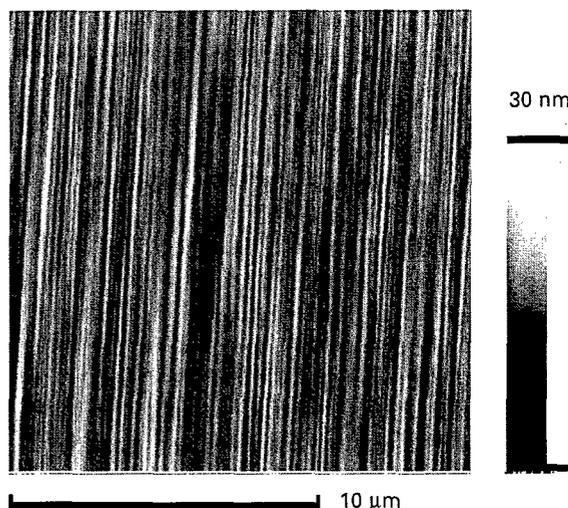
Fig. 7. AFM top view of 0.25 μm GaAs layer grown at 545°C with atomic hydrogen irradiation over a 0.25 μm GaAs buffer layer grown at 625°C without hydrogen.
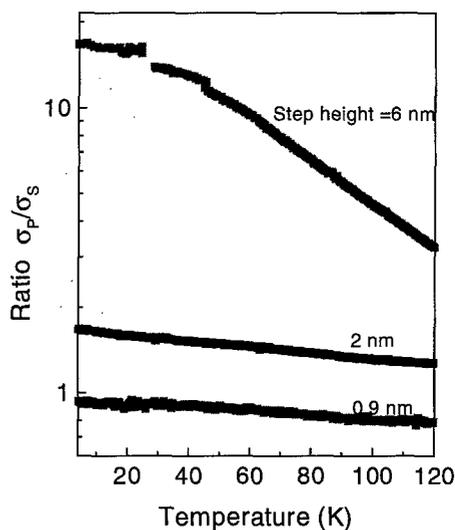

Fig. 8. Anisotropy of conductivity as a function of temperature in n-type Si-modulation-doped GaAs/$Al_xGa_{1-x}As$ heterostructures with step-bunched interface corrugations on GaAs(331)A. The height of the interface corrugation is indicated.

erostructures are currently analysed by magnetotransport measurements at low temperature and are partly presented at this conference by Friedland et al. [27].

In samples with the highest steps the anisotropy of the conductivity is maintained up to room temperature

reaching a value of two for the two orthogonal conductivities. Novel transistor concepts based on the lateral modulation of carrier density and mobility in the conductive channel can be realized with these structures. In such devices the conductivity is controlled without changing the carrier density which limits the switching speed of conventional field-effect transistors.

4.3. Uniform quantum dot arrays by step bunching across sidewall quantum wires

The natural multiatomic step arrays on planar GaAs (3 1 1)A generated by atomic hydrogen are oriented along $[\bar{2} 3 3]$ [21]. This is perpendicular to the QWR direction formed in patterned growth at the fast growing sidewall of shallow $[0 1 \bar{1}]$ mesa stripes [3], which is described in Section 4.1. Therefore, superimposing the natural step arrays in hydrogen-assisted MBE to the sidewall quantum wires on patterned (3 1 1)A substrates in conventional MBE generates an additional periodic thickness modulation of the wires and produces an ordered array of quantum dots along the mesa sidewall (see Fig. 9 for illustration) [5]. The lateral periodicity of the step arrays in hydrogen-assisted MBE of about 40 nm is comparable in size to the width of the wires in conventional patterned growth. Hence, the lateral extension of the dots is 40–50 nm in both the $[0 1 \bar{1}]$ and the $[\bar{2} 3 3]$ direction with the height determined by the thickness of the GaAs layer. Detailed AFM studies on structures starting with a 3 nm GaAs quantum well revealed that in hydrogen-assisted MBE the surface profile shows the same convex curvature of the fast growing sidewall. The natural one-dimensional surface corrugation generated by atomic hydrogen is established on the convex-curved part of the sidewall without degradation of uniformity. The spatial correlation of the step bunches continues over the sidewall from the mesa top to the mesa bottom without any displacement at the bottom step corner. In this way, the QWR thickness is modulated equidistantly yielding a periodic array of quantum dots with a uniform size.

The existence of uniform QD arrays manifests itself in PL spectra dominated by a single sharp peak, as shown in Fig. 10[5]. These micro-PL spectra are dominated by one single sharp peak without any background emission over an energy of 20–30 meV, indicating the superior size uniformity of this QD
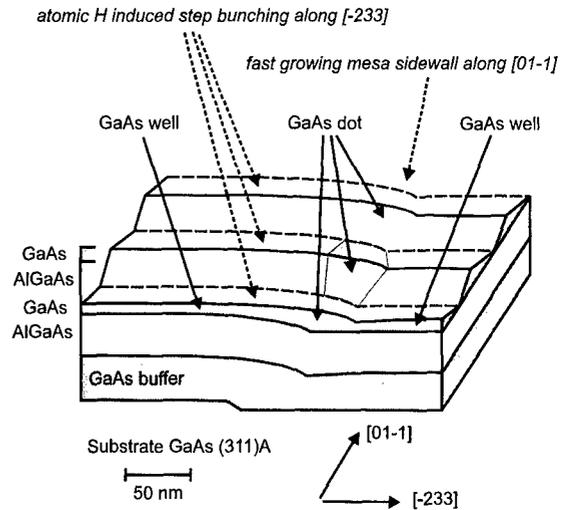
Fig. 9. Schematic illustration of GaAs quantum dot array formed by the natural quasi-periodic step bunching along $[\bar{2} 3 3]$ in atomic hydrogen-assisted MBE perpendicular to the quantum wire formation in patterned growth at the fast growing sidewall of shallow $[0 \bar{1} 1]$ mesa stripes.
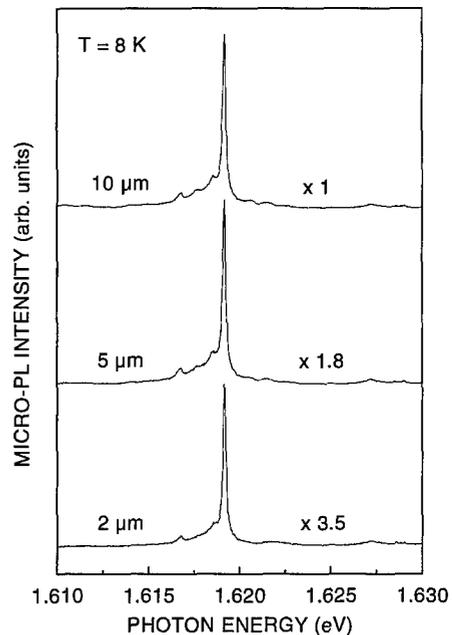

Fig. 10. High-resolution μ-PL spectra of the GaAs dot array along the sidewall for three different diameters of the optical probe area. The observed line width of 60–70 μeV is limited by the spectrometer resolution.

array. The periodic step bunching in hydrogen-assisted MBE across the sidewall, where only a segment of 50 nm width across the wire region is relevant, leads to a very regular thickness modulation of the wire. The emission of the resulting QD array is thus governed by the peaked energy dependence of the density of states characterized by ultrasharp PL at a well-defined energy position. The PL spectra of the QD array do not change when the diameter of the optical probe area is increased from 2 to 10 μm (Fig. 10). This corresponds to a change in contribution from several 10 to several 100 dots. Therefore, the combination of hydrogen-induced one-dimensional surface corrugation with patterned growth on GaAs(3 1 1)A substrate produces dense arrays of uniform QDs over areas which are already large enough for further processing towards device structures.

5. Conclusion

At the 6th Mauterndorf Winterschool held in 1990, H. Sakaki [2] has summarized some of the most stringent requirements on the geometrical features of QWR and QD structures for their device application and some important needs in fabricating such structure with uniform feature sizes of a few ten nanometer or less. In this paper we have demonstrated that the fabrication of high-quality GaAs QWR and QD structures meeting above criteria is achievable by MBE using the intelligent combination of selforganized growth on high-index substrates with lithographic patterning and atomic-hydrogen assistance. In this way, GaAs single QWR and periodic QWR arrays have been fabricated using the fast growing sidewalls of [0 1 1̄] oriented shallow mesa stripes on GaAs(3 1 1)A. The smooth convex surface profile without any faceting and the strong carrier confinement with lateral confinement energies as high as 220 meV result in unique electronic properties persistent up to room temperature. The addition of hydrogen induces step bunching perpendicular to the [0 1 1̄] oriented shallow mesa stripes on (3 1 1) substrates, which gives rise to very uniform GaAs QD arrays. These quantum dots exhibit minimized size fluctuations which is demonstrated by the extremely narrow single PL line having a linewidth of 60 μeV at 8 K (limited by the spectrometer resolution). Finally, and this is very important for most

device application, our new approach presented here allows the precise positioning of the QWRs and QDs on the wafer.

Note added in proof

In answering one of the referees on the PL linewidth of our GaAs quantum dots we are really surprised which way this discussion now takes. I like to remind this referee that the narrowest PL linewidth of single GaAs quantum *wells* that we have measured quite some time ago (J. Cryst. Growth 111, p. 344) was less than 200 μeV for a 27.3 nm wide well. Since in quantum dots the density of states is a singularity, the homogeneous excitonic PL feature should, in principle, exhibit no broadening, which, however, is smeared out for dot arrays with statistic size fluctuations. That was one of the motivations to undertake all the expensive efforts for the fabrication of quantum dot arrays of uniform size needed for applications.

The whole discussion of the uniformity of our GaAs quantum dots is based on the high-resolution spectra displayed in our review article in addition to the overview spectra showing the dot emission together with the emission from the connecting quantum wells published now in Nature, 392, p. 56. Even in the high-resolution spectra it is seen, that the emission is composed from the contribution of many dots. There are weak shoulders and a background, however, in a very narrow spectral range – a single dot or defect would have no background and shoulders. The spectral narrowing and the dominance of one single peak for the dot array compared to the quantum well emission is shown in Fig. 4a of above reference at three different positions at the sidewall. At positions which reveal some residual size fluctuations a statistics in a strongly reduced spectral range is visible demonstrating the contribution of many dots which can merge in the resolution limited sharp peak. The emission from the quantum well varies randomly at different positions at the sidewall (due to the random fluctuations along the step bunches) while the dot emission is centered at a certain energy, additionally proving the size distribution.

The emission from wires grown by conventional MBE is around 10 meV wide due to random interface fluctuations along the wires and is thus observed to

split into sharp lines at random energy positions. It is the introduction of the quasi-periodic step bunching to translate the random interface fluctuations into the quasi-periodic thickness modulations which thus generates the uniform dot array characterized by spectral narrowing of the PL. Important is that the emission from the dots dominates the spectra in intensity and does appear almost at the center of the envelope of the quantum-dot PL when the excitation area is increased above several 10 µm. This excludes the emission to be due to a single cluster or defect. Since, however, the real line width of the dots is not known and is not resolved by the spectrometer, it is not possible to discuss how the sharp lines superimpose when their energy distribution is improved in the observed way, i.e., what the resulting lineshape and peak intensity should be.

Acknowledgements

References

[1] For a recent review see, K. Eberl, P.M. Petroff, P. Demeester (Eds.), Low Dimensional Structures Prepared by Epitaxial Growth or Regrowth on Patterned Substrates, Kluwer, Dordrecht, 1995 NATO Advanced Science Institute Series E, vol. E, p. 298.

[2] H. Sakaki, in: F. Kuchar, H. Heinrich, G. Bauer (Eds.), Localization and Confinement of Electrons in Semiconductors, Springer, Berlin, 1990, Springer Ser. Solid-State Sci. 97 (1990) 2.

[3] R. Nötzel, J. Menniger, M. Ramsteiner, H.P. Schönherr, L. Däweritz, K.H. Ploog, Appl. Phys. Lett. 68 (1996) 1132; Jpn. J. Appl. Phys. 35 (1996) L297.

[4] H.P. Schönherr, J. Fricke, Z.C. Niu, K.J. Friedland, R. Nötzel, K.H. Ploog, Appl. Phys. Lett. 72 (1998) 566.

[5] R. Nötzel, Z.C. Niu, M. Ramsteiner, H.P. Schönherr, A. Trampert, L. Däweritz, K.H. Ploog, Nature 392 (1998) 56.

[6] W.T. Tsang, A.Y. Cho, Appl. Phys. Lett. 30 (1977) 293.

[7] E. Kapon, D.M. Hwang, R. Bhat, Phys. Rev. Lett. 63 (1989) 430.

[8] A. Madhukar, Thin Solid Films 231 (1993) 8.

[9] M. Walther, T. Röhr, G. Böhm, G. Tränkle, G. Weimann, J. Cryst. Growth 127 (1993) 1045.

[10] R. Nötzel, N. Ledentsov, L. Däweritz, M. Hohenstein, K.H. Ploog, Phys. Rev. Lett. 67 (1991) 3812.

[11] L. Goldstein, F. Glas, J.Y. Marzin, M.N. Charasse, G. Le Roux, Appl. Phys. Lett. 47 (1985) 1099.

[12] S. Guha, A. Madhukar, K.C. Rajkumar, Appl. Phys. Lett. 57 (1990) 2110.

[13] D. Leonard, M. Krishnamurty, C.M. Reaves, S.P. DenBaors, P.M. Petroff, Appl. Phys. Lett. 63 (1993) 3203.

[14] J.M. Moison, F. Houzay, F. Barthe, L. Leprince, E. Andre, O. Vatel, Appl. Phys. Lett. 64 (1994) 196.

[15] R.L. Schwoebel, J. Appl. Phys. 40 (1969) 614; Y. Tokura, H. Saito, T. Fukui, J. Crystal Growth 94 (1989) 1373.

[16] M. Sato, H. Maehashi, H. Asahi, S. Hasegawa, H. Nakashima, Superlattices Microstruct. 7 (1990) 279.

[17] T. Fukui, H. Saito, Jpn. J. Appl. Phys. 29 (1990) L483 and L731.

[18] R. Nötzel, L. Däweritz, K.H. Ploog, Phys. Rev. B 46 (1992) 4736.

[19] Y. Okada, T. Fujita, M. Kawabe, Appl. Phys. Lett. 67 (1995) 676.

[20] Y. Morishita, Y. Nomura, S. Goto, Y. Katayama, Appl. Phys. Lett. 67 (1995) 2500.

[21] R. Nötzel, M. Ramsteiner, J. Menniger, A. Trampert, H.P. Schönherr, L. Däweritz, K.H. Ploog, Proc. 24th Int. Symp. Compound Semicond., San Diego, 8–11 September 1997, Eds. M. Melloch, M.A. Reed, IEEE 97 TH 8272, Inst. Phys. Publ., Bristol, 1998, p. 65.

[22] R. Nötzel, M. Ramsteiner, J. Menniger, A. Trampert, H.P. Schönherr, L. Däweritz, K.H. Ploog, J. Appl. Phys. 80 (1996) 4108.

[23] R. Nötzel, J. Menniger, M. Ramsteiner, A. Trampert, H.P. Schönherr, L. Däweritz, K.H. Ploog, J. Crystal Growth 175/176 (1997) 1114.

[24] A. Richter, G. Beheme, M. Süptitz, Ch. Lienau, T. Elsässer, M. Ramsteiner, R. Nötzel, K.H. Ploog, Phys. Rev. Lett. 79 (1997) 2145.

[25] R. Nötzel, U. Jahn, Z.C. Niu, A. Trampert, J. Fricke, H.P. Schönherr, T. Kruth, D. Heitmann, L. Däweritz, K.H. Ploog, Appl. Phys. Lett. 72 (1998) 2002.

[26] R. Nötzel, M. Ramsteiner, Z.C. Niu, H.P. Schönherr, L. Däweritz, K.H. Ploog, Appl. Phys. Lett. 70 (1997) 1578.

[27] K.J. Friedland, H.-P. Schönherr, R. Nötzel, K.H. Ploog, Phys. Rev. B (1998), submitted.

Atomically precise quantum dots fabricated by two-fold cleaved edge overgrowth: from artificial atoms to molecules

W. Wegscheider *, G. Schedelbeck, M. Bichler, G. Abstreiter

Walter Schottky Institut, Technische Universität München, Am Coulombwall, D-85748 Garching, Germany

Abstract

The optical properties of different quantum dot structures consisting of individual dots, pairs of coupled dots as well as of linear arrays of dots have been studied experimentally. The $7 \times 7 \times 7 \, nm^3$ size GaAs quantum dots form at the intersection of three orthogonal quantum wells fabricated by two-fold application of the cleaved edge overgrowth method. The high degree of control over shape, composition and position of the dots, achievable by this technique, allows a detailed investigation of the influence of coupling between almost identical zero-dimensional objects. In this way an "artifical molecule", characterized by the existence of bonding and antibonding states, can be assembled from two of such "artificial atoms". The coupling strength between the "artificial atoms" directly adjusts the "interatomic" distance and is reflected in the energetic separation of the bonding and antibonding levels and the line widths of the corresponding interband transitions observed by means of microscopic photoluminescence spectroscopy. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 73.20.Dx; 73.20.At; 78.55.Cr

Keywords: Quantum dots; Excitons; Coupling

The optical absorption and emission spectra of atoms exhibit sharp lines which directly reflect the discrete energy level scheme of electrons in the different shells. In contrast, as a result of the periodic potential, electrons in crystalline solids reside within bands which are separated by forbidden energy gaps. Consequently, the optical properties of semiconductors are usually dominated by wide absorption bands, resulting from the smooth and continuous density of states profiles. Semiconductor quantum wires (QWRs) and dots (QDs) with quantum confinement of charge carriers to one (1D) or zero dimensions (0D), respectively, have attracted extensive research activities recently, as they possess unique properties which arise due to sharp peaks in their density of states. While charge carriers in 1D systems can still move freely along the QWR axis and, thus, still have a continuous energy level spectrum, the complete confinement present in 0D systems results in discrete δ-function-like density of states peaks. The successful realization of different types of QDs in semiconductors over the last five years [1–18] led to the discovery of effects of a "shell-like" energy staircase

* Corresponding author. Fax: +49 89 3206620; e-mail: wegscheider@wsi.tu-muenchen.de.

for single-electron charging [18] or extremely narrow homogeneous line widths for optical transitions [9] which resemble in many ways the characteristic features of atoms. This analogy manifests itself in the designation "artificial atom" for such 0D objects.

A diverse range of technologies has been implemented to fabricate these "artificial atoms" from molecular beam epitaxially prepared layers which often serve as a starting material since quantum confinement along the growth direction can be readily achieved. Lateral patterning of two-dimensional (2D) quantum well (QW) layers, however, severely affects the optical quality of the structures and typical photoluminescence (PL) line widths of single QDs prepared in this way exceed 0.5 meV [1–3]. In addition, substantial fluctuations in composition and size from one QD to another cannot be avoided. The realization of QDs confining excitons within about 10 nm, the equivalent of the hydrogen-like Bohr orbit for excitons in GaAs, whose emission is governed by narrow "atom-like" lines was up to now limited to natural formation techniques. These rely on monolayer thickness fluctuations in narrow QWs [6–9] or on a thermodynamically driven instability during strained layer epitaxy which leads to spontaneous island formation [10–13]. These natural and self-assembled QDs do not suffer from crystal imperfections or free surfaces which are inevitably accompanied by a high number of nonradiative recombination centers. However, as a consequence of the inherent randomness in the formation process highly local spectroscopic techniques are essential in order to extract the properties of individual objects from an ensemble of such QDs which differ in size and shape and whose position is not under precise experimental control.

Molecular beam epitaxial (MBE) growth on the cleavage face of a previously prepared GaAs/AlGaAs multilayer, a technique pioneered by Pfeiffer et al. [19], has recently attracted increasing attention because of its capability for the fabrication of high-quality QWR structures suitable for optical and transport investigations [20–25]. The high perfection of these structures led to the demonstration of enhancement of exciton binding [21,22] and lasing from excitons in 1D [21]. As predicted in Ref. [26] and theoretically modelled recently [27], cleaved edge overgrowth (CEO) can also be employed to produce QDs, in which charge carriers are confined in all three

dimensions due to atomically precise QW potentials and, therefore, are highly uniform in size and shape. Here we report on the optical properties of different structures consisting of individual QDs, pairs of coupled QDs as well as of linear arrays of QDs which were obtained by two-fold application of the CEO method [28,29]. The complex sample designs make explicit use of atomic control of both size and position of two-fold CEO QDs. The latter seem to be ideally suited to perform fundamental studies on a novel class of well isolated or controllably coupled 0D objects.

The existence of 0D quantum mechanical bound states at the right angle intersection of three QWs is visualized in Fig. 1a. Similar to the formation of a 1D state at the T-shaped intersection of two QWs "expansion" of the electron and hole wavefunctions into the third QW due to confinement relaxation leads to a lowering of the corresponding ground state energy. Carriers which can move freely in two directions within the QWs and along one direction in the QWRs are completely localized in such QDs. If the layer sequence grown along the first growth direction consists of multiple QWs, the same number of QDs, connected via a single QWR result. The two different QD structures investigated in detail here are sketched in Fig. 1b and c. The QD superlattice structure of Fig. 1b consists of a 22-period $GaAs/Al_{0.35}Ga_{0.65}As$ multiple QW (MQW_1) structure with well and barrier thicknesses of 7 and 30 nm, respectively, which was two times in situ cleaved along the (1 1 0) and ($\bar{1}$ 1 0) cleavage planes and subsequently overgrown with two 7 nm wide GaAs single QWs (QW_2, QW_3) and $Al_{0.35}Ga_{0.65}As$ barriers. Details of the three step growth procedure are similar to those described in Ref. [21]. As can be seen in Fig. 1a three types of QWRs (QWR_{12}, QWR_{13}, QWR_{23}), resulting from intersections of MQW_1 with QW_2, MQW_1 with QW_3 and QW_2 with QW_3 form. At a length of about 800 nm 22 equally spaced and nominally identical QDs, each of which is constituted by the junction of three QWs, are located along QWR_{23} about 930 nm below the (0 0 1) sample surface. The distances of the QD array from the (1 1 0) and ($\bar{1}$ 1 0) sample surfaces are approx. 340 nm and 140 nm, respectively. In order to study the transition from an isolated QD ("artificial atom") to a system of two QDs with adjustable coupling strength ("artificial molecule"), a regime that is inaccessible to experiments on real atoms, the QD
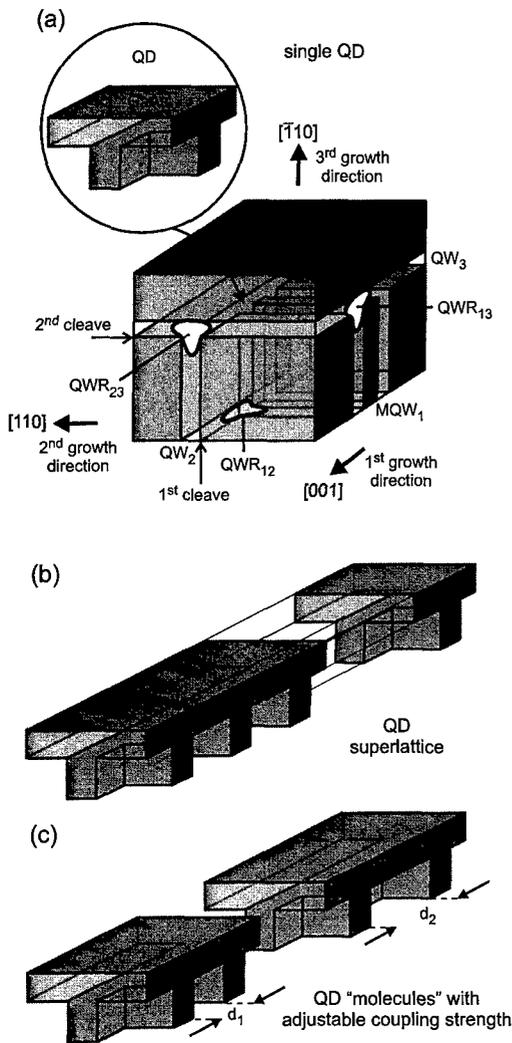
Fig. 1. (a) Schematic illustration of the quantum dot (QD) structure (not to scale) obtained after three growth steps separated by two in situ cleaves. The junction of three quantum wells and wires, at which a QD forms, is shown in the magnified part of the figure. The T-shaped contours are lines of constant probability for electrons confined in the quantum wires. The two sample structures investigated here consist of a QD superlattice (b) and a series of three coupled QDs with different coupling strengths as well as an individual QD (c).

structure of Fig. 1c has been prepared. The layer sequence of the first growth step consists in this case of a single QW followed by three pairs of QWs embedded in $Al_{0.35}Ga_{0.65}As$ barrier material. All GaAs QW layers are again 7 nm wide and are, similar to the

QD superlattice structure described previously, overgrown after cleavage along the two $\langle 1\,1\,0 \rangle$ directions by another 7 nm wide QW and an $Al_{0.35}Ga_{0.65}As$ barrier. The barriers separating adjacent QWs and also QDs are $d = 15$, 30 nm and 60 nm. The spatial distance between the QDs adjusts the coupling strength of the zero-dimensional QD levels. The center-to-center distance from the single QD to the first and the following coupled QDs was chosen to be 800 nm. The relevant distances of the QDs from the $(1\,1\,0)$, $(\bar{1}\,1\,0)$ and $(0\,0\,1)$ surfaces are identical to those given for the QD superlattice structure.

The CEO QD samples were investigated using microphotoluminescence (μPL) and μPL excitation (μPLE) spectroscopy combining high spatial and spectral resolution. Excitation was performed by means of a tunable dye laser which was focused through a microscope objective lens (numerical aperture $NA = 0.75$) onto the sample. The power density was kept below $50\,W/cm^2$. Under such conditions less than one exciton at a time is expected to be in the QDs. The samples were mounted on the cold-finger of an He cryostat and studied at nominally 5 K. The emitted light was collected by the same objective lens and directed to a confocal imaging system, which defined a nearly diffraction limited detection range of about 800 nm. Finally, the PL signal was dispersed using a triple grating Raman spectrometer (spectral resolution $\approx 40\,\mu eV$) and detected with a liquid nitrogen cooled charge-coupled device camera. We would like to point out that in spite of diffraction limitation, spectroscopy of single QDs is possible using conventional optical microscopy since the CEO method allows the fabrication of QD structures with adjustable QD separations larger than the spatial resolution of the instrument. This is the case for the sample, designed to monitor subsequent phases in the formation of "artificial molecules" out of "artificial atoms", in which all relevant QD objects, i.e. the single QD and the three coupled QD systems, can be probed individually without the use of non-diffraction limited probes like near-field optical spectroscopy [7,9].

Fig. 2 shows typical μPL spectra of the two samples recorded from the $(0\,0\,1)$ and $(\bar{1}\,1\,0)$ surfaces as indicated in the inset. The spectrum of the superlattice QD sample (lower curve) was taken at a position in the vicinity of the QDs. As a result of the excitation geometry given in the inset of Fig. 2 all 22 QDs are
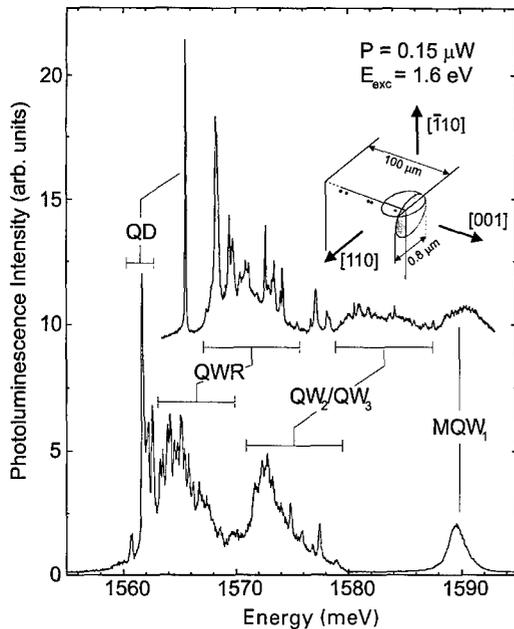
Fig. 2. µPL spectra of the samples containing the 22-period QD superlattice (lower curve) and the individual and coupled QDs (upper curve). The spectra were obtained by excitation and detection from the (0 0 1) (lower curve) and ($\bar{1}$ 1 0) (upper curve) sample surfaces, respectively, as indicated by the shaded circles in the inset (not to scale) which also schematically depicts the positions of the QDs (black dots). The two spectra are vertically offset against each other for clarity.

excited simultaneously and contribute to the signal in the spectral range of 1560–1563 meV. It should be noted at this point that due to the large excitation area, which is unavoidable in diffraction limited spectroscopy, and exciton diffusion over a length of about 0.5 µm [30], the carriers, most of which are generated in the QW regions, recombine in the quantum dots, wires and wells. The relatively strong response from the QDs in spite of their extremely small volume results from efficient collection of excitons from the adjacent QWR and QW regions. Comparison with spectra taken at positions excluding excitation of the dots or wires identifies the PL response observed at 1589 meV and in the range of 1570–1580 meV as emission of MQW_1 and the two wells QW_2/QW_3. The PL energies of QW_2 and QW_3 of 1574 and 1576 meV determined independently in this way almost coincide, but are, as a result of the large heavy hole mass for $\langle 1\ 1\ 0 \rangle$ type confinement di-

rections, well below the MQW_1 response. Emission of the three types of QWRs present in the QD structure is detected around 1565 meV. Again, the PL maxima of QWR_{12} and QWR_{13}, determined by moving the detection window away from the QD region coincide within about 1 meV. With the high spatial resolution used for these experiments we find that the QWR PL and partly that of QW_2 and QW_3 starts to decompose into sharp lines as a result of exciton localization. This effect becomes even more pronounced in the µPL spectrum of the second sample which has been taken at a position where only one QD can contribute to the signal (upper curve). Consequently, the number of localized 1D states within the excitation and collection volume is also reduced substantially and individual sharp lines start to emerge from the broad QWR background. The QD response, however, is in this case characterized by a single sharp line at an energy of 1565.9 meV, which indicates that the corresponding 0D exciton ground state is ≈ 5 meV deep with respect to the 1D QWR exciton states centered around an energy of 1571 meV. This value agrees well with the theoretical prediction for our QD structure of 6 meV [27]. Comparison of the two spectra in Fig. 2 also reveals that all PL transitions associated with the QDs, QWRs and QW_2/QW_3 appear at lower energies in the QD superlattice sample. This result, which originates from slightly larger widths of QW_2 and QW_3 in this sample, is typical for thickness variations encountered in different CEO samples. The unambigous assignment of the peaks to the different low-dimensional structures in our samples follows from a detailed three-dimensional µPL mapping which is presented in extracts below.

Fig. 3a shows µPL spectra taken at different positions from the ($\bar{1}$ 1 0) surface of the QD superlattice sample together with grey-scale intensity plots (pixel size $200 \times 200\,nm^2$) at selected PL energies. For the position chosen in the upper panel of this figure, marked by a white cross in the grey scale image, emission from MQW_1 and QWR_{13} can be detected. The grey-scale intensity plot recorded with the detection system kept fixed at the MQW_1 PL line, indicated by an arrow in the spectrum, clearly reveals quenching of the multiple quantum well emission in the vicinity of the QDs due to exciton diffusion into wires and dots. The fact that the MQW_1 PL signal is not accordingly suppressed at positions away from the QDs but still on top of QWR_{13} is a result of the finite

excitation and collection depth of our probe which exceeds the exciton diffusion length. In accordance with the detection position used for the spectrum displayed in the middle panel of Fig. 3a, only signals originating from recombination within the two single QWs (QW$_2$, QW$_3$) and the single wire (QWR$_{23}$) which is formed by these two wells can be detected. Again, bleaching

(a)

of the QW emission in the vicinity of the multiple quantum wires and dots is observed, as can be seen in the corresponding grey scale plot. Exclusively at the intersection of the wells and wires, i.e. on top of the QD array, the characteristic sharp peaks, interpreted before as PL originating from the QDs, emerge (lower panel of Fig. 3a). Although the relative intensities of these peaks have changed with respect to the spectrum shown in Fig. 2 (lower PL curve), their energetic positions exactly coincide. The intensity decrease of the QD emission in all directions is limited by the experimental resolution, as demonstrated by the grey scale image taken at one of the sharp resonances. Similar intensity distributions are recorded for the other sharp peaks. They are, however, along the extension of the QD array slightly offset against each other. This is shown in Fig. 3b, where the intensities of two peaks measured during a high-resolution linescan, marked in the inset, are plotted. The displacement of the two curves of about 190 nm is well within the length of the QD array of 800 nm. Similar offsets are observed for the other QD transitions. This result implies that the different QD resonances can be attributed to exciton recombination in individual spatially separated QDs of the array. The energy variations might be explained by slight differences in the widths of the participating QWs at the dot positions.

Near resonant μPL and μPL excitation (μPLE) spectra of this sample are collected in Fig. 4. When the excitation is tuned to the energy labeled (4) in the upper panel of Fig. 4, the PL signal of the QDs is much clearer resolved and almost free of background.
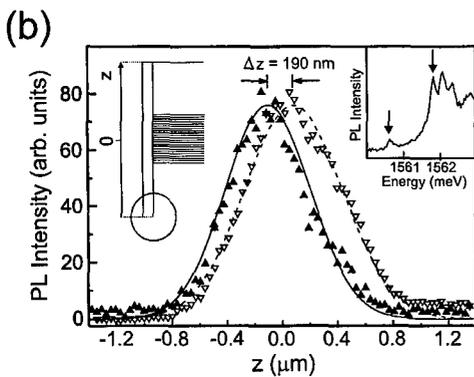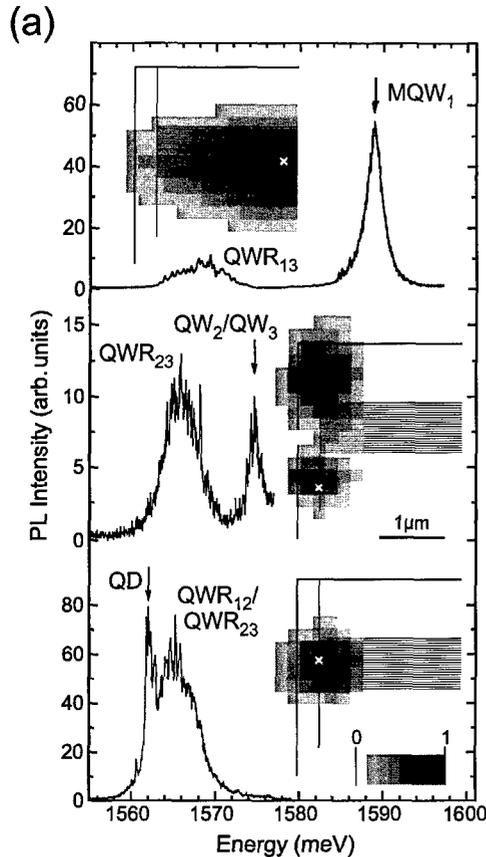
Fig. 3. (a) μPL spectra ($E_{exc} = 1.6$ eV, $P = 0.15$ μW) and grey scale intensity plots recorded from the ($\bar{1}$ 1 0) surface of the QD superlattice sample. The positions on the sample where the spectra were taken are indicated by the white crosses in the grey scale images which are superimposed on a schematic view of the sample. The arrows define the spectral position of the corresponding grey scale plot. (b) High-resolution PL intensity linescans of two, spectrally well separated quantum dot peaks at 1560.7 meV (filled symbols) and 1561.9 meV (hollow symbols) marked by the arrows in the magnified portion of the μPL spectrum (right inset). The orientation of the linescan is given in the left inset by the circle, indicating the detection position, which has been moved parallel to the z-axis along QWR$_{23}$. The solid curves are a guide to the eye.
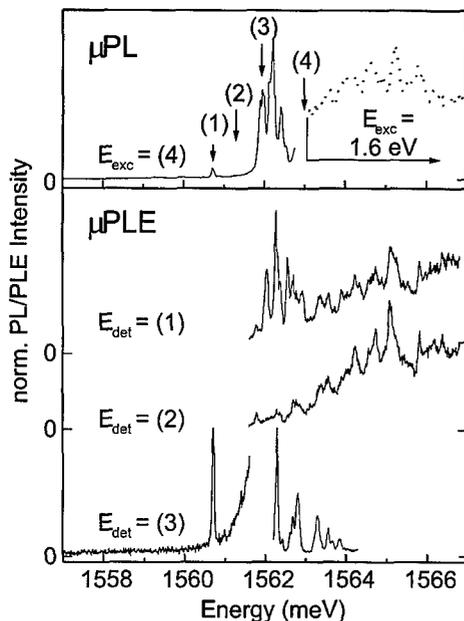
Fig. 4. Microscopic PL and PLE spectra ($P = 0.4\,\mu\text{W}$) taken from the ($\bar{1}\,1\,0$) surface of the QD superlattice sample at a position centered on top of the quantum dot array. The arrows mark the energetic positions used for excitation (μPL) or detection (μPLE).

The μPLE spectrum displayed in the lower panel of Fig. 4 with the detection at the energy labeled (1), the lowest QD level, exhibits sharp resonances exactly at the QD PL energies. For energies above 1563 meV the line shape of the μPLE signal coincides with the QWR PL finestructure (dotted line). If the detection energy is slightly detuned with respect to the QD levels, for instance to the energy labeled (2), the sharp QD resonances are completely suppressed. In contrast, the QWR μPLE signal remains unchanged. For the lowest μPLE spectrum displayed in Fig. 4 the detection was moved to the energy labeled (3). Again, sharp resonances appear exactly at the energies of the other dot levels. Surprisingly, even for excitation energies smaller than the detection energy it is possible to observe such a sharp resonance at the lowest dot level. The pronounced excitation of individual QDs under resonant pumping of QD states which are slightly above or below in energy and localized at spatially separated QDs has to be interpreted in terms of coupling of the 0D states. Such a coupling seems to be possible although the QDs are separated

by 30 nm wide barriers, because the effective barrier height, represented by QWR_{23} is only ≈ 5 meV. This has been confirmed by a one-dimensional Kronig–Penney-type superlattice calculation, that resulted in a ground state bandwidth of 1.7 meV. Assuming a temperature slightly higher than the nominal 5 K, the thermal energy of about 0.7 meV is comparable to the energy difference from the lowest to the higher QD levels. The observed anti-Stokes luminescence is then explained by thermally activated tunneling of excitons from one QD to another.

Deeper insight into the effects of quantum mechanical coupling between 0D exciton states in two QDs can be gained by a detailed analysis of the sample which contains a single QD as well as a series of pairs of QDs with differently adjusted coupling strength. Fig. 5 shows the spectral image of a linescan along QWR_{23} in this sample on an enlarged energy scale. In additon the PL intensities integrated over the spectral regions of maxima in this image are plotted as a function of the position along the linescan in the bottom portion of this figure. The image is dominated at the right-hand side by the single line, which is also visible in the upper PL spectrum of Fig. 2, and which originates from excitons recombining within the single QD. Moving the excitation and detection window to the coupled QDs, a series of double peaks appears. The spatial positions of all peaks exactly coincide with the designed positions of our individual and coupled QDs as can be seen by the integrated PL intensity plots in which the nominal QD positions are marked by arrows. The spatial width of these peaks is comparable to that observed for the different transitions in the QD super-lattice sample (Fig. 3b) and is limited by the spatial resolution of the μPL setup. This clear correlation of the sharp lines with position directly proves that they originate from excitonic recombination in the CEO QDs. The energetic spread of the peaks in the range of 1 to 2 meV can be well described by monolayer thickness fluctuations in the overgrown QWs. However, the wave functions of the coupled QDs probes only a very short part of the constituting QWs with respect to the length scale of monolayer fluctuations and coupling between almost identical objects can be studied.

The appearance of two energetically distinct peaks from coupled QDs is clear evidence that their excitonic levels are coupled. This scenario of an exciton
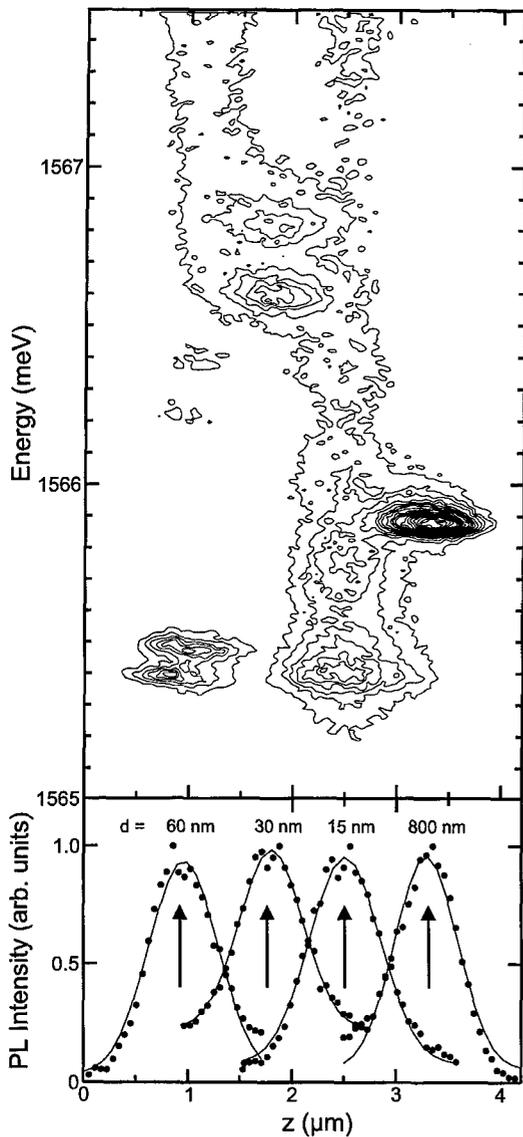
Fig. 5. Spectral image resulting from a high-resolution μ-PL linescan (E_{exc} = 1588 meV) along QWR$_{23}$, which connects the single QD with the three different coupled QDs, taken from the ($\bar{1}$ 1 0) sample surface (upper panel). The contour lines in this image represent regions of constant PL intensity of 10%, 15%, 20%, ..., 95% of the maximal intensity. In the lower panel of this figure normalized PL intensities integrated over the spectral regions of maxima in the image, centered at about 1565.42 meV (d = 60 nm), 1566.63 meV (d = 30 nm), 1565.70 meV (d = 15 nm) and 1565.89 meV (d = 800 nm, single QD) are plotted as a function of the position along the linescan. The nominal positions of the different QD structures are indicated by the arrows in this plot. The solid curves are just guides to the eye.
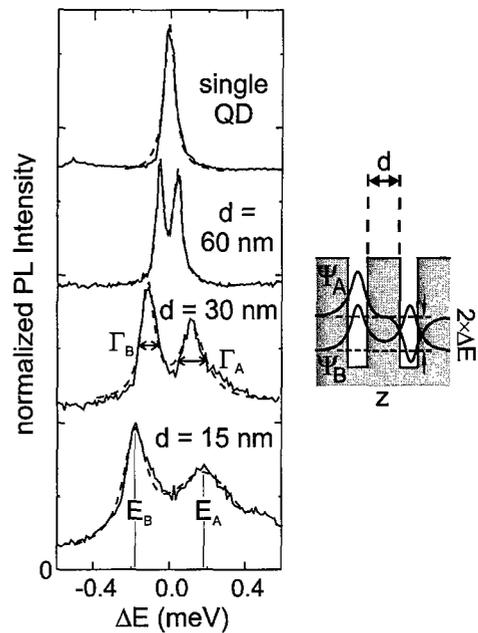
Fig. 6. Selected normalized μPL spectra from the linescan of Fig. 5 of the single and the coupled QDs (at the positions marked by the arrows), plotted on a relative energy scale and offset vertically against each other for clarity. The one-dimensional potential profile for excitons in the coupled systems, the resulting bonding (ψ_B) and antibonding wave functions (ψ_A), and the energetic splitting $2 \times \Delta E$ of such a two-level system are displayed on the right-hand side of this figure.

bound to two QDs can be compared to the synthesis of a positively charged hydrogen molecule H_2^+ from an electron and two protons, a system that can be described in the framework of elementary quantum mechanics by the formation of bonding and antibonding states. The influence of coupling on the excitonic levels, in the order of 100 μeV, is, however, small compared to the binding energy of an exciton confined to a CEO QD of more than 10 meV. Therefore, the exciton is considered as one particle whose internal structure is not affected by coupling. In analogy to the two coupled electronic levels of the H_2^+ molecule, the wave function ψ of our system of two coupled excitonic 0D levels can be approximated by the sum and difference of the wave functions of individual 0D states. These wave functions, ψ_B and ψ_A, correspond to the bonding and antibonding states as visualized in the schematic drawing on the right-hand side of Fig. 6. From first-order perturbation theory, the energy levels of the

coupled system can be written as $E_{B/A} = E_0 \mp \Delta E$, where E_0 is the energy level of an individual QD and ΔE is a matrix element that describes the energetic splitting $2 \times \Delta E$ caused by coupling. Following this argument, we assign the two energetically distinct PL lines of each coupled QD to emission from bonding and antibonding energy levels, which are separated by $2 \times \Delta E$. The observed systematic dependence of the energetic splitting on the distance between the coupled dots (Fig. 6) confirms our interpretation in terms of bonding and antibonding states. The enhancement of the energetic splitting with decreasing distance between the coupled QDs directly reflects the increasing coupling strength of the 0D levels.

In agreement with results obtained from naturally formed QDs [9] the spectra shown in Fig. 6 are well fitted with Lorentzian functions (dashed curves). This implies that the PL line widths corresponding to the bonding and antibonding states are life-time broadened. These line widths increase systematically with increasing coupling strength. The broadening is especially pronounced for emission from the antibonding level. This decrease of the exciton lifetime for the antibonding state with increasing level separation might be due to acoustical phonon scattering from the upper into the lower state as suggested recently [31]. In such a simple two-level picture one would expect a cubic dependence of the line width of the upper level on the energy level separation, which is also observed experimentally.

In conclusion, we have shown that the CEO method has evolved to a state where two-fold cleaved and overgrown structures can be prepared reliably. The second overgrowth step adds another degree of flexibility in the design of the structures. The excellent optical quality, manifested in extremely narrow emission lines, as well as the high degree of homogeneity of QD structures prepared in this way permit a variety of experiments on adjustable coherently coupled nanoscale structures inaccessible by other means.

References

[1] K. Brunner, U. Bockelmann, G. Abstreiter, M. Walther, G. Böhm, G. Tränkle, G. Weimann, Phys. Rev. Lett. 69 (1992) 3216.

[2] Y. Nagamune, H. Watabe, M. Nishioka, Y. Arakawa, Appl. Phys. Lett. 67 (1995) 3257.

[3] M. Bayer, A. Schmidt, A. Forchel, F. Faller, T.L. Reinecke, P.A. Knipp, A.A. Dremin, V.D. Kulakovskii, Phys. Rev. Lett. 74 (1995) 3439.

[4] N.C. van der Vaart, S.F. Goodijn, Y.V. Nazarov, C.J.P.M. Harmans, J.E. Mooij, L.W. Moolenkamp, C.T. Foxon, Phys. Rev. Lett. 74 (1995) 4702.

[5] F.R. Waugh, M.J. Berry, D.J. Mar, R.M. Westervelt, K.L. Campman, A.C. Gossard, Phys. Rev. Lett. 75 (1995) 705.

[6] A. Zrenner, L.V. Butov, M. Hagn, G. Abstreiter, G. Böhm, G. Weimann, Phys. Rev. Lett. 72 (1994) 3382.

[7] H.F. Hess, E. Betzig, T.D. Harris, L.N. Pfeiffer, K.W. West, Science 264 (1994) 1740.

[8] K. Brunner, G. Abstreiter, G. Böhm, G. Tränkle, G. Weimann, Appl. Phys. Lett. 64 (1994) 3320; Phys. Rev. Lett. 73 (1994) 1138.

[9] D. Gammon, E.S. Snow, B.V. Shanabrook, D.S. Kratzer, D. Park, Phys. Rev. Lett. 76 (1996) 3005; Science 273 (1996) 87.

[10] J.-Y. Marzin, J.-M. Gérard, A. Izraël, D. Barrier, G. Bastard, Phys. Rev. Lett. 73 (1994) 716.

[11] R. Leon, P.M. Petroff, D. Leonard, S. Safard, Science 267 (1995) 1966.

[12] M. Grundmann, J. Christen, N.N. Ledentsov, J. Böhrer, D. Bimberg, S.S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V.M. Ustínov, A. Yu. Egorov, A.E. Zhukov, P. S. Kopév, Zh. I. Alferov, Phys. Rev. Lett. 74 (1995) 4043.

[13] A. Kurtenbach, K. Eberl, T. Shitara, Appl. Phys. Lett. 66 (1995) 361.

[14] M.A. Kastner, Phys. Today 46 (1993) 24.

[15] O. Klein, C. de Chamon, D. Tang, D.M. Abusch-Magder, U. Meirav, X.-G. Wen, M.A. Kastner, S.J. Wind, Phys. Rev. Lett. 74 (1995) 785.

[16] R.C. Ashoori, H.L. Stormer, J.S. Weiner, L.N. Pfeiffer, K.W. Baldwin, K.W. West, Phys. Rev. Lett. 71 (1993) 613.

[17] R.C. Ashoori, Nature 379 (1996) 413.

[18] S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, L.P. Kouwenhoven, Phys. Rev. Lett. 77 (1996) 3613.

[19] L. Pfeiffer, K.W. West, H.L. Störmer, J.P. Eisenstein, K.W. Baldwin, D. Gershoni, J. Spector, Appl. Phys. Lett. 56 (1990) 1697.

[20] A.R. Goñi, L.N. Pfeiffer, K.W. West, A. Pinczuk, H.U. Baranger, H.L. Stormer, Appl. Phys. Lett. 61 (1992) 1956.

[21] W. Wegscheider, L.N. Pfeiffer, M.M. Dignam, A. Pinczuk, K.W. West, S.L. McCall, R. Hull, Phys. Rev. Lett. 71 (1993) 4071.

[22] T. Someya, H. Akiyama, H. Sakaki, Phys. Rev. Lett. 74 (1995) 3664.

[23] H. Gislason, W. Langbein, J.M. Hvam, Appl. Phys. Lett. 69 (1996) 3248.

[24] J. Hasen, L.N. Pfeiffer, A. Pinczuk, S. He, K.W. West, B.S. Dennis, Nature 390 (1997) 54.

[25] A. Yacoby, H.L. Stormer, N.S. Wingreen, L.N. Pfeiffer, K.W. Baldwin, K.W. West, Phys. Rev. Lett. 77 (1996) 4612.

[26] W. Wegscheider, L.N. Pfeiffer, K.W. West, in: R. Helbig (Ed.), Festkörperprobleme/Advances in Solid State Physics, vol. 35, Vieweg, Wiesbaden, 1996, p. 155.

[27] M. Grundmann, D. Bimberg, Phys. Rev. B 55 (1997) 4054.

[28] W. Wegscheider, G. Schedelbeck, G. Abstreiter, M. Rother, M. Bichler, Phys. Rev. Lett. 79 (1997) 1917.

[29] G. Schedelbeck, W. Wegscheider, M. Bichler, G. Abstreiter, Science 278 (1997) 1792.

[30] R.D. Grober, T.D. Harris, J.K. Trautman, E. Betzig, W. Wegscheider, L. Pfeiffer, K. West, Appl. Phys. Lett. 64 (1993) 1421.

[31] P. Platzman, personal communication.

Electronic states in quantum dot atoms and molecules

S. Tarucha [a,*], T. Honda [a], D.G. Austing [a], Y. Tokura [a], K. Muraki [a], T.H. Oosterkamp [b], J.W. Janssen [b], L.P. Kouwenhoven [b]

[a] NTT Basic Research Laboratories, 3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-0124, Japan
[b] Department of Applied Physics and DIMES, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands

Abstract

We study electronic states in disk-shaped semiconductor artificial atoms and molecules containing just a few electrons. The few-electron ground states in the artificial atom show atomic-like properties such as a shell structure and obey Hund's rule. A magnetic field induces transitions in the ground states, which are identified as crossings between single particle states, singlet–triplet transitions and spin polarization. These properties are discussed in conjunction with exact calculation in which the effect of finite thickness of the disk is taken into account. An artificial molecule is made from vertically coupling two disk-shaped dots. When the two dots are quantum mechanically strongly coupled, the few-electron ground states are de-localized throughout the system and the electronic properties resemble those of a single artificial atom. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 73.20.Dx; 72.20.My; 73.40.Gk

Keywords: Artificial atom; Artificial molecule; Quantum dot; Excitation spectroscopy; Coulomb oscillations

1. Introduction

Semiconductor quantum dots are often referred to as artificial atoms since their electronic properties, for example the ionization energy and discrete excitation spectrum, resemble those of real atoms [1,2]. We have recently fabricated a circular disk-shaped quantum dot, and observed atomic-like properties by measuring Coulomb oscillations [3]. Electrons bound to a nuclear potential experience sufficiently strong quantum mechanical confinement and mutual Coulomb in-

teractions that they are well arranged in ordered states. This leads to the ordering of atoms in the periodic table. The ionization energy has large principal maxima for atomic numbers 2, 10, 18, ... when certain orbitals are completely filled with electrons. In addition, for the filling of electrons in similar orbitals Hund's rule favours parallel spins until the set of orbitals is half-filled. This also gives rise to secondary maxima in the ionization energy. The disk-shaped quantum dot we can fabricate is formed in a laterally gated double barrier structure, and contains a tunable number of electrons starting from zero. Associated with the rotational symmetry of the lateral confinement, we observe a two-dimensional (2D) "shell structure" from the addi-

* Corresponding author. Tel.: +81 462 40 3445; fax: +81 462 40 4723; e-mail: tarucha@will.brl.ntt.co.jp.

tion energies, analogous to the three-dimensional (3D) shell structure for atomic ionization energies. In addition, spin effects such as a pairing of Coulomb oscillation peaks due to spin degeneracy, and modifications of the pairing in line with Hund's rule are all observed. In real atoms, electrons are so strongly trapped that their quantum mechanical properties are not accessible by means of conventional spectroscopic techniques. In contrast, the electrons in our quantum dot are bound in a relatively large region of the order of 100 nm. This allows us to use readily accessible magnetic fields not only to identify the quantum mechanical states, but also to induce transitions in the ground states which are expected but never tested in real atoms on earth [4]. In this paper, we first discuss the addition energy spectrum of the ground states at zero magnetic field, and magnetic field induced transitions in the ground states for a different number of electrons, N, in a disk-shaped dot. We employ an exact diagonalization technique incorporating many body interactions, and the effect of a finite thickness of the dot disk, to understand the magnetic field induced transitions of ground states. The effect of finite thickness weakens the Coulomb interactions relative to the quantum mechanical confinement. In our previous calculation [4] we neglected this effect, and assumed that the Coulomb interactions were weaker than those reproduced in the experiment. We show here that a good agreement with experiment is now obtained with a realistic interaction model. For the next set after artificial atoms, we outline how vertically coupled disk-shaped dots can be employed to study the filling of electrons in artificial molecules. We show that in a quantum mechanically strongly coupled double dot system the electronic states are delocalized.

2. Device fabrication and experimental set up

Fig. 1a shows a schematic diagram of the device which consists of an n-doped GaAs substrate, with undoped layers of 7.5 nm $Al_{0.22}Ga_{0.78}As$, 12 nm $In_{0.05}Ga_{0.95}As$, and 9.0 nm $Al_{0.22}Ga_{0.78}As$, and $a \approx$ 500 nm n-doped GaAs top layer. A sub-micrometer pillar (geometrical diameter, D) is fabricated using
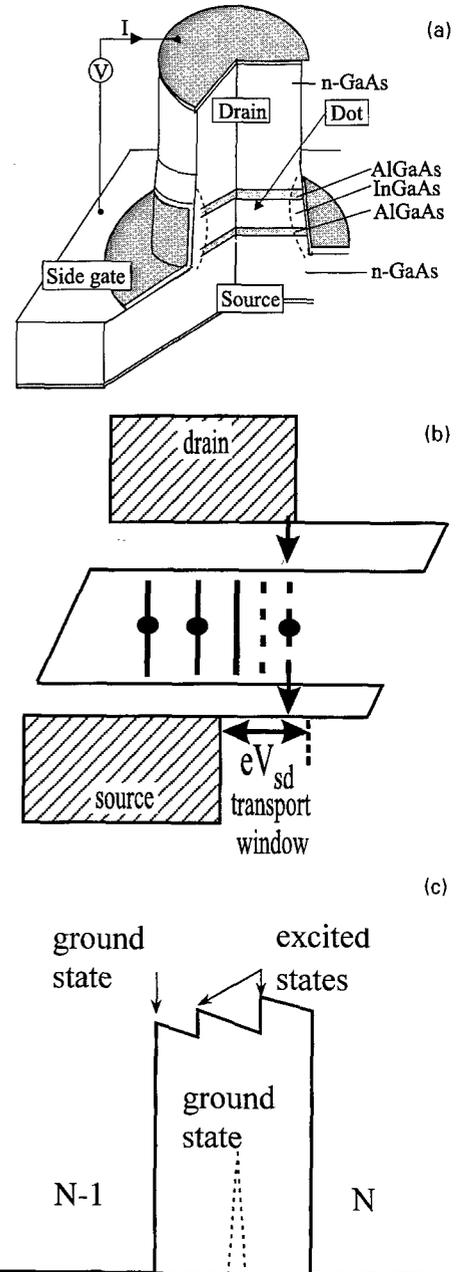
Fig. 1. (a) Schematic diagram of the device containing a disk-shaped dot. (b) Schematic energy (horizontal axis) diagram along the vertical axis of the pillar. Hatched regions are occupied electron states in source and drain contacts. For the case shown, two electrons are permanently trapped in the quantum dot. The third electron can choose to tunnel through the $N = 3$ ground state (solid line) or through one of the two excited states which lie in the transport window. (c) Schematic of tunneling current vs. gate voltage for a sufficiently small V_{sd} (dashed line) that only the ground state contributes to the current, and for a larger V_{sd} (solid line) that allows both the ground and excited states to contribute to the current. For the case of the large V_{sd}, a small jump in the current stripe identifies where an excited state enters the transport window.

electron beam lithography and etching techniques [5]. Source and drain electrical wires are connected to the top and substrate contacts. A third wire is attached to the metal that is wrapped around the pillar. This electrode is the side gate. The energy landscape along the vertical axis is shown in Fig. 1b. The $In_{0.05}Ga_{0.95}As$ layer has a disk shape. By making the gate voltage, V_g, more negative we can electrically squeeze the effective diameter of this disk from a few hundred nanometers down to zero. Application of a bias voltage, V_{sd}, between the source and drain opens a "transport window" between the Fermi energies of the source and drain for detecting both the ground and excited states in the dot (Fig. 1b). Ground and excited states lying within the transport window can contribute to the current. If the gate voltage is made more positive, then the levels in Fig. 1b shift down in energy. When V_{sd} is smaller than the energy difference between the ground and lowest excited state, only the ground state contributes to the current, I, because the electron tunneling into the excited state is blocked by the charging of the ground state. This is the usual case for the measurement of Coulomb oscillations, so we see a series of current "peaks" as a function of gate voltage corresponding to the one-by-one change of electrons in the ground states of the dot (Fig. 1c). When V_{sd} is large enough, however, both ground and excited states can be within the transport window, and contribute to the current. Electron tunneling into the excited states can occur following the electron escape from a ground state. I vs. V_g therefore becomes a series of current "stripes". Small jumps inside the stripe measure when excited states enter the transport window. Each current stripe falls off when the ground state leaves the transport window, so it has a width in energy given by eV_{sd}. We employ this technique to study the excitation spectrum. Our samples are measured while mounted in a dilution refrigerator. Due to pick-up of noise the effective electron temperature is about 100 mK.

3. Electronic states in quantum dot atoms

3.1. Ground state spectroscopy

Fig. 2a shows I vs. V_g for single quantum dot ($D = 0.5\,\mu m$) measured at a small V_{sd} of 0.15 mV so that only the ground states contribute to the current.
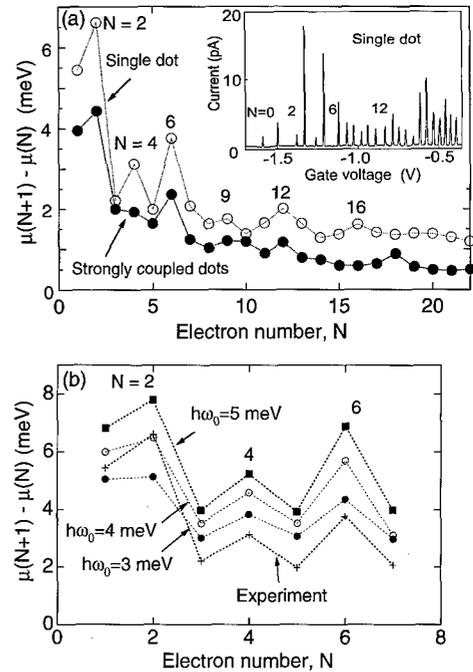
Fig. 2. (a) Change of electrochemical potential vs. electron number for a single dot ($D = 0.5\,\mu m$ open circles) and also for two strongly coupled dots ($D = 0.56\,\mu m$ solid circles; see Section 4 in the text). The inset shows the Coulomb oscillations measured for $V_{sd} = 0.15\,mV$ at $B = 0\,T$ for the single dot. (b) Change of the electrochemical potential calculated for the single dot shown in Fig. 2a. Three different $\hbar\omega_0$ values of 3, 4 and 5 meV are taken as parameters. The experimental data for the single dot in Fig. 2a is also shown by the crosses for comparison.

The current oscillations arise from the one-by-one change of electrons trapped in the dot. The absolute values of N can be identified in each zero-current region (Coulomb blockade region) between the peaks, starting from $N = 0$, because for $V_g < -1.6\,V$ the dot is empty. When N becomes smaller than 20, the oscillation period depends strongly on N. In contrast, Coulomb oscillations observed for a large dot containing more than 100 electrons look very periodic (data not shown; see Ref. [6]), as expected from classical Coulomb blockade theory. The current peak to the left of a Coulomb blockade region with N trapped electrons thus measures the N-electron ground state energy (or electrochemical potential, $\mu(N)$), of the N-electron dot). For example, the first, second and third peaks from the left measure the one, two and three electron ground state energies, respectively. The peak

spacing labeled by "N" therefore corresponds to the energy difference $\mu(N+1) - \mu(N)$ between the N and $N+1$ electron ground states. This energy difference, which can also be determined from measurement of the widths of the so-called "Coulomb diamonds" [2], is plotted as a function of N in Fig. 2a. In correspondence to the spacings between the Coulomb oscillations, the energy difference is unusually large for $N = 2$, 6 and 12, and is also relatively large for $N = 4$, 9 and 16. 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 in accordance with Hund's rule [3]. We compare the data in Fig. 2a with an exact calculation for $N = 1$ to 7 shown in Fig. 2b. This calculation incorporates the effect of the finite thickness of the disk. We note that this effect weakens the Coulomb interactions in the dot, and thus makes the shell structure more visible. The thin disk thickness freezes the electrons in the lowest state in the vertical direction. We therefore only have to consider the confinement in the plane of the disk for which we take a parabolic potential $V(r) = \frac{1}{2} m^* \omega_0^2 r^2$, where $m^* = 0.06m_0$ is the effective mass of electrons in the InGaAs disk, ω_0 is the characteristic frequency of the lateral confinement and r is the distance measured from the center of the disk. Details of the calculation technique are given in Ref. [7]. The strength of the Coulomb interactions can be represented by a parameter $Q = e^2/\varepsilon l_0$, where l_0 $(= \sqrt{\hbar/m^* \omega_0})$ is the spatial extension of the lowest state's wave function for parabolic confinement. As ω_0 becomes large, or as the quantum mechanical confinement becomes strong, the strength of Coulomb interactions relative to that of quantum confinement $(= Q/\hbar\omega_0)$ becomes progressively small $(\propto \omega_0^{-3/2})$ [8]. Consequently, as $\hbar\omega_0$ varies from 3 to 5 meV in the calculation, the $N = 2$ and 6 peaks linked to the complete filling of shells become significantly large as compared to the $N = 4$ peak. If we inspect carefully the peaks relative to the background, we find that the agreement with the experimental data is good for $\hbar\omega_0 > 4$ meV when $N \leqslant 2$, and becomes better for a smaller $\hbar\omega_0$ value as N increases. Note that in the calculation the $N = 2$ peak is almost missing for $\hbar\omega_0 = 3$ meV and the $N = 6$ peak is too high for $\hbar\omega_0 = 5$ meV to reproduce the experiment. These arguments on $\hbar\omega_0$ are also supported by the experiment
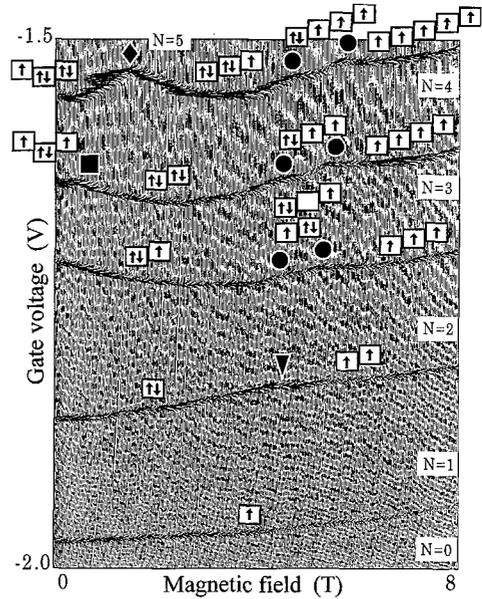

Fig. 3. $I(V_g, B)$ for $N = 0$ to 5 measured with small $V_{sd} = 0.1$ mV such that only ground states contribute to the current. Different types of ground state transitions are indicated by different labels. The arrows in the squares indicate the spin configuration. The lowest square corresponds to a single particle state with angular momentum $l = 0$. For squares to the right l increases to 1, 2, 3, etc. For $N = 4$ and 5, near $B = 0$ T, also the $l = -1$ square is shown to the left of the $l = 0$ square.

on the B-field dependence of the Coulomb oscillation peaks (see Ref. [3] and Section 3.3). Decrease of $\hbar\omega_0$ with increasing N can be explained by the effect of Coulomb screening from the leads and gates which is not incorporated in the calculation [9]. We note that the background relative to the peaks is significantly smaller in the calculation than that in the experiment. This can also be due to the screening effect.

Fig. 3 shows the magnetic field dependence of the first five current peaks for $V_{sd} = 0.1$ mV. The single dot device is similar to, but not the same as, that used for the experiment of Fig. 2a. The B-field dependence of the peak positions in gate voltage reflects the evolution of the ground state energies. Besides an overall smooth B-field dependence, we see several kinks, which we indicate by different labels. As we discuss in the following section, these kinks are assigned to transitions in the ground states, so for the regions between the kinks, we can identify the quantum numbers, including the spin configurations.
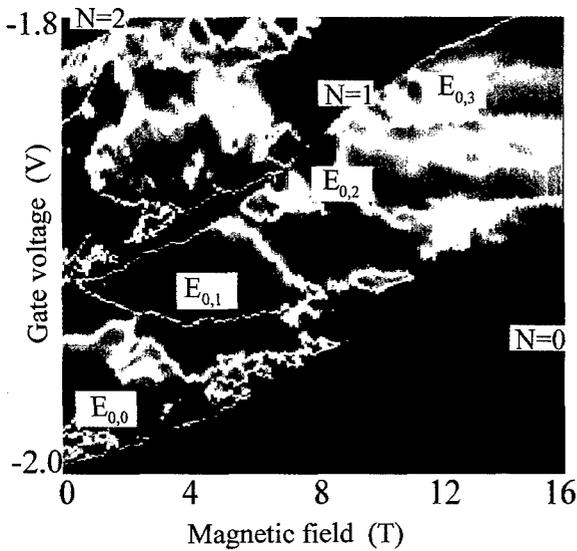
Fig. 4. $I(V_g, B)$ for $N = 0$ to 2 measured with $V_{sd} = 5$ mV up to 16 T. $I < 0.1$ pA in the dark blue regions and $I > 10$ pA in the dark red regions. Both the ground state and the first few excited states can contribute to the current. Current stripes between the Coulomb blockade regions (black) for $N = 0$ and 1 electrons, and for $N = 1$ and 2 electrons are the first and second current stripes, respectively. The states in the first stripe are indexed by the quantum numbers (n, l).

Fig. 5. Exact calculations of the ground states (thick lines) and excited states (thin lines) for $N = 1$ and 2. The $N = 1$ states are indexed by single-particle eigenenergies $E_{n, l}$. The $N = 2$ many-body states are indexed by the total spin and total angular momentum (boxed number). The lowest thick line of each stripe is the ground state energy. The upper thin dashed line of each stripe is the ground state energy shifted upwards by 5 meV. The solid and dotted lines indicate $S = 0$ and 1 configurations, respectively. The label ▲ is discussed in the text.

3.2. Excitation spectroscopy

To investigate what kind of many-body states are responsible for the kinks observed in Fig. 3, we measure I vs. V_g for a large V_{sd} of 5 mV. The data are shown in Fig. 4 for $N = 1$ and 2. For this particular voltage, the two stripes just touch at $B = 0$ T. A pronounced current change, as indicated by the colour change from dark blue to red (i.e. from < 1 pA to > 10 pA), enters the upper edge of the first stripe at $B = 0.2$ T. This change identifies the position of the first excited state for the $N = 1$ dot (we discuss the index $E_{0,1}$ below). Note that at higher B values two higher excited states also enter from the upper edge of the stripe at 5.7 and 9.5 T, respectively. The energy separation between the ground state and the first excited state can be read directly from the relative position inside the stripe. So, the excitation energy ($= \hbar\omega_0$) is slightly larger than 5 meV at $B = 0$ T and decreases for increasing B. Note that even over this wide magnetic field range of 16 T, the first excited state never crosses with the ground

state (lower edge of the stripe). In the second stripe of Fig. 4, however, we see the first excited state cross with the ground state at $B = 4.15$ T, i.e. the first excited state for $B < 4.15$ T (seen as the current change from blue to red inside the second stripe) becomes the ground state for $B > 4.15$ T. Located exactly at this magnetic field is the kink labeled by ▼ in Fig. 3, so it is assigned to a crossing between the ground state and the first excited state. In a similar fashion, we are able to identify a crossing between the ground state and the first excited state corresponding to each kink in Fig. 3 [4].

3.3. Exact calculation of many-body states

For a few electrons the energy spectrum can be calculated exactly [10]. Fig. 5 shows exact calculations of the electrochemical potential vs. B for the $N = 1$ and 2 ground states (thick lines) and excited states (thin lines). We use $\hbar\omega_0$ of 5.5 meV and make the same assumptions as those taken above in Section 3.1. For

each stripe the lower thick line is the ground state, and the upper thin dashed line is the lower thick line shifted upwards by 5 meV. This indicates the transport window for $V_{sd} = 5\,mV$. For the first electron the exact calculation gives single particle states. The eigenenergies with radial quantum number $n = 0, 1, 2, \ldots$ and angular momentum quantum number $l = 0, \pm1, \pm2, \ldots$ are given by [11]

$$E_{n,l} = (2n + |l| + 1)h\sqrt{\left(\tfrac{1}{4}\omega_c^2 + \omega_0^2\right)} - \tfrac{1}{2}lh\omega_c, \quad (1)$$

where the cyclotron frequency $\omega_c = eB/m^*$. (We neglect the much smaller Zeeman energy.) The states lying within this transport window can be compared with the observed current changes seen in the first stripe of Fig. 4. The agreement is very good for both the ground state and the first excited state over the whole B-field range. For the $N = 2$ case, many excited states are lying within the transport window. We index the ground and excited states by the total spin, S, and the total angular momentum, M (boxed number). The solid and dotted lines indicate the $S = 0$ (spin singlet) and $S = 1$ (spin triplet) states, respectively. Whilst the $N = 1$ ground state $E_{0,0}$ never crosses with the first excited state, we see a crossing (labeled by ▲) between the ground state with $(S, M) = (0, 0)$ and the excited state with $(S, M) = (1, 0)$, which is referred to as singlet–triplet transition [12]. For $\hbar\omega_0 = 5.5\,meV$ this singlet–triplet transition is expected at $B = 4.0\,T$, which is in good agreement with the experimental value in Figs. 3 and 4. Note that the calculated excited states with $(S, M) = (1, -1)$ and $(0, 2)$ for $N = 2$ can also be seen in the second stripe of Fig. 4 (i.e. the lines between blue and red current regions which show a maximum near $\sim 2\,T$). The $(1, -1)$ is located $\approx 3\,meV$ above the ground state of $(0, 0)$ at $B = 0\,T$. This position is well predicted by the calculation of Fig. 5. The excitation energy is significantly smaller than the single-particle excitation energy of $E_{0,1} (= \hbar\omega_0)$ due to the exchange effect of the parallel spin of the electrons for the $(S, M) = (1, -1)$ state.

Now we discuss the energy spectrum for $N = 3$ and 4. Exact calculations of the ground and excited states are shown in Fig. 6. The three electron ground state has two transitions labeled by ●. On increasing B, the total spin and the total angular momentum of the many-body states change from $(S, M) = (1/2, 1)$ to $(1/2, 2)$ at 4.3 T, and then to $(3/2, 3)$ at 4.8 T. The transitions to
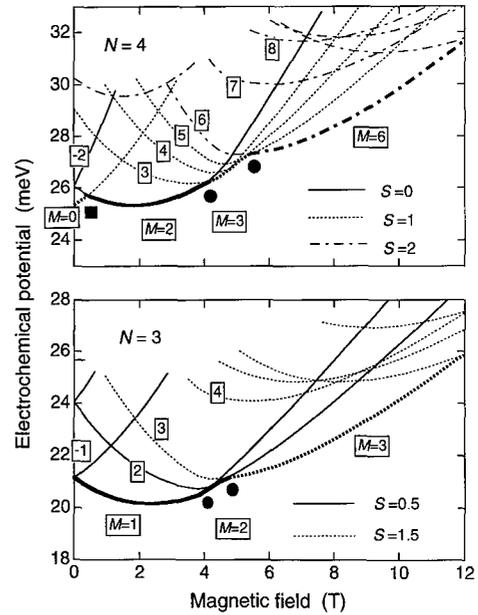

Fig. 6. Exact calculations of ground states (thick lines) and excited states (thin lines) for $N = 3$ and 4. Many-body states are indexed by the total spin and angular momentum (boxed numbers). For $N = 3$, the solid and dotted lines indicate $S = 0.5$ and 1.5, respectively. For $N = 4$, the solid, dotted and dot-dashed lines indicate $S = 0$, 1 and 2 respectively. The ■, ● labels are discussed in the text.

larger angular momentum states reduce the Coulomb interactions. In addition, the total spin increases to gain exchange energy. A double transition in the ground state energy is indeed observed in the form of two kinks in the third peak trace of Fig. 3. In most regions in Fig. 3 there is one main configuration for the occupation of single-particle states. For $N = 3$, however, in the region between the two ● labels there are two important configurations, which both have the same total spin, and total angular momentum. The four electron ground state has five transitions: $(S, M) = (1, 0)$ to $(0, 2)$ at 0.43 T, then to $(1, 3)$ at 4.0 T, to $(1, 4)$ at 4.9 T, to $(1, 5)$ at 5.2 T, and finally to $(2, 6)$ at 5.4 T, respectively. The first transition (labeled by ■) is associated with breakdown of Hund's rule [3] and the other transitions can be understood in the same way as discussed above for the three electron ground state. These transitions are indeed observed as kinks in Fig. 3. The transitions between the two ● labels are not so evident in the experimental data. This is prob-

ably because different states in this region lie close in energy, and the ground state does not show a critical change in electrochemical potential at the transition points (see Fig. 6). We note that the first few excited states in the calculation of Fig. 6 are also observed in the measurement of the excitation spectra [4]. For B larger than the right most ● for $N = 3$ and 4 there is again a distinct ground state in which the electrons are fully spin-polarized and occupy sequential momentum states.

4. Electronic states in quantum dot molecules

A quantum dot molecule can be realized in the same vertical device configuration as for a quantum dot atom except that the double barrier structure is replaced by a triple barrier structure [13]. The outer barriers have the same thickness of 7.0 nm. Quantum mechanical coupling between the two dots form symmetric and anti-symmetric states. By changing the thickness of the central barrier, b, from 7.5 to 2.5 nm we are able to increase the energy splitting between symmetric and anti-symmetric states, $\Delta_{S\,AS}$, from about 0.09 to 3.4 meV. Quantum mechanically, we consider the dots separated by a 7.5 nm barrier to be "weakly" coupled, and the dots separated by a 2.5 nm barrier to be "strongly" coupled. As a rough guide, for the case of two electrons trapped in the system ($N = 2$), the lateral confinement energy, $\hbar\omega_0$, $= 4$ meV, a typical average "classical" charging energy, $E_{classical} = 3$ meV, and an electrostatic coupling energy, $E_{electrostatic} = 0.7$ meV respectively are comparable to $\Delta_{S\,AS}$ for $b = 2.4, 2.8$, and 4.8 nm. The $\hbar\omega_0$ value is slightly smaller than that for a single dot probably because the lateral electrostatic confinement is weaker as the system is larger. We note that quantum mechanical coupling is not the only coupling mechanism in *artificial molecules*. In the regime where $(\hbar\omega_0 >) E_{electrostatic} \gg \Delta_{S\,AS}$, it is electrostatic coupling between the dots which becomes important [14]. Competition between the two mechanisms as b is varied is expected to have a profound effect on the transport properties of the two dot system. Here we focus on the strongly coupled double dot system.

Fig. 7 shows a grey scale plot of dI/dV_{sd} in the V_{sd}–V_g plane for a $D = 0.56\,\mu m$ "strongly" coupled double dot device ($b = 2.5$ nm). Black (positive values of dI/dV_{sd}) and white (negative values of dI/dV_{sd}) lines criss-crossing the plot and running parallel to the sides of the diamonds identify bound and excited states – details of which will be published elsewhere. Well formed Coulomb diamonds (grey regions where $I = 0$ pA) close to zero bias from $N = 1$ to 22 are evident. The symmetry of the diamonds with respect to the bias direction confirms that the states responsible are indeed delocalized over both dots. Notice that the $N = 2$, and $N = 6$ diamonds are unusually large compared to the adjoining diamonds. As for the single dot the half width of the Nth diamond is a direct measure of $\mu(N + 1) - \mu(N)$. The $\mu(N + 1) - \mu(N)$ values obtained from Fig. 7 are shown by the black circles in Fig. 2b. For this double dot device we see the same magic numbers 2, 6, and 12 as for the single dot device although, intriguingly, and for reasons which are not yet understood, 4, 9, and 16 are apparently absent. The 3 meV value of $E_{classical}$ is also in

Fig. 7. Grey scale plot of dI/dV_{sd} in the V_{sd}–V_g plane for a $D = 0.56\,\mu m$ *quantum mechanically* "strongly" coupled double dot device. Coulomb diamonds similar to those for a single dot are formed from $N = 1$ to 22 close to zero bias. The half width in eV_{sd} of a diamond shaped region is a direct measure of the change of electrochemical potential when one more electron is added to the double dot system.

line with the experimental data for $N = 1, 2$, and 3 in Fig. 2b. Note that for $N > 15$, $\mu(N + 1) - \mu(N)$ is approximately half that of the single dot. This is reasonable because the double structure dot occupies roughly twice as much volume. Finally, for this *artificial molecule*, there is no evidence from Figs. 2b and 7 for the occupation of anti-symmetric states for $N \leqslant 12$, i.e. the first 12 electrons all occupy the symmetric states and are delocalized. This might look inconsistent with a single-particle picture as the symmetric and anti-symmetric states can only be distinguished in the presence of quantum mechanical coupling, and both sets of lateral states have an identical single-particle energy spectrum with a characteristic confining energy of $\hbar\omega_0$. Putting electrons consecutively into the symmetric states costs much single-particle excitation energy. For example, the complete filling of electrons in the second shell costs $4\hbar\omega_0$, which is much greater than Δ_{SAS}. However, besides Δ_{SAS}, there are a number of important interaction effects that determine the filling of electrons. These effects are the exchange effects within either the symmetric states or within the anti-symmetric states, are between symmetric and anti-symmetric states, screening effects, and direct Coulomb repulsion. Coulomb repulsion favours the filling of laterally delocalized electrons, i.e. p-type electrons rather than s-type electrons, so the consecutive filling of electrons in to the symmetric states is favoured. This consecutive filling is also promoted by the screening effect since it reduces the lateral confinement energy with increasing N. These three factors help to explain the experimental data for this strongly coupled dot device. The exchange effect between the symmetric and anti-symmetric states favours the filling of electrons in to the anti-symmetric states. However, this effect can be weaker than the other effects described above. Exact calculations incorporating many-body interactions are necessary for more detailed arguments. These are underway and will be discussed elsewhere.

5. Conclusions

We have studied the atomic-like properties of a single disk shaped dot and the molecular-like properties of two vertically coupled disk-shaped dots. For the single dot the addition energy spectrum for the few electron ground states at $B = 0\,\mathrm{T}$ and in the presence of a magnetic field induced transitions in the ground states and these compare well to the exact calculations of many-body states in an artificial atom. For the strongly coupled double dot device the few electron ground states show properties similar to those of a single dot, indicating that the first few electrons only occupy the symmetric states and are delocalized throughout the whole system.

References

[1] M. Reed, Scientific American 268 (1993) 118; M. Kastner, Physics Today 46 (1993) 24; R.C. Ashoori, Nature 379 (1996) 413.

[2] See for a review: Proc. Advanced Study Institute on Mesoscopic Electron Transport, Curacao, June 1996, Series E, Kluwer, Dordrecht, 1997.

[3] S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, L.P. Kouwenhoven, Phys. Rev. Lett. 77 (1996) 3613.

[4] L.P. Kouwenhoven, T.H. Oosterkamp, M.W.S. Danoesastro, M. Eto, D.G. Austing, T. Honda, S. Tarucha, Science 278 (1997) 1788.

[5] D.G. Austing, T. Honda, S. Tarucha, Semiconductor Sci. Technol. 11 (1995) 212.

[6] S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, L.P. Kouwenhoven, Jpn. J. Appl. Phys. 36 (1997) 3917.

[7] Y. Tokura, L.P. Kouwenhoven, D.G. Austing, S. Tarucha, Physica B 246–247 (1998) 83.

[8] This ratio of $Q/\hbar\omega_0$ is for example 1.5 for a parabolic potential with $\hbar\omega_0 = 5\,\mathrm{meV}$. It is assumed to be 1 in our previous paper (see Ref. [4]).

[9] Y. Tanaka, H. Akera, J. Phys. Soc. Japan 66 (1997) 15.

[10] See for example: J.J. Palacios, L. Martin-Moreno, G. Chiappe, E. Louis, C. Tejedor, Phys. Rev. B 50 (1994) 5760; See for more references the review by N.F. Johnson, J. Phys.: Condens. Matter 7 (1995) 965.

[11] V. Fock, Z. Phys. 47 (1928) 446; C.G. Darwin, Proc. Cambridge Phil. Soc. 27 (1930) 86.

[12] See for theory: M. Wagner, U. Merkt, A.V. Chaplik, Phys. Rev. B 45 (1992) 1951; P. Hawrylak, Phys. Rev. Lett. 71 (1993) 3374; See for experiment: B. Su, V.J. Goldman, J.E. Cunningham, Phys. Rev. B 46 (1992) 7644; R.C. Ashoori et al., Phys. Rev. Lett. 71 (1993) 613; T. Schmidt et al., Phys. Rev. B 51 (1995) 5570.

[13] D.G. Austing, T. Honda, Y. Tokura, S. Tarucha, Jpn. J. Appl. Phys. 34 (1995) 1320.

[14] D.G. Austing, T. Honda, K. Muraki, Y. Tokura, S. Tarucha, Physica B 249–251 (1998) 152.

Raman spectroscopy of quantum wires and dots: magnetoplasmons and edge-spin-density modes

Christian Schüller *

Institut für Angewandte Physik und Zentrum für Mikrostrukturforschung, Universität Hamburg, Jungiusstraße 11, D-20355 Hamburg, Germany

Abstract

We review recent magneto-Raman experiments on electronic excitations in GaAs–AlGaAs quantum wires and dots. We observe simultaneously collective spin-density and charge-density excitations (SDEs and CDEs), and also single-particle-like excitations. In a magnetic field, the CDEs (magnetoplasmons) exhibit a complex anticrossing behavior, which, in quantum wires, strongly depends on the direction of lateral wave-vector transfer. In quantum dots, we observe a splitting of the lowest spin-density mode in a magnetic field, very similar to the splitting of the magnetoplasmons. The lower branch can be interpreted as an edge-spin-density mode, where the electrons move on skipping orbits around the edge of the dot. In a microscopic model, we can identify transitions between electronic shells, which contribute to the observed low-energy SDEs of the quasiatomic system. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Quantum wires; Quantum dots; Electronic excitations; Raman spectroscopy

1. Introduction

Low-dimensional electron systems, realized in semiconductor nanostructures, have evolved to an important field in modern solid-state research during the past decades. Resonant inelastic light scattering is an ideal tool to study the electronic elementary excitations in such systems, since it allows to distinguish between different types of excitations by polarization selection rules [1]: collective charge-density excitations (CDEs) can be observed if the polarization directions of the incident and scattered photons are parallel to each other (polarized geometry), and collective spin-density excitations (SDEs) are measured for perpendicular polarizations (depolarized geometry). In most cases, CDEs have larger energies than the corresponding single-particle transitions due to direct and exchange Coulomb interactions, whereas SDEs are shifted to lower energies because they are only affected by the exchange interaction. In addition, the Raman experiments offer also the possibility to transfer and vary a finite wave vector q, which can be transferred to the excitations.

The first observation of intersubband SDEs and CDEs in GaAs–AlGaAs heterostructures and quantum wells was reported by Abstreiter [2] and Pinczuk [3] in their pioneering works. Since then, a wealth of

* Tel.: + 49 40 4123 5308; fax: + 49 40 4123 6332; e-mail: schueller@physnet.uni-hamburg.de.

experiments on quasi-two-dimensional (2D) electron systems followed, which demonstrated the power of the resonant light scattering technique [4]. In 1989, Pinczuk observed in high-mobility quantum wells surprisingly an additional excitation, which energetically is in between the intersubband SDE and CDE, and which occurs in *both* polarization configurations [5]. This excitation showed all features, which one expects from unrenormalized single-particle excitations (SPEs) and was so far interpreted as SPE. During the past decade, the light scattering method has also been successfully applied to quantum wires [6–13] and dots [14–17].

In this article we review our recent Raman experiments on modulation-doped deep-etched GaAs–AlGaAs quantum wires and dots, where we have investigated the wave vector (q) and magnetic-field (B) dependence of electronic excitations [11,12,15,17]. In quantum wires, we observe for a wave vector transfer parallel to the wires a well-pronounced negative B dispersion of the intraband CDE that arises from the skipping-orbit motion of the individual electrons within the collective excitation. At higher frequencies, we find a set of additional modes that represent a superposition of freely propagating oscillations along the wire and confined oscillations perpendicular to the wires. These modes exhibit in a magnetic field a strong interaction with harmonics, $n = 2, 3 \ldots$, of the cyclotron frequency $n\omega_c$. In quantum dots, we can identify, by a detailed analysis of q- and B-dependent measurements, transitions between different electronic shells of the quasiatomic systems, which predominantly contribute to the observed low-energy SDEs and resemble an atomic fine structure. In a magnetic field, the lowest SDE branch can, in a classical picture, be interpreted as an edge-spin-density mode.

2. Electronic elementary excitations in quantum wires and dots

Quantum-wire and quantum-dot samples were prepared by a reactive-ion-deep-etching process, starting from one-sided modulation-doped GaAs–Al$_{0.3}$–Ga$_{0.7}$As single quantum wells with 25 nm well width. The low-temperature carrier densities and mobilities of the quantum-well samples were in the range of

$(4–8) \times 10^{11}$ cm^{-2} and 7×10^5 cm^2 V^{-1} s^{-1}, respectively. The etch masks have been defined by holographic lithography, using the 458 nm line of an Argon laser. The periods of the wire and dot samples are in the range of (240–800) nm and the geometrical structure sizes are between (120–270) nm. This results in about 4–12 occupied quasi-one-dimensional (1D) subbands for the different wire samples, and electron numbers in the order of 100 per dot in the dot samples. The Raman experiments have been performed using a tunable Ti : sapphire laser and a triple Raman spectrometer with a charge-coupled-device camera for detection. All measurements were done at temperatures between 2 K and 12 K.

The lateral confining potential in the deep-etched samples is determined by the homogeneously distributed ionized remote donors in the AlGaAs barrier, and negatively charged surface states. By analytical calculations one can show that this lateral potential, which acts on a test electron within the structure (*bare* potential), is for the case of dots [18] in x and y direction, and, in wires, for the direction perpendicular to the wires (in the following y direction) in good approximation parabolic. We will therefore start our investigations by assuming that the electron motion in the wires is in y direction quantized by a parabolic potential and in x direction quasifree. The two quantum numbers, which then characterize this problem, are a discrete 1D subband quantum number $i = 0, 1, 2, \ldots$ for the quantized movement, and a quasi-continuous quantum number k_x. The corresponding energy eigenvalues are given by

$$E_{i, k_x} = \hbar \Omega_0 (i + \tfrac{1}{2}) + \frac{\hbar^2 k_x^2}{2m^*}. \tag{1}$$

Ω_0 is the quantization energy and m^* the effective mass. This situation is sketched in the insets of Fig. 1a for the case of 3 occupied 1D subbands. Of course, the presence of many other electrons in the samples leads to a screening of the potential, which acts on the test electron. Theoretically, this is described, e.g., in a self-consistent Hartree calculation of the ground state. Thus, the *effective* lateral potential is flattened in the center of the wire, i.e. it gets more and more square-well-like. This results in a decrease of the single-particle subband spacings compared to the case of the bare potential, where only one test electron is
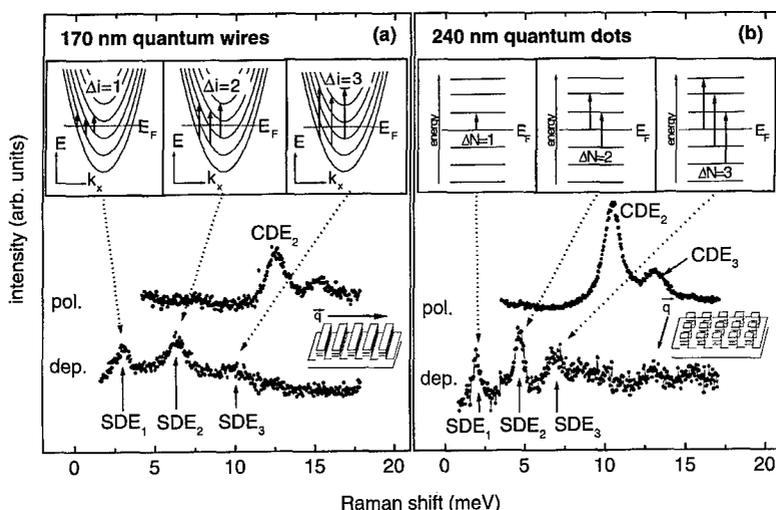
Fig. 1. Depolarized and polarized Raman spectra of electronic excitations in (a) a quantum-wire and (b) a quantum-dot sample. The insets show schematic diagrams of the single-particle transitions, which predominantly contribute to the observed excitations. The wave vector transfer q in the experiments was $q \approx 1.3 \times 10^5$ cm^{-1}.

assumed within the structure. For the bare potential of quantum dots with cylindrical symmetry, the corresponding single-particle energies are given by the eigenvalues of a two-dimensional harmonic oscillator

$$E_{nm} = \hbar\Omega_0(2n + |m| + 1) = N\hbar\Omega_0. \qquad (2)$$

Here the two relevant quantum numbers are the radial quantum number $n = 0, 1, 2,\ldots$ and the azimuthal or angular-momentum quantum number $m = 0, \pm 1, \pm 2,\ldots$. We define a *lateral* quantum number $N = (2n + |m| + 1)$, which characterizes the $2N$-fold degenerate discrete levels, as sketched in the insets in Fig. 1b. Also here, as in the case of wires, the effective potential will certainly deviate from a parabolic shape due to screening. This will result in renormalizations of energy levels, and lifting of degeneracies, as we will see in Section 3. For simplicity, we stay for the moment in the simple picture, where we assume that also the effective, or Hartree, potential is parabolic. In such a single-particle picture, the electronic excitations, which can be created in the inelastic light scattering process, are transitions of electrons from occupied to unoccupied states, as, e.g., sketched in the insets in Fig. 1a and b.

In Fig. 1, polarized and depolarized spectra of electronic excitations in a quantum-wire (Fig. 1a) and a quantum-dot sample (Fig. 1b) are shown. In the ex-

periments on wires, the wave vector q is transferred perpendicular to the wires. This is achieved by tilting in backscattering geometry the sample normal with respect to the directions of incoming and scattered light (see also inset in Fig. 4a in Section 3). Several peaks can be observed, which can be identified as either SDEs or CDEs due to polarization selection rules. These spectra were recorded at a laser frequency well above the effective band gap of the underlaying 2D structure. From our investigations at different laser frequencies [15] we know that here the collective excitations are created by a third-order excitonic scattering mechanism [21,22]. The index j of the labels SDE$_j$ or CDE$_j$ in Fig. 1 gives the change in lateral quantum number for single-particle transitions, which dominantly contribute to the observed collective excitations [11]. The quite large differences in the energetic positions of the SDE$_j$ and CDE$_j$ are due to many-particle interactions [15]. In a macroscopic view, the CDEs in wires can be regarded as electron-density oscillations perpendicular to the wire direction with, e.g., one node in the induced density for the CDE$_1$, and two nodes for the CDE$_2$. Within this picture, these CDEs are also entitled *confined plasmons*. The peculiarity in wire, and also dot, structures with parabolic bare potential is that the energy of the CDE$_1$, the so-called Kohn's mode, is exactly equal to the single-particle spacing
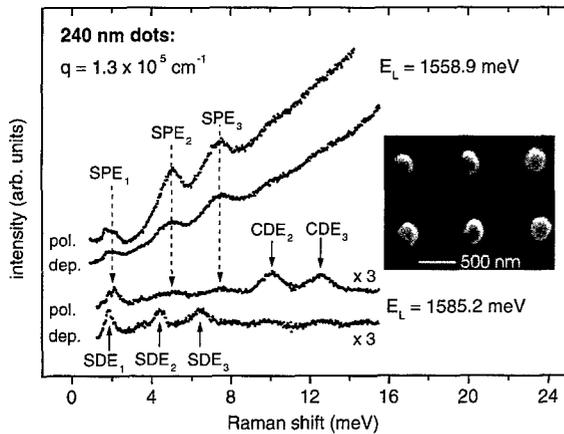
Fig. 2. Depolarized and polarized Raman spectra of electronic excitations in a quantum-dot sample for different laser frequencies E_L. The background in the upper spectra is due to hot luminescence. The inset shows a scanning electron micrograph of the investigated quantum-dot sample.

of the bare potential. The reason is the so-called generalized Kohn theorem [19]: the CDE_1 is in the long-wavelength limit a rigid center-of-mass oscillation of all electrons. Therefore, this mode has a large dipole moment and can be observed in far-infrared transmission experiments [20]. For the SDEs we find in our samples that the energies are very close to unrenormalized single-particle transitions of the self-consistently screened *effective* potential, [1] as sketched in the insets of Fig. 1: This can also be deduced from Fig. 2. Here spectra of quantum dots are shown for a laser frequency which is further lowered towards the band gap (lower part), and at conditions of extreme resonance (upper part), which means that the laser frequency is close to the effective band gap. Here excitations occur (marked by dashed arrows), which get more and more resonant, and finally dominate the spectra, by lowering the laser frequency towards the effective band gap. We find for all our samples – quantum wells, wires, and dots – that at conditions of extreme resonance these excitations, which can be interpreted as SPEs, can be observed in both polarization configurations [15]. Nevertheless, we are very well aware of

----

[1] We note that in these samples we have about 9 occupied 1D subbands, or, about 200 electrons per dot, in the case of dots. For simplicity we have considered in the insets smaller electron numbers.

the fact that in a many-particle system all excitations must be many-particle excitations. So we believe that a rigorous theoretical treatment, by inclusion of the valence band and many-particle interaction, which so far has not been performed, should prove that these excitations also have small but finite collective shifts.

3. Magnetoplasmons and edge-spin-density modes

In order to learn something about the internal structure of the observed excitations, we have also performed experiments in an external magnetic field B, which was oriented perpendicular to the plane of the wires or dots. As examples we discuss here (i) the anticrossing of the CDEs in quantum wires [11,12] due to the coupling with Bernstein modes [23], which can in a similar manner also be observed in quantum dots, and (ii) the formation of edge-spin-density modes in quantum dots, which reveals insight into the electronic structure of the artificial atoms [17].

Fig. 3a shows the magnetic-field dispersions (solid symbols) of confined plasmons for a wave vector transfer perpendicular to the wires, as discussed in the previous section. The CDE_1, which is in the long-wavelength limit the Kohn's mode, i.e., a rigid center-of-mass oscillation of all electrons perpendicular to the wires, shows no significant splitting in the measured B range. This is consistent with the fact that the external lateral potential is to a good approximation parabolic because in this case the generalized Kohn theorem holds. Interestingly, the second confined plasmon, the CDE_2, exhibits a strong anticrossing at $2\omega_c$ due to the coupling with Bernstein modes. The CDE_2 is a Raman-active mode (see below) and has no dipole moment at all [11]. This implies that it purely consists of relative motions of the interacting electrons, so that it is clear that here Kohn's theorem cannot be applied. The small open circles in Fig. 3 are results of a calculation. Here the effective lateral potential was determined in a self-consistent Hartree calculation and the Raman response was calculated within the random-phase approximation (RPA). The curves displayed in Fig. 3 show the dispersions of the strongest modes. These calculations also demonstrate that in the anticrossing regime the modes can no longer be characterized, as for other magnetic fields, by specific single-
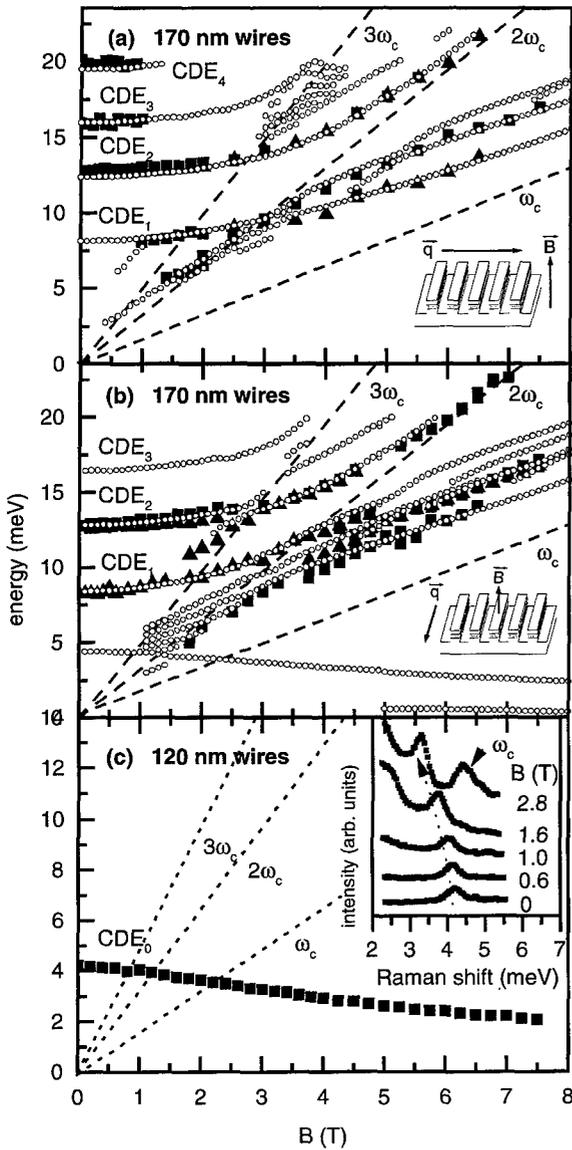
Fig. 3. Experimental (full symbols) and theoretical (open symbols) magnetic-field dispersions of magnetoplasmons in quantum-wire samples for different directions of wave vector transfer q with respect to the wire direction.

particle transitions, but rather a strong intermixing of different single-particle contributions occur [11].

In Fig. 3b, the results for a wave vector transfer q in wire direction are displayed [12]. Here the macroscopic density oscillations, which are characteristic for the excitations, consist of a travelling wave along the wires and, at the same time, an oscillation perpendicular to the wires. As we can see from Fig. 3b, in this case also the Kohn's mode splits in the vicinity of $2\omega_c$, which demonstrates that Kohn's theorem does not hold, because we are no longer in the long-wavelength limit $q \approx 0$.

The theoretical curves in Fig. 3b also show one mode, starting at about 4 meV at $B = 0$, which shows a negative B dispersion. This mode is a 1D plasmon, i.e., a plasma wave which moves along the wires [24], or, in a microscopic picture, it is an *intraband* plasmon (CDE_0). The negative B dispersion results from a skipping-orbit motion of the individual electrons at the edges of the wires. We could not observe this excitation in wire samples with relatively large periods and wire widths, as, e.g., displayed in Figs. 3a and b. Therefore we have prepared a series of samples with smaller periods and wire widths, as for example shown in Fig. 3c. In these samples we indeed observe this mode with a negative B dispersion. For illustration, polarized spectra are displayed in the inset. We have also calculated the dispersion of this skipping-orbit mode. The calculations, which are performed for isolated wires, reproduce the negative dispersion, however, they do not give the correct energy. The calculated energy is typically only half of the experimentally observed one. We attribute this experimentally observed energy increase to the coupling between wires in these small period arrays. Such an increase due to coupling is suggested by the calculation of Li and DasSarma [25].

We now come to the discussion of SDEs in quantum dots [17]. Fig. 4a shows depolarized spectra of a quantum-dot sample with 240 nm dots for different wave vector transfer q in lateral direction. This wave vector transfer was achieved by tilting the sample normal with respect to the directions of incoming and scattered light, as displayed in the inset of Fig. 4a. In a symmetric system, Raman-allowed excitations have even parity because of the two-photon nature of the Raman process. At small q, only the Raman allowed SDE_2 is observed. This enhancement of the *second* mode is even more pronounced for the CDE_2 in dots and also in wires, as can be seen in Fig. 1 in the previous section. By increasing q, the symmetries of the excitations are broken and also Raman-forbidden excitations with odd parity occur (SDE_1 and SDE_3 in Fig. 4a). Furthermore, an additional fine structure can
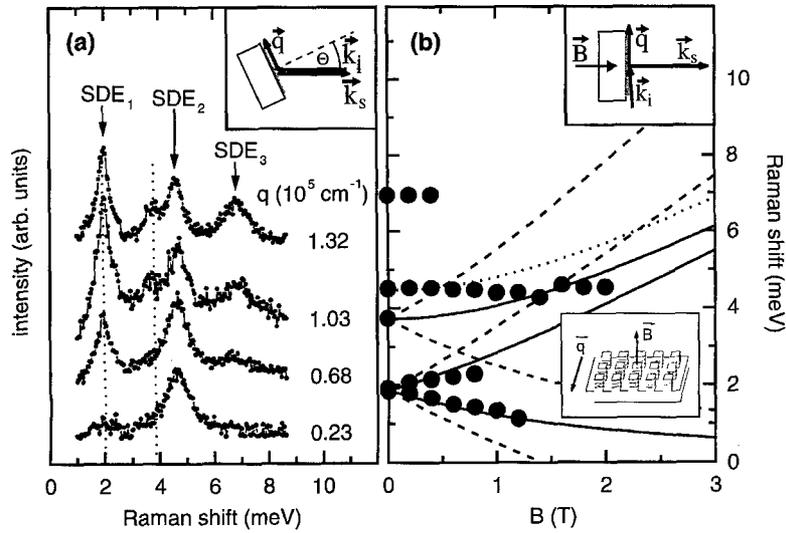
Fig. 4. (a) Spectra of spin-density excitations in a 240 nm quantum-dot sample for different wave vector transfer q. (b) Magnetic-field dispersions of the spin-density modes. The lines are calculated for a parabolic effective potential. The insets show the scattering geometries, which were used for the experiments (a) without and (b) with magnetic field. In (b) the wave vector transfer was $q = 0.8 \times 10^5$ cm^{-1}.
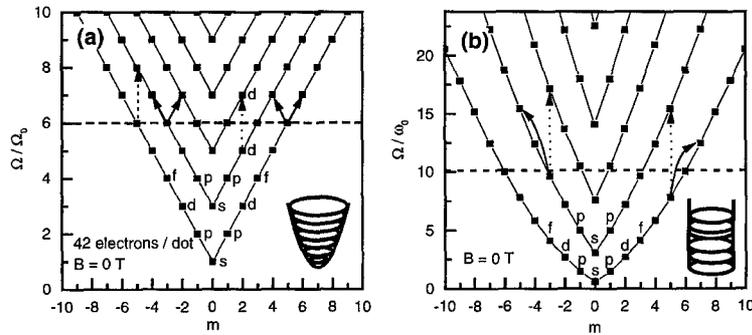

Fig. 5. Energy spectrum for (a) a two-dimensional harmonic oscillator, and (b) a hard-wall potential with cylindrical symmetry. The potentials are sketched in the insets. Ω_0 is the quantization energy of the parabolic potential, and $\omega_0 = \hbar/(2m^*)(\pi/a)^2$ characterizes the hard-wall potential with a being the diameter.

be observed at the low-energy side of SDE$_2$ at large q. Fig. 4b displays the measured B dispersions of the spin-density modes (full circles). In these experiments in a split-coil cryostat, we used the scattering geometry sketched in the inset of Fig. 4b. Therefore we are restricted to a wave vector transfer $q \approx 0.8 \times 10^5$ cm^{-1}. At this relatively small q, we cannot follow all peaks, especially the fine structure peak, to finite magnetic fields. Also the signals of the SDE$_3$ are very weak and broad so that we cannot definitely say if there is a splitting or not. As given by Eq. (2), the single-

particle states in a quantum dot are characterized by a radial quantum number $n = 0, 1, 2, \ldots$ and an angular-momentum quantum number $m = 0, \pm 1, \pm 2, \ldots$. We start the interpretation of our observations in a simple model, by assuming that also the *effective* lateral potential in our quantum dot is parabolic. The energy spectrum of a parabolic quantum dot, given by Eq. (2), is displayed in Fig. 5a. Here the lines connect points which belong to the same n, where n rises from the bottom to the top. By the labels, e.g., "s" or "p", we have indicated the character of the quasiatomic orbitals as

quasi-s or quasi-p orbitals. Of course, these flat quasi-atomic orbitals have different symmetries than orbitals of real 3D atoms. From the observed splitting of the SDE_1 in a magnetic field (see Fig. 4b), we can deduce that the SDE_1 dominantly consists of transitions with changes $(\Delta n, \Delta m)$ in radial and azimuthal quantum numbers $\Delta n = 0, 1$, and $|\Delta m| = 1$. Some of these transitions are marked by full arrows in Fig. 5a: in a magnetic field, the spectrum of the parabolic quantum dot is given by

$$E_{nm}(B) = (2n + |m| + 1)\hbar\sqrt{\Omega_0^2 + \left(\frac{\omega_c}{2}\right)^2} + m\hbar\frac{\omega_c}{2}.$$
(3)

The full lines in Fig. 4b, which start at 1.9 meV, give the splitting of the above given transitions, calculated with Eq. (3). All other possible transitions with $\Delta N = 1$ and $|\Delta m| \geq 2$ result in splittings $\geq 2\omega_c$, which are experimentally not observed. For example, the forbidden next higher transitions are indicated by dashed lines, also starting at 1.9 meV, in Fig. 4b. In analogy to the edge magnetoplasmons [26,20], the lower branch of the SDE_1 can be interpreted as an edge-spin-density mode, where the individual electrons within the excitation perform skipping-orbit motions at the edge of the dot.

The SDE_2 shows neither an energy shift nor a significant splitting with B. The only transitions with $\Delta N = 2$, which show such a behavior, have $(\Delta n = 1, \Delta m = 0)$, as indicated, for $B = 0$, by dotted arrows in Fig. 5a. The calculated B dispersion of this mode is given by the solid line starting at 3.8 meV in Fig. 4b. For a parabolic potential, also the experimental points should start at this energy. In contrast, we find the SDE_2 at higher energy (≈ 4.5 meV), and we observe that the additional peak has almost exactly twice the energy of the SDE_1. We claim that this additional peak just stems from transitions with $\Delta n = 0$ and $|\Delta m| = 2$, which are energetically degenerate with transitions $(\Delta n = 1, \Delta m = 0)$ for a parabolic effective potential: in our dots we have approximately 200 electrons per dot and thus screening should be important. So, the real effective potential should be in between a parabolic and a hard-wall potential. In Fig. 5b, we show the energy spectrum for a hard-wall potential with cylindrical symmetry, which should be the extreme limit. This graph shows that (i) transi-

tions with $\Delta n = 0$ and $|\Delta m| = 2$ (solid arrows in Fig. 5(b)) have lower energies than transitions with $\Delta n = 1$ and $\Delta m = 0$ (dotted arrows), and (ii) transitions with $|\Delta m| = 2$ have almost twice the energy than transitions with $|\Delta m| = 1$. This is indeed observed in the experiments as indicated by the two vertical dotted lines in Fig. 4a, and confirms our interpretation.

4. Conclusions

We have investigated the q and B dependence of electronic excitations in quantum wires and dots. In quantum wires with parabolic external potential, we find that for no wave vector transfer *parallel* to the wires, higher index magnetoplasmons show strong coupling to Bernstein modes, whereas the Kohn's mode does not interact. On the other hand, if a finite q is transferred parallel to the wires, also the Kohn's mode splits in the vicinity of $2\omega_c$. Our experimental results are confirmed by self-consistent Hartree-RPA calculations. Furthermore, we observe in short-period wire samples a 1D plasmon, which shows a negative B dispersion due to a skipping-orbit motion of the individual electrons. In these samples, evidence for a coupling between adjacent wires is found. The detailed analysis of the q and B dependence of SDEs in quantum dots allows to identify the transitions between quasiatomic shells, which dominantly contribute to the observed excitations. The low-energy branch of the SDE_1 in a magnetic field is interpreted as an edge-spin-density mode.

C. Schüller / Physica E 3 (1998) 121–128

the Graduiertenkolleg 'Physik nanostrukturierter Festkörper'.

References

[1] D. Hamilton, A.L. McWhorter in: G.B. Wright (Ed.), Light Scattering Spectra of Solids, Springer, New York, 1969, p. 309.

[2] G. Abstreiter, K. Ploog, Phys. Rev. Lett. 42 (1979) 1308.

[3] A. Pinczuk, H.L. Störmer, R. Dingle, J.M. Worlock, W. Wiegmann, A.C. Gossard, Solid State Commun. 32 (1979) 1001.

[4] For an overview see: A. Pinczuk, G. Abstreiter, in: M. Cardona, G. Güntherodt (Eds.), Light Scattering in Solids V, Topics in Applied Physics, vol. 66, Springer, Berlin, 1988, p. 153.

[5] A. Pinczuk, S. Schmitt-Rink, G. Danan, J.P. Valladares, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 63 (1989) 1633.

[6] J.S. Weiner, G. Danan, A. Pinczuk, J. Valladares, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 63 (1989) 1641.

[7] T. Egeler, G. Abstreiter, G. Weimann, T. Demel, D. Heitmann, P. Grambow, W. Schlapp, Phys. Rev. Lett. 65 (1990) 1804.

[8] A.R. Goñi, A. Pinczuk, J.S. Weiner, J.S. Calleja, B.S. Dennis, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 67 (1991) 3298.

[9] A. Schmeller, A.R. Goñi, A. Pinczuk, J.S. Weiner, J.S. Calleja, B.S. Dennis, L.N. Pfeiffer, K.W. West, Phys. Rev. B 49 (1994) 14778.

[10] A.R. Goñi, A. Pinczuk, J.S. Weiner, B.S. Dennis, L.N. Pfeiffer, K.W. West, Phys. Rev. Lett. 67 (1993) 1151.

[11] C. Steinebach, R. Krahne, G. Biese, C. Schüller, D. Heitmann, K. Eberl, Phys. Rev. B 54 (1996) R14281.

[12] E. Ulrichs, G. Biese, C. Steinebach, C. Schüller, D. Heitmann, K. Eberl, Phys. Rev. B 56 (1997) R12760.

[13] C. Steinebach, C. Schüller, G. Biese, D. Heitmann, K. Eberl, Phys. Rev. B 57 (1998) 1703.

[14] R. Strenz, U. Bockelmann, F. Hirler, G. Abstreiter, G. Böhm, G. Weimann, Phys. Rev. Lett. 73 (1994) 3022.

[15] C. Schüller, G. Biese, K. Keller, C. Steinebach, D. Heitmann, P. Grambow, K. Eberl, Phys. Rev. B 54 (1996) R17304.

[16] D.J. Lockwood, P. Hawrylak, P.D. Wang, C.M. Sotomayor Torres, A. Pinczuk, B.S. Dennis, Phys. Rev. Lett. 77 (1996) 354.

[17] C. Schüller, K. Keller, G. Biese E. Ulrichs, L. Rolf, C. Steinebach, D. Heitmann, K. Eberl, Phys. Rev. Lett. 80 (1998) 2673.

[18] U. Merkt, Physica B 189 (1993) 165.

[19] P. Maksym, T. Chakraborty, Phys. Rev. Lett. 65 (1990) 108.

[20] D. Heitmann, J.P. Kotthaus in Physics Today, June 1993, p. 56, and references therein.

[21] G. Danan, A. Pinczuk, J.P. Valladares, L.N. Pfeiffer, K.W. West, C.W. Tu, Phys. Rev. B 39 (1989) 5512.

[22] A.O. Govorov, J. Phys. Condens. Matter 9 (1997) 4681.

[23] I.B. Bernstein, Phys. Rev. 109 (1958) 10.

[24] T. Demel, D. Heitmann, P. Grambow, K. Ploog, Phys. Rev. Lett. 66 (1991) 2657.

[25] Q.P. Li, S. DasSarma, Phys. Rev. B 43 (1991) 11768.

[26] T. Demel, D. Heitmann, P. Grambow, K. Ploog, Phys. Rev. Lett. 64 (1990) 788.

Application of self-organized quantum dots to edge emitting and vertical cavity lasers

D. Bimberg [a,*], N.N. Ledentsov [a], M. Grundmann [a], F. Heinrichsdorff [a], V.M. Ustinov [b],
P.S. Kop'ev [b], M.V. Maximov [b], Zh.I. Alferov [b], J.A. Lott [c],

[a]Institut für Festkörperphysik, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany
[b]A.F. Ioffe Physical Technical Institute, 194021, St. Petersburg, Russian Federation
[c]Air Force Institute of Technology, DECE 2950, Wright-Patterson AFB, 45433 Ohio, USA

Abstract

Self-organised quantum dots (QDs) are used successfully as active media of edge emitting and vertical cavity lasers. Threshold current densities at room temperature (RT) of 60 A/cm^2 for edge emitting lasers are measured in four side cleaved geometry. High internal (>96%) and differential (70%) efficiencies are obtained for InGaAs–AlGaAs lasers based on vertically coupled QDs. 1.5 W continuous wave operation at RT is demonstrated. Vertical cavity surface emitting lasers (VCSELs) based on QDs demonstrate threshold current densities down to 170 A/cm^2, comparable to the best values obtained for quantum well (QW) VCSELs, and total threshold currents as low as 67 µA. QD VCSELs demonstrate a reduction in the total current vs. aperture size down to 1 µm apertures due to reduced nonequilibrium carrier spreading out of the injection area as compared to QW VCSELs. QD lasers exhibit a much larger gain, differential gain and smaller line width enhancement factor as compared to conventional QW devices. These are clear inidcations that QD lasers will significantly overcome the performance of QW lasers in the future. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 42.55.Px; 68.55.−a; 68.65.+g; 81.05.Ea; 81.15.Hi; 78.60.Fi; 78.66.−w

Keywords: Quantum dot; Injection laser; Self-organized growth

1. Introduction

Significant improvements in characteristics of injection lasers have been foreseen for structures using quantum dots (QDs) as the gain medium [1]. Qunatum dots provide an ultimate limit of size quantization in solids and an extremely large change of electronic properties as compared to quantum wells and wires. The δ-function like density of states and the strong confinement of electron and hole wave functions in a QD result in the reduction of threshold current density, improved temperature stability of threshold current, ultrahigh material and differential gain of QD lasers [2]. On the other hand, until recently, there was a lack of fruitful ideas to fabricate QDs of a lateral size of

* Corresponding author. Tel.: +49 30 314 22783; fax: +49 30 314 22569; e-mail: bimb0431@mailszrz.zrz.physik.tu-berlin.de.

10–20 nm necessary for realistic RT devices. Previous ways to fabricate QDs by selective etching, or inter-mixing of quantum well structures suffered either from insufficient resolution, or from defects introduced by the patterning procedure, or both.

2. Self-organized growth of quantum dots

A solution of the problem of reliable quantum dot fabrication is based on an effect previously consid-ered as undesirable by crystal growers. A layer of a material having a lattice constant different from that of the substrate, after some critical thickness is de-posited, can spontaneously transform to an array of *three-dimensional* islands [3]. More recently it was shown, that there can exist a range of deposition pa-rameters, where the islands are small ($\sim$10 nm), have a similar size and shape [4,5] and form dense arrays. Theoretical considerations show that due to the *strain-induced renormalization* of the surface energy of the facets, an array of *equisized and equishaped* 3D is-lands can represent a *stable state* of the system [6]. This is directly confirmed for the most extensively studied MBE growth in the InAs–GaAs material sys-tem by the reversibility of the islanding – surface planarization process. The reversible transition occurs when the surface stoichiometry is changed from As-rich (islanding) to In-rich (planarization), and vice versa [7]. High arsenic pressure results in the reduction of the characteristic QD size and in the formation of high concentrations of dislocated InAs clusters. Thus, a stable array of 3D InAs islands on a GaAs (1 0 0) substrate exists only in some part of the "arsenic over-pressure – substrate temperature" phase diagram as is demonstrated also for MOCVD growth [8]. When the surface density of the QDs is high, interaction of the islands via the substrate also makes their *lateral or-dering* favourable [6]. If strained InAs islands are cov-ered with a thin GaAs layer, the islands in the second sheet are formed over the dots in the first sheet, re-sulting in a *three-dimensional ordered array* of QDs being either isolated [3,9] or strongly *vertically cou-pled* [3,7,10,11] resolved by transmission electron mi-croscopy and X-ray diffraction. The size and the shape of InAs islands can be changed by changing the de-position mode. Bright luminescence (300 K) from In-GaAlAs QDs in (Al,Ga)As matrrices can be turned in

the range 0.7–1.36 µm. Modification of the QD PL en-ergy is also possible via *post-growth annealing* [12].

3. Electronic spectrum

Experimentally observed energy levels for elec-trons and holes agree with theoretical calculations based solely on the QD geometry derived by HRTEM images. A remarkable property of a QD is a nar-row (FWHM < 0.1 meV) *atomic-like luminescence line* [13,14] which does not show broadening with temperature [14] up to 80 K and directly manifests formation of an electronic QD.

The transparency condition in a QD with a two-fold degenerate electron and hole ground state is realized when it captures one exciton. Then, the probability to emit or to absorb light is equal [15]. This statement is correct if the separation between the exciton and biexciton energy levels is smaller than the inhomo-geneous broadening. If this separation is larger a QD occupied by the exciton cannot absorb a photon hav-ing the exciton energy but will show stimulated emis-sion. Thus, in this case, which is, e.g. realized in II–VI materials, finite gain appears also for *single* excitons. Increase in the excitation density results in saturation of QD excitonic gain followed by its decrease, due to the larger fraction of QDs filled with biexcitons. On the other hand, biexciton absorption first increases and then drops to zero and converts to gain, when most of the QDs become populated with two electron–hole pairs. Charged excitons and biexcitons provide finite gain since ground-state absorption is no longer pos-sible. If there is no transport between QDs (particu-larly at low temperatures) the population of QDs with electrons, holes and excitons is defined by the capture and recombination probabilities and is not a function of temperature [15]. The gain-current relation depends on the capture mechanism [15], whether electrons and holes are captured in a correlated or uncorrelated man-ner. We note that the recombination current is *not* simply given by the product of electron and hole den-sity [16]. At high temperatures, thermal evaporation of carriers from dots may result in the predominant population of deeper QDs. Gain behavior is different in these cases: the gain maximum keeps its maximum energy with excitation density in the first case while it shifts to higher photon energies in the second [15].

4. Capture and relaxation of carriers in a quantum dot

Carrier capture efficiency is very high in properly designed structures with high concentration of QDs [17]. No degradation in PL efficiency at low temperatures and only moderate decrease at RT were found in deep etched mesa structures fabricated from samples with QDs when the mesa size was varied between 1 and 0.2 μm. On the contrary, no luminescence at RT can be recorded from deep mesas of similar size fabricated from quantum well samples. The nonequilibrium carrier capture time to a QD was estimated in Ref. [17] as being below 1 ps.

Carrier relaxation from an InAs–GaAs QD excited to the *ground state* was found to be about 40 ps for excitation in the GaAs mtarix and 25–40 ps for direct excitation in the QD excited state. In the latter case the lifetime of the exciton in the QD exited state and the population time of the QD ground state coincide [18]. Lower values are measured when the energy separation between the QD ground and excited state fits to an integer number of InAs LO phonons. These results are in agreement with PLE data also showing enhanced relaxation in this case [5,19]. The values measured agree favorably with the ground state population time in high quality QWs which are about 100 ps [20], while it is much longer than the higher sub-band de-population time ($\sim$1 ps) in the QW case, making QDs potentially important for application in IR lasers based on electron interlevel transitions.

5. Realization of quantum dot lasers

InGaAs–GaAs QD lasers operating under photopumped excitation both at low and at RT were fabricated in 1993 [21]. The first report on injection lasing already revealed the most important features of QD lasers [22]. At low observation temperatures very high characteristic temperatures ($T_0 > 400$ K) are observed. At high temperatures, however, evaporation of excitons from dots results in a decrease of QD gain for the same drive current, and to keep lasing conditions fulfilled (gain equals to total losses), one needs to increase the current.

Electroluminescence from QDs in edge geometry is significantly depolarised [7] due to hole quantization in all three dimensions, in contrast to QW structures, where heavy hole emission is completely TE polarized. Lasing resonant to the ground state (GS) of QDs and low threshold current densities (J_{th}) have been demonstrated in this range as well. At higher temperatures, evaporation of carriers from InGaAs QDs into a GaAs matrix resulted in a decrease in the QD gain, increase in J_{th}, and at 300 K, the lasing energy came closer to the wetting layer energy [23]. To overcome this problem, vertically electronically coupled QDs have been applied (see Fig. 1). This allowed reducing of the J_{th} both at low and at high temperatures down to 90 A/cm^{-2} (300 K, N=10) for four-side cleaved samples resulting in GS QD lasing up to room temperature (RT). Using MOCVD, ground state lasing for a single QD layer at RT was obtained for GaInP barriers [8]. Ultrahigh material gain [24] and suppression of gain saturation in coupled QDs [25] have been demonstrated. Further improvements were obtained by putting coupled QDs in an AlGaAs matrix, providing better localization of carriers in QDs. J_{th} was further reduced down to 60 A/cm^{-2} [9,18] for four-side cleaved samples and a RT CW lasing with 1 W has been realized for a ridge waveguide laser with 10 stacks of vertically coupled InGaAs–AlGaAs QDs (VCQDs) [26]. Even higher output power has been recently demonstrated for thrice stacked InAs–AlGaAs VCQDs [27]. In Fig. 2a we show the dependence of the threshold current density vs. reverse cavity length for this laser. Threshold current densities down to 180 A/cm^2 are realized at RT for sufficiently long (2 mm) devices. A characteristic feature of present QD lasers is the strong dependence of the lasing emission wavelength on cavity length which allows tuning of the lasing energy in a wide spectral range (see Fig. 2b) for the same structure. For long cavity length the lasing wavelength coincides with the maximum of the photoluminescence spectrum recorded from the top of the surface manifesting the QD ground state lasing. For smaller cavity lengths and higher external losses, higher injection current densities are required to maintain lasing. Higher role of excited states in this case causes the characteristic blue shift of the lasing energy.

In Fig. 3 we show light output vs. current dependence (left) and electroluminescence spectra at different drive currents. Maximum total output power of 1.5 W is realized with a differential efficiency of 55%. The two main effects now limiting the laser
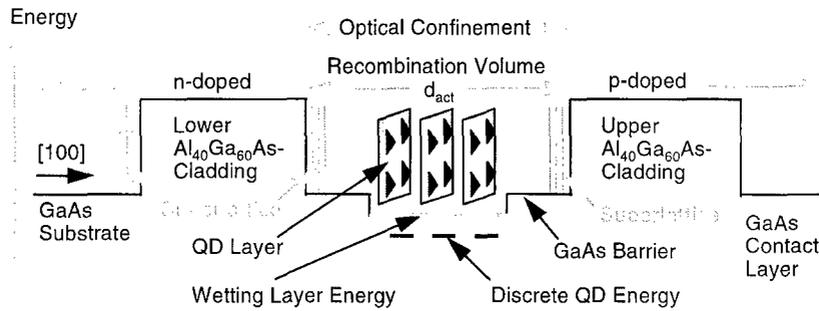
Fig. 1. Energy diagram under forward bias of an edge-emitting laser based on vertically stacked quantum dots.
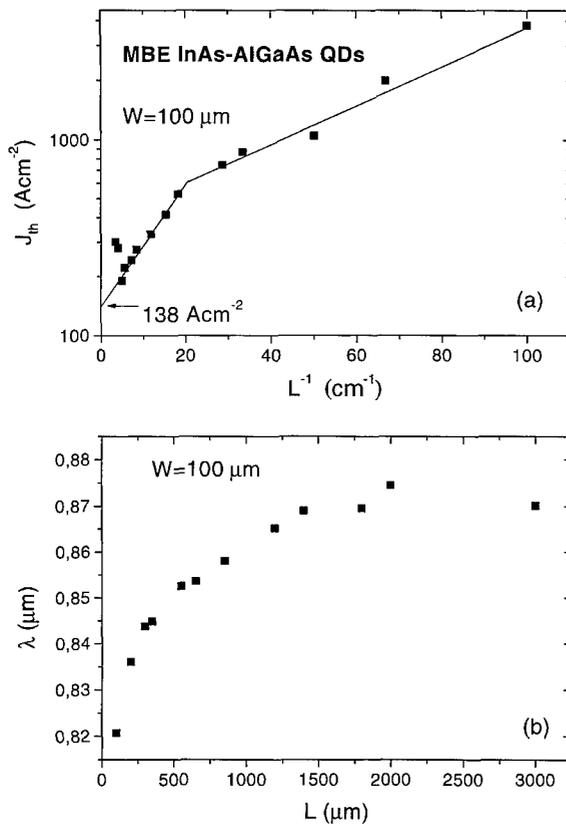

Fig. 2. (a) Threshold current density of an edge-emitting laser based on 3-period vertically coupled InAs–AlGaAs quantum dots vs. inverse cavity length. (b) Dependence of the lasing wavelength vs. cavity length. The Quantum dot laser provides a possibility of tuning of the emission wavelength in a broad spectral range (here about 100 meV) by selecting the cavity length.

be expected for lasers having lower series and contact resistance, using more efficient heat sinks. Higher internal ($>96\%$) and differential (70%) efficiencies at 300 K were recently obtained for InGaAs–AlGaAs lasers based on VCQDs [27]. High quantum efficiency is realized also in Ref. [28] indicating the competitiveness of edge-emitting QD lasers to the best QW analogues.

6. Basic parameters of quantum dot edge-emitting lasers

6.1. Edge-emitting lasers

Low-transparency currents and complete temperature insensitivity of the threshold current are the predicted basic advantages of QD l asers. For currently achieved 10% QD size dispersion the theoretical threshold current density [29] was found to be $10 \, \text{A/cm}^{-2}$ using a Master equation approach [15,16] and much lower values might be obtained in structures with improved QD homogeneity. J_{th} values of 12–$16 \, \text{A/cm}^2$ are measured for QD lasers in the temperature range up to 180 K [8].

Saturation material gain as high as $150\,000 \, \text{cm}^{-1}$ is reported in Ref. [24] as compared to QW values of about $3000 \, \text{cm}^{-1}$. This observation agrees well with calculated values [30].

Maximum differential gain reported for QD lasers approaches $10^{-12} \, \text{cm}^2$ at threshold. Under operating conditions it exceeds the QW laser values by about two orders of magnitude [24].

Chirp. Any absorption or gain peak causes modulation of the refractive index near the resonance energy

performance are related to laser heating and the differential efficiency. Much higher output powers can
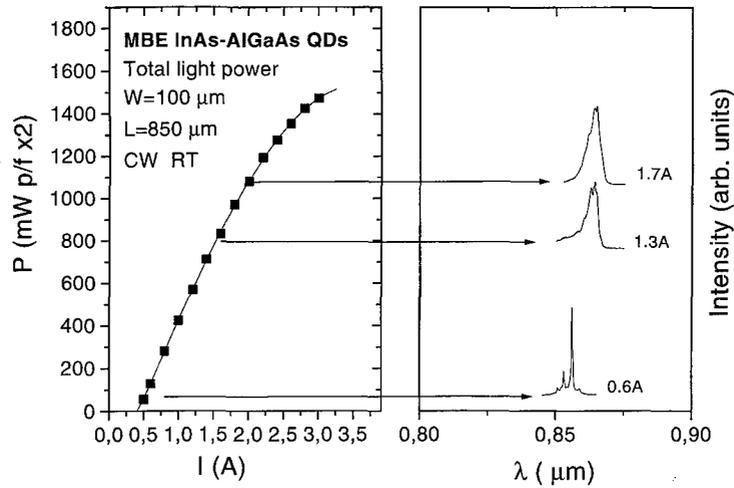
Fig. 3. Light output as a function of continuous wave drive current (left) and electroluminescence spectra at several drive currents of the laser based on InAs–AlGaAs quantum dots. The lasers are fabricated in the shallow mesa stripe geometry. The cavity length (L) and the stripe width (W) are 850 and 100 μm, respectively. The onset voltage on the $I–V$ curve is 1.2 V.

in agreement with Kramers–Kronig relation. The refractive index increases on the low energy side and decreases on the high energy side of the absorption maximum. A gain peak causes an opposite behavior. Thus, the photon wavelength in a crystal changes during the excitation pulse, causing lasing intensity jumps between Fabry–Perot modes, described by the line width enhancement factor (α). Highly asymmetric absorption and gain profiles in QWs, which also change their shape with excitation density, cause the α factor to be about two. For a symmetric QD gain (or absorption), the change in the refractive index near the gain maximum is zero and α is zero. In real QDs, due to the finite contribution of QD excited states to the total gain, α is presently found to be less than or about 0.5 [2].

Dynamics. From direct observation of relaxation oscillations cut-off frequencies close to 10 GHz have been determined [31,32].

Exciton waveguiding. The QD exciton absorption peak can be used intentionally to induce resonant refractive index enhancement on the low energy side of the absorption peak and can lead to entirely exciton-induced waveguiding and lasing [33]. This approach is particularly attractive in materials, where no suitable lattice-matched heterostructure with significantly lower refractive index exists, such as in some II–IV

compounds, diamond, silicon, etc. Physical and technical advantages of QD lasers are summarized in Tables 1 and 2, respectively.

7. Surface-emitting lasers

The first surface-emitting laser (VCSEL) operating via the QD GS was reported for photopumped excitation at low temperatures [34]. Injection VCSELs based on stacked InGaAs–AlGaAs QDs were realized later [35]. The lasing at 300 K occured via excited states and J_{th} was about $\sim$5000 A/cm^{-2}. Using an oxide defined 7 μm mesa and a single sheet InGaAlAs–GaAs QD structure, injection lasing via the QD GS at $\sim$1000 A/cm^{-2} (300 K) has been realized with a $\sim$0.5 mA total threshold current [36]. Using the novel concept of electronically -coupled InGaAs–GaAs QDs and oxide-defined 10 μm mesas, QD GS CW lasing with a total current of 180 μA has been recently manifested (180 A/cm^{-2}, 300 K) [37]. The structure is presented schematically in Fig. 4 . The active region consists of a GaAs cavity with a three-period InGaAs–GaAs VCQD structure. Maximum efficiencies amount to 16%. Dependence of the threshold current (upper part) and the threshold current density (lower part) vs. mesa size in our structures is shown in Fig. 5 by

Table 1
Comparison of physical parameters of QD and QW lasers quoting typical values

Laser parameter	Quantum well	Quantum dot
Threshold current density J_{th}	43 A/cm^2, 300 K (four-side cleaved) [40]	60 A/cm^2, 300 K (four-side cleaved) [18]
Theoretical temperature dependence (T_0) of J_{th}	285 K [1]	Infinity [1]
Saturation material gain	$\sim$3000 cm^{-1}	150 000 cm^{-1} (three-fold stacked)
Maximum differential gain	5×10^{-16} cm^2	$\sim 10^{-12}$ cm^2 (at threshold)
Lasing mechanism	e–h plasma or excitonic photon-assisted (wide gap)	Excitonic/biexcitonic (symmetric gain curve, low chirp, excitonic waveguide)
Ground state population time	20–100 ps	20–40 ps (single QD) <15 ps (coupled QDs)
Intersubband/interlevel relaxation time	1 ps	20–40 ps (single QD)

Table 2
Technological advantages of quantum dot lasers

Laser parameter	Quantum well	Quantum dot
Nonequilibrium carrier spreading out of the stripe (or cavity) region	Yes	No
Facet overheating due to surface recombination	Yes	No
Nonradiative recombination enhanced dislocation growth (due to carrier diffusion to dislocations)	Yes	Suppressed
Free-standing microstripes or microcavities	Not possible	Possible
1.3 μm at RT for GasAs/InGaAs/GaAs (1 0 0) structures	Not possible	Possible
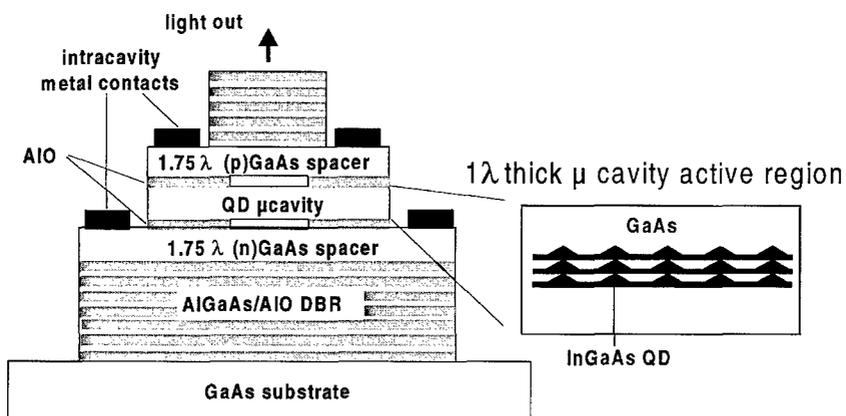
Fig. 4. Scheme of a vertical cavity surface emitting laser with vertically coupled quantum dots and oxide-defined current aperture. AℓO–AlGaAs Bragg reflectors are used as mirrors.
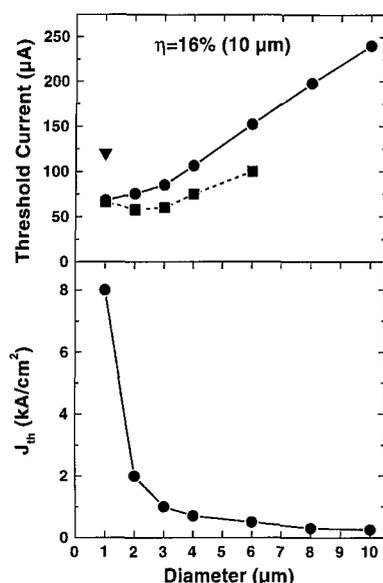
Fig. 5. Threshold current density (upper) and threshold current density vs. size of oxide-defined aperture. Original data are shown by closed circles (see text).

solid circles. The best data available in literature for the quantum well VCSELs of similar geometry are shown for comparison. Solid triangle is a result obtained in Ref. [38] and solid squares correspond to Ref. [39]. Minimum threshold currents of 68 μA are measured in our structures for 1 μm apertures. It is important to note that the threshold current in QD VCSELs decreases monotonically with mesa size, while QW VCSELs show an increase in threshold current below some minimum value. We relate this effect to the reduced spreading of nonequilibrium carriers in QD lasers out of the aperture region due to confinement of electrons and holes in QDs. The spreading effect should be even more suppressed for InGaAs QDs in an AlGaAs matrix, where sub-micron apertures become possible.

8. Conclusion

It look about seven years since the first realization of photopumped lasing in QWs until QW lasers with improved parameters with respect to double heterostructure lasers were created. The recent progress in the area of injection lasers based on self-organized QDs

leads to the expectation that QD lasers with properties improved as compared to best conventional QW lasers will be available soon. Tables 1 and 2 summarizes the basic characteristics of QD lasers in comparison with QW lasers.

Further progress in the area of QD lasers is related to further optimization of the structure geometry and growth conditions.

References

[1] Y. Arakawa, H. Sakaki, Appl. Phys. Lett. 40 (1982) 939; M. Asada, Y. Miyamoto, Y. Suematsu, IEEE J. Quantum. Electron. QE-22 (1986) 1915.
[2] D. Bimberg, N.N. Ledentsov, N. Kirstaedter, O. Schmidt, M. Grundmann, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, S.S. Ruvimov, U. Gösele, J. Heydenreich, Jpn. J. Appl. Phys. 35 (1996) 1311; D. Bimberg, N. Kristaedter, N.N. Ledentsov, Zh.I. Alferov, P.S. Kop'ev, V.M. Ustinov, IEEE J. Selected Topics Quantum Electron. 3 (1997) 1.
[3] L. Goldstein, F. Glas, J.Y. Marzin, M.N. Charasse, G. Le Roux, Appl. Phys. Lett. 47 (1985) 1099.
[4] J.M. Moison, F. Houzay, F. Barthe, L. Leprice, E. Andre, O. Vatel, Appl. Phys. Lett. 64 (1994) 196; D. Leonard, M. Krishnamurthy, C.M. Reaves, S.P. Denbaars, P.M. Petroff, Appl. Phys. Lett. 63 (1993) 3203; D. Bimberg, M. Grundmann, N.N. Ledentsov, S.S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V.M. Ustinov, P.S. Kop'ev, Zh.I. Alferov, Thin Solid Films 267 (1995) 32.
[5] N.N. Ledentsov, M. Grundmann, N. Kirstaedter, J. Christen, R. Heitz, J. Böhrer, F. Heinrichsdorff, D. Bimberg, S.S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V.M. Ustinov, A.Yu. Egorov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, Proc. ICPS-22, Vancouver, 1944, vol. 3, World Scientific, Singapore, 1995, p. 1855.
[6] V.A. Shchukin, N.N. Ledentsov, P.S. Kop'ev, D. Bimberg, phys. Rev. Lett. 75 (1995) 2968; Surf. Sci. 352 (1996) 117.
[7] N.N. Ledentsov, M. Grundmann, N. Kirstaedter, O. Schmidt, R. Heitz, J. Böhrer, D. Bimberg, V.M. Ustinov, V.A. Shchukin, P.S. Kop'ev, Zh.I. Alferov, S.S. Ruvimov, A.O. Kosogov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, Solid State Electron. 40 (1996) 785.
[8] F. Heinrichsdorff, M.-H. Mao, N. Kirstaedter, A. Krost, D. Bimberg, A.O. Kosogov, P. Werner, Appl. Phys. Lett. 71 (1997) 22.

[9] Q. Xie, A. Madhukar, P. Chen, N.P. Kobayashi, Phys. Rev. Lett. 75 (1995) 2542.

[10] N.N. Ledenstov, V.A. Shchukin, M. Grundmann, N. Kirstaedter, J. Böhrer, O. Schmidt, D. Bimberg, S.V. Zaitsev, V.M. Ustinov, A.E. Zhukov, P.S. Kop'ev, Zh.I. Alferov, A.O. Kosogov, S.S. Ruvimov, P. Werner, U. Gösele, J. Heydenreich, Phys. Rev. B 54 (1996) 8743; G.S. Solomon, J.A. Trezza, A.F. Marshall, J.S. Harris, Phys. Rev. Lett. 76 (1996) 952.

[11] A.A. Darhuber, V. Holy, J. Stangl, G. Bauer, A. Krost, F. Heinrichsdorff, M. Grundmann, D. Bimberg, V.M. Ustinov, P.S. Kop'ev, A.O. Kosogov, P. Werner, Appl. Phys. Lett. 70 (1997) 955; A.A. Darhuber, V. Holy, J. Stangl, G. Bauer, A. Krost, M. Grundmann, D. Bimberg, V.M. Ustinov, P.S. Kop'ev, A.O. Kosogov, P. Werner, Jpn. J. Appl. Phys. 36 (1997) 4084.

[12] A.O. Kosogov, P. Werner, U. Gösele, N.N. Ledentsov, D. Bimberg, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, P.S. Kop'ev, N.A. Bert, Zh.I. Alferov, Appl. Phys. Lett. 69 (1996) 3072; R. Leon, Yong Kim, C. Jagadish, M. Gal, J. Zou, D.H. Cockayne, Appl. Phys. Lett. 69 (1996) 1888.

[13] J.-Y. Marzin, J.M. Gerard, A. Izraël, D. Barrier, G. Bastard, Phys. Rev. Lett. 73 (1994) 716.

[14] M. Grundmann, J. Christen, N.N. Ledentsov, J. Böhrer, D. Bimberg, S.S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, P.S. Kop'ev, Zh.I. Alferov, Phys. Rev. Lett. 74 (1995) 4043.

[15] M. Grundmann, D. Bimberg, Jpn. J. Appl. Phys. 36 (1997) 4181.

[16] M. Grundmann, D. Bimberg, Phys. Rev. B 55 (1997) 9740.

[17] N.N. Ledentsov, M.V. Maximov, P.S. Kop'ev, V.M. Ustinov, M.V. Belousov, B.Ya. Meltser, S.V. Ivanov, V.A. Shchukin, Zh.I. Alferov, M. Grundmann, D. Bimberg, S.S. Ruvimov, W. Richter, P. Werner, U. Gösele, U. Heidenreich, P.D. Wang, C.M. Sotomayor Torres, Microelectronics J. 26 (1995) 871.

[18] N.N. Ledentsov, Proc. ICPS23, Berlin 1996, p. 19 and references therein.

[19] R. Heitz, M. Grundmann, N.N. Ledentsov, L. Eckey, M. Veit, D. Bimberg, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, P.S. Kop'ev, Zh.I. Alferov, Appl. Phys. Lett. 68 (1996) 361.

[20] B. Devaud, F. Clérot, N. Roy, K. Satzke, B. Sermage, D.S. Katzer, Phys. Rev. Lett. 67 (1991) 2355.

[21] N.N. Ledentsov, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, M.V. Maximov, I.G. Tabatadze, P.S. Kop'ev, Semiconductors 28 (1994) 832.

[22] N. Kirstaedter, N.N. Ledentsov, M. Grundmann, D. Bimberg, U. Richter, S.S. Ruvimov, P. Werner, J. Heydenreich, V.M. Ustinov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, Electronics Lett. 30 (1994) 1416.

[23] K. Kamath, P. Bhattacharya, T. Sosnowski, T. Norris, J. Phillips, Electron. Lett. 32 (1996) 1374.

[24] N. Kirstaedter, O.G. Schmidt, N.N. Ledentsov, D. Bimberg, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, Appl. Phys. Lett. 69 (1996) 1226.

[25] O.G. Schmidt, N. Kirstaedter, N.N. Ledentsov, M.-H. Mao, D. Bimberg, V.M. Ustinov, A.Y. Egorov, A.E. Zhukov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, Electron. Lett. 32 (1996) 1302.

[26] Yu.M. Shernyakov, A.Yu. Egorov, A.E. Zhukov, S.V. Zaitsev, A.R. Kovsh, I.L. Krestnikov, A.V. Lunev, N.N. Ledentsov, M.V. Maximov, A.V. Sakharov, V.M. Ustinov, Zhao Zhen, P.S. Kop'ev, Zh.I. Alferov, D. Bimberg, Pis'ma v Zh. Tekhn. Fiz. 23 (1) (1997) 51.

[27] M.V. Maximov et al., to be published.

[28] R. Mirin, A. Gossard, J. Bowers, Electron. Lett. 32 (1996) 1732.

[29] L.V. Asryan, R.A. Suris, Semicond. Sci. Technol. 11 (1996) 1.

[30] Y. Arakawa, A. Yariv, IEEE J. Quantum. Electron. QE22. (1986) 1887.

[31] K. Kamath, J. Phillips, H. Jiang, J. Singh, P. Bhattacharya, Appl. Phys. Lett. 70 (1997) 2952.

[32] M.-H. Mao, F. Heinrichsdorff, A. Krost, D. Bimberg, Electron. Lett. 33 (1997) 1641.

[33] N.N. Ledentsov, I.L. Krestnikov, M.V. Maximov, S.V. Ivanov, S.L. Sorokin, P.S. Kop'ev, Zh.I. Alferov, D. Bimberg, C.M. Sotomayor Torres, Appl. Phys. Lett. 69 (1996) 1343; N.N. Ledentsov, I.L. Krestnikov, M.V. Maximov, S.V. Ivanov, S.L. Sorokin, P.S. Kop'ev, Zh.I. Alferov, D. Bimberg, C.M. Sotomayor Torres, Appl. Phys. Lett. 70 (1997) 2766; M. Straßburg, V. Kutzer, U.W. Pohl, A. Hoffmann, I. Broser, N.N. Ledentsov, D. Bimberg, A. Rosenauer, U. Fischer, D. Gerthsen, I.L. Krestnikov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, Appl. Phys. Lett. (1998) (in press).

[34] R. Schur, F. Sogawa, M. Nishioka, S. Ishida, Y. Arakawa, Jpn. J. Appl. Phys. 35 (1997) L357.

[35] H. Saito, K. Nishi, I. Ogura, S. Sugou, Y. Sugimoto, Appl. Phys. Lett. 69 (1996) 3140.

[36] D.L. Huffaker, O. Baklenov, L.A. Graham, B.G. Streetman, D.G. Deppe, Appl. Phys. Lett. 70 (1997) 2356.

[37] J.A. Lott, N.N. Ledentsov, V.M. Ustinov, A.Yu. Egorov, A.E. Zhukov, P.S. Kop'ev, Zh.I. Alferov, D. Bimberg, Electron. Lett. 33 (1997) 1150.

[38] T.H. Oh, D.L. Huffaker, D.G. Deppe, Appl. Phys. Lett. 69 (1996) 3152.

[39] D.L. Huffaker, D.G. Deppe, Appl. Phys. Lett. 70 (1997) 1781.

[40] Zh.I. Alferov, S.V. Ivanov, P.S. Kop'ev, N.N. Ledentsov, M.E. Lutsenko, M.I. Nemenov, B.Ya. Melster, V.M. Ustinov, S.V. Shaposhnikov, Fizika i Tekn. Poluprovodn. 24 (1990) 152; Sov. Phys. Semicond. 24 (1990) 92.

Quantum transport in ballistic quantum dots

D.K. Ferry *, R.A. Akis, D.P. Pivin Jr., J.P. Bird, N. Holmberg,
F. Badrieh, D. Vasileska

*Center for Solid State Electronics Research and Department of Electrical Engineering, Arizona State University,
Tempe, AZ 85287-5706, USA*

Abstract

Carriers in small 3D quantum boxes take us from unintentional quantum dots in MOSFETs (arising from the doping fluctuations) to single-electron quantum dots in semiconductor heterostructures. In between these two extremes are the realm of open, *ballistic* quantum dots, in which the transport can be quite regular. Several issues must be considered in treating the transport in these dots, among which are: (1) phase coherence within the dot; (2) the transition between semi-classical and fully quantum transport, (3) the role of the contacts, vis-à-vis the fabricated boundaries, and (4) the actual versus internal boundaries. In this paper, we discuss these issues, including the primary observables in experiment, the intrinsic nature of oscillatory behavior in magnetic field and dot size, and the connection to semi-classical transport emphasizing the importance of the filtering by the input (and output) quantum point contacts. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Quantum transport; Ballistic quantum dot

1. Introduction

Recently, a class of open systems, the *ballistic quantum dots*, has come to be studied for the manner in which regular versus chaotic behavior can be evident in the conductance properties [1]. Unfortunately, the rush to see chaos in every natural phenomena has cast a cloud over the relevant understanding of the transport behavior. Instead, these systems are properly understood through a careful application of the Landauer formula and the concepts of semi-classical trajectories. Indeed, the variation in the *total trans-mission* that is provided by the input and output quantum points contacts dramatically affects the overall conductance through the quantum dot. The fact that these contacts are actually *quantum* point contacts, with modes, leads to collimation of the input beam, dramatically affecting the transport through the dot [2,3]. In this paper, we will attempt to present a view of the experimental and theoretical understanding of the transport, both classical and quantum, through these ballistic quantum dots, concentrating on those with square dimension, although this behavior carries over to more generic shapes.

In general, these regular structures may be viewed quite reasonably by the extension of many years of studies of regular (and chaotic) behavior, in which the

* Corresponding author. Tel.: +1 602 9652570; fax: +1 602 9658058; e-mail: ferry@asu.edu.

conductance of the system is dominated by the prop-
erties of the density of states at the Fermi energy. The
latter is defined by the expansion of the delta function
density of states in the closed dot into a summation
over the closed semi-classical trajectories within the
dot. Experimentally, the square ballistic quantum dots
are characterized by a magnetoresistance (or magneto-
conductance) which exhibits regular, periodic fluctu-
ations. The correlation function of these fluctuations
consequently also exhibits periodic oscillations, and
these, along with the fluctuations themselves, are be-
lieved to be endemic of the density of states in the
quantum dot. However, the fact that the dot is ex-
cited by quantum point contacts means that the prop-
erties of the latter *filter* the total density of states
and will greatly affect the transport properties of the
dot.

In the next section, we review the essentials of
the experimental observations, and relate these to the
properties of the quantum point contacts. Then, a dis-
cussion of the periodic properties of the dots, and how
this affects the semi-classical and regular behavior,
will be presented. Finally, we extend this to discuss
precisely the importance of the environment in which
the dot is embedded and how it will affect the observed
conductance behavior. We then turn to scaling of the
oscillations with dot size, and estimate the connection
between the periodicity and the dominant trajectories.
In this way, it is also pointed out that the common
approaches to magnetic behavior in the semi-classical
world are incorrect and misleading.

2. The experimental observations

A quantum dot is a confined region of a semicon-
ductor (we deal only with semiconductors here), in
which the dynamics is quantized; that is, the motion
is essentially constrained in all three dimensions. In
the inset of Fig. 1, we illustrate an electrostatically de-
fined dot, with the normal direction quantized by the
heterostructure interface. That is, the basic material
structure is a GaAs/AlGaAs heterostructure in which
a quasi-two-dimensional electron gas forms at the in-
terface (on the GaAs side). This is then confined fur-
ther, as indicated in Fig. 1 (inset) by metal Schottky
gates, which deplete the region under the metal. This
leaves the electrons in the open area between the two
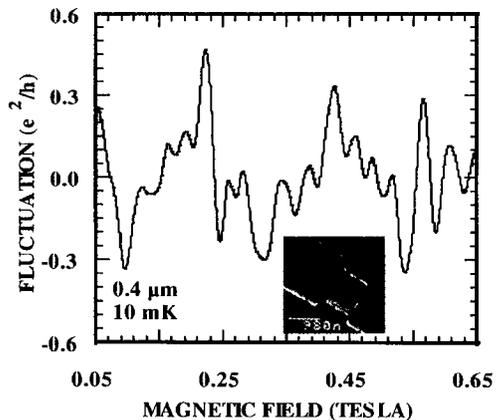
Fig. 1. Experimental fluctuations observed in the magnetoconduc-
tance of a GaAlAs/GaAs dot. The inset is a view of the device
itself.

quantum point contacts (fingers in the image). In this
regard, the eigenvalue spectrum of the *closed* dot at
low temperature (typically a few mK) is a series of
δ-functions located in energy.

In closed quantum dots, it has been possible to
study the energy level structure, and therefore the os-
cillating density of states, by careful measurements
of the tunneling current in a single-electron tunneling
experiment [4,5]. In open quantum dots, it has been
suggested that the discrete nature of the spectrum is
washed out by interactions with the environment [6].
However, it must be remembered that these latter au-
thors wished to study the chaos in the dot, and a dis-
crete spectrum would have made chaos quite unlikely.
If one assumes Neumann boundary conditions on the
dot rather than Dirichlet, so that the wave functions
can penetrate easily into the area outside the dot, the
mean level separation is only slightly modified from
that of the closed dot, at least in the Thomas–Fermi
approximation [7]. This clearly means that merely al-
lowing the wave functions to connect to the outside
world does not eliminate the discretized levels nor
the fluctuations in the density of states. One must in-
stead carefully study the manner in which the dot
states are perturbed by their interaction with the en-
vironment states [8,9], an issue to which we return
below. In Fig. 1, we show the magnetoconductance
fluctuations in a dot like that illustrated in the inset.
These oscillations have been studied for sizes ranging
from 2.0 μm down to 0.4 μm (the electrical sizes are
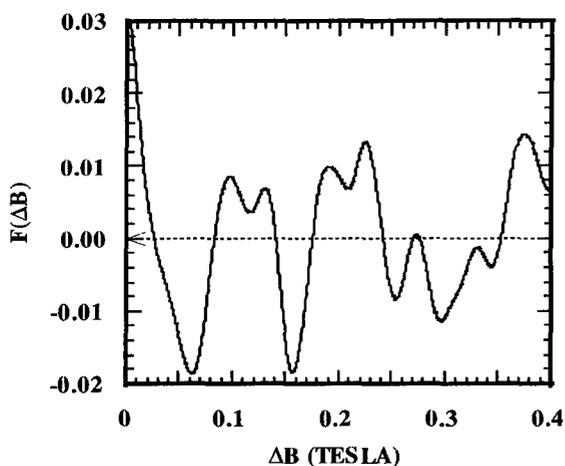
Fig. 2. The correlation function of the fluctuations shown in Fig. 1 above.

Fig. 3. Computation of the conductance through a wire (a) and a wire with a dot (b).

somewhat smaller than the physical size due to depletion of carriers around the edges of the gates). They have been studied in GaAs/AlGaAs heterostructures with both Schottky gates and in-plane gates, and in InGaAs/GaAs strained heterostructures with in-plane gates [10]. The gate design in each case was such as to trap the carriers within the dots, although the quantum point contacts were not closed off and had 1–5 modes propagating through them. Note particularly the position of the openings in the inset, as this is crucial to the excitation of the dots, as will be discussed below.

The data in Fig. 1 is for a GaAs/AlGaAs dot with Schottky gates, and a temperature of 10 mK. Reproducible oscillations persist across a wide range of magnetic field. In addition, the correlation function for the fluctuations is plotted in Fig. 2; the latter is averaged over a range of initial magnetic fields. It is clear that the basic nature of the reproducible fluctuations is very oscillatory, as represented also in the correlation function. These basic oscillatory properties are found to be quite universal, although the specific frequency content is very dot dependent. It will become clear that these magnetoconductance oscillations arise from the fluctuations in the density of states of the dot, which in turn are a result of the quasi-periodic orbits of the carriers inside the quantum dot, and are therefore a reflection of the intrinsic properties of the dots themselves.

3. Fluctuations and the density of states

We can examine the source of the fluctuations by simulating the transport through the quantum dot. This is done utilizing the recursive Green's function and is shown in Fig. 3. Here, curve (a) represents transport through a quantum wire with no barriers (other than hard wall boundary conditions), while curve (b) results when the quantum dot is created between two barriers. Here, the opening in the resulting two quantum point contacts is the same as in the wire of (a). The steps in (a) are replicated in (b), albeit with a shift and significant superimposed oscillations (which are seen in experiment as well [11]). The shift is a result of the fact that the quantum point contacts are not isolated wires, and are affected by the dot presence. The oscillations arise from the density of states within the dot, which result in "trapping" resonances, hence the apparent drop in conductance. Clearly, these are not washed out with the open dot ($\geqslant 1$ mode propagating through the structure), and we can use this fact to begin to understand the source of the oscillations/fluctuations. Previously, it has been supposed that the conductance is related to trajectories that connect the input and output ports [6]. In fact, this is an incomplete description of what is important. The conductance is easily shown to be an integral over the density of states at the Fermi energy, and this density of states is defined by *closed* orbits within the dot. When the dot is opened, it is those closed orbits which are almost closed, but which connect to the ports that

are responsible for the transport. It is important to note that these must be derived from the closed orbits of the closed dots and make a large, but finite, number of orbits around the trajectory before exiting; that is, these trajectories return to a starting point, as has been conjectured in much semi-classical transport theory [7].

The connection between the density of states and semi-classical trajectories is easily obtained through semi-classical quantum mechanics. If the δ-function is replaced by its Fourier representation (in energy space) via the Poisson summation formula [7]

$$\sum_n f(n) = \sum_{M=-\infty}^{\infty} \int_0^{\infty} f(n)e^{i2\pi nM}\, dn + \frac{1}{2} f(0), \quad (1)$$

then the integrals can be evaluated by the saddle-point method, and the density of states can be expanded into the form [7]

$$\rho(E) = \frac{m^* a^2}{2\pi\hbar^2} \sum_{M_1, M_2 = -\infty}^{\infty} J_0\left(\frac{1}{\hbar} S_{M_1, M_2}\right)$$
$$- \frac{a}{4\pi\hbar} \sqrt{\frac{2m^*}{E}} \sum_{M=-\infty}^{\infty} \cos\left[\frac{2Ma}{\hbar}\sqrt{2m^*E}\right] \quad (2)$$

for a two-dimensional square dot of side a. Here, M_1 and M_2 are integers defining the number of basic repetitions of a which are in the total trajectory length while S is the action integral over the trajectory. *What is often missed is that this integral must be quantized in the dot, and hence only certain values of the energy (momentum) are allowed.* As a result, only some terms contribute at each energy E. *It is this quantization of the energy that will allow a particular term in Eq. (2) to recur as the energy is varied* (or as the magnetic field is varied). It has been suggested that one can introduce the magnetic field variation by multiplying each term in Eq. (2) by a factor given by [12]

$$\cos\left(\frac{n\Phi}{2}\right), \qquad \Phi = \frac{ea^2 B}{M_1 M_2 \hbar}, \quad (3)$$

but *this is incorrect*. The problem with this formulation may be seen from Fig. 4. The two closed trajectories, labeled (a) and (b), both have $M_1 = M_2 = 1$, and therefore have the same length around the orbit. Hence, their action integral is the same in the absence of a magnetic field, and they are degenerate at the same energy level. However, *these two orbits enclose different areas*, so that their behavior in
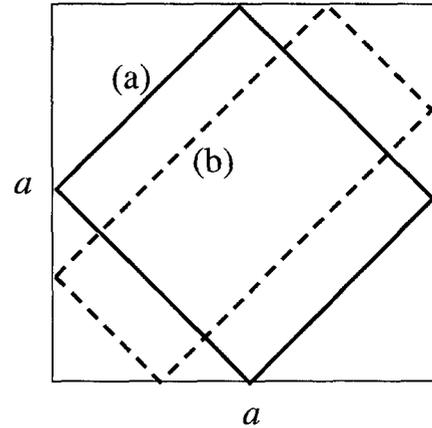
Fig. 4. Two trajectories embedded in a dot of side a. The trajectories (a) and (b) both have the same length and action integral in the absence of a magnetic field. However, they have different areas and will respond differently in a magnetic field.
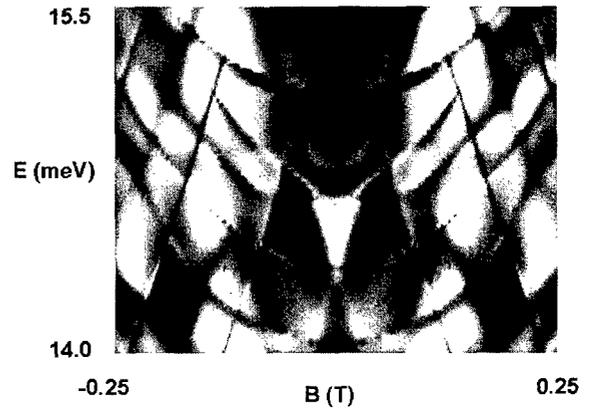
Fig. 5. The calculated conductance through a 0.3 μm dot. The lighter areas represent higher conductance.

a magnetic field will be quite different. In essence, the magnetic field raises the degeneracy of the levels, and (for a circular dot) the different areas correspond to different angular momentum in the orbit. In the square, this degeneracy is still raised, but the results are not thought to be pure angular momentum states. Eq. (3) defines the magnetic dependence only through M_1 and M_2, but clearly another quantum number is required to specify the area or something like the angular momentum.

In Fig. 5, we plot the conductance calculated for an open square dot with four propagating modes in the quantum point contact leads. Here, the conductance is computed with a stabilized version of the transfer matrix approach and hard wall potentials. We have also coupled this with a real potential determined from full Schödinger/Poisson solutions of the actual structure, with essentially similar results. The lighter areas correspond to higher values of the conductance, while darker areas correspond to higher values of the resistance. There are structured lines in the conductance which are reminiscent of the energy levels in the Darwin–Fock spectra [13]. Moreover, it is clear that the magnetic field indeed splits degeneracies which exist in the absence of this field. The diagonal lines seem to be uniformly separated, and provide direct evidence for "closed" orbits. These lines cross at zero magnetic field, which corresponds to the degeneracy discussed in the previous paragraph, and the lifting of this degeneracy by the field leads to the diverging lines. The periodicity is also thought to be the cause of the periodic scars that are observed in the wave functions for these structures [2,3] (we note that, that contrary to many discussions in the chaos literature, scars are generally thought to arise from residuals of the classical regular orbits, which is certainly the case in these dots). One might at first glance be led to believe that these highly structured conductances are unique to the square geometry studied here, but they have been found in stadium structures [14] and in more re-entrant structures where careful coupling of the Schödinger/Poisson solution derived potential with the quantum simulations reproduce the experimentally observed behavior [15].

4. Modulating the spectrum

We have now established that the spectrum of the oscillations arises from the *partial* density of states that is excited by the quantum point contacts and their collimation of the incoming beam of particles. This can be carried further by varying the excitation of each of the modes passing through the quantum point contacts. In Fig. 6, we show a dot that is embedded in a special gated structure so that each quantum point contact is excited by a quantum wire, whose modes
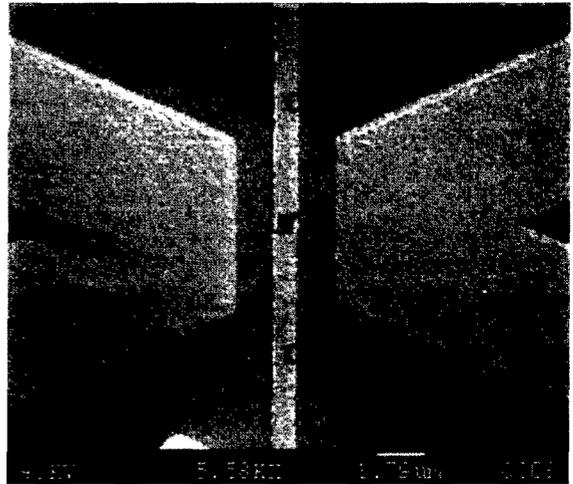
Fig. 6. Photomicrograph of a dot embedded in a pair of quantum wires. The various gates and leads are labeled accordingly.
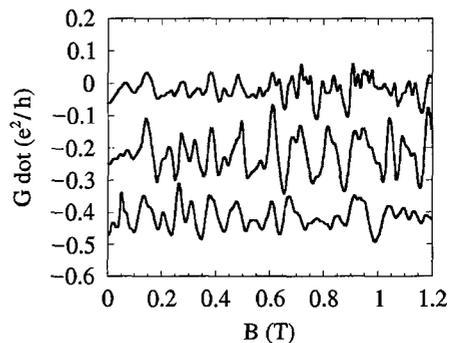
Fig. 7. Conductance fluctuations as the wire is biased toward fewer modes. The upper curve is for the most open wire, while the lower curve has only 1–2 modes in the wire.

propagate at right angles to those of the contacts [16]. Hence, varying the number of modes in the wire will dramatically affect the coupling to the individual modes in the quantum point contacts and thereby affect the conductance oscillations themselves. In Fig. 7, we demonstrate the variations in the conductance as the number of wire modes is reduced. The number of modes in the quantum point contact is 1–2, based upon measurements when the wire is not formed. As the wire-defining gate is biased more negative, we see that the fluctuations vary significantly.

The coupling of the modes in the quantum point contact to the dot can be seen more graphically by
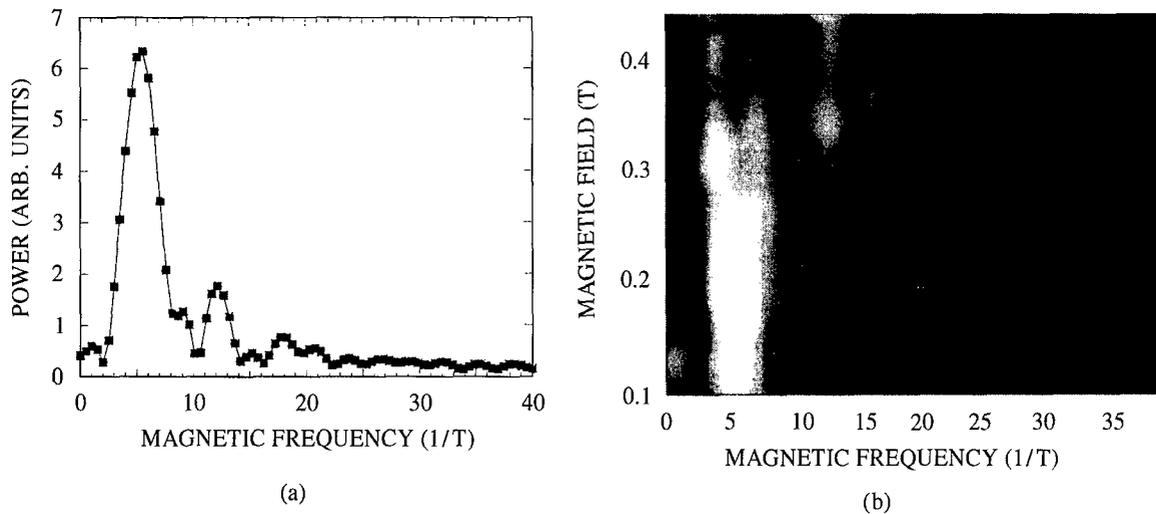
Fig. 8. (a) A calculated Fourier spectrum for a 0.2 μm (actual) dot. (b) The summarized spectrum for a range of magnetic fields. These are discussed in the text.

looking at the Fourier transform of the conductance autocorrelation function. As can be inferred from Eq. (2), different magnetic frequencies correspond to different trajectories, and hence to different effective areas enclosed by these trajectories. While it is quite difficult to determine exactly which trajectories correspond to each of the magnetic frequencies, one can at least study the manner in which these trajectories are excited by the dots. In Fig. 8a, we plot one such (theoretical) Fourier spectra, and in Fig. 8b, we plot the family of these spectra: the horizontal axis is the magnetic frequency, while the vertical axis is the magnetic field at which the correlation function is taken. Each plot such as Fig. 8a corresponds to one line scan horizontally through Fig. 8b.

In Fig. 9, we plot the spectra for our wire-excited dot, for the three curves of Fig. 7. The number in each panel corresponds to the relative amplitude to which "white" corresponds. It may be clearly seen that as the wire is pinched down, the first effect is that fewer dot modes seem to be excited, but these are much more strongly excited. However, as the number of the modes in the wire is reduced, these have great difficulty in coupling to the quantum point contacts, and hence to the dot itself, which is signified by the very weak amplitude of the Fourier peaks in the third panel of the figure. In the top panel, we see that there are several frequencies which have signifi-

cant amplitude. As the wire becomes pinched down, however, only one dominant mode remains strongly excited, and its amplitude has increased five-fold. Finally, for the smallest wire (most negative bias), the coupling to the dot is hindered, and the dominant mode is broken up into many harmonics, which could well be characteristic of heavy damping of this mode.

This behavior can be seen in the data of Fig. 7 as well. However, in both of these figures, some care must be taken, as the dot size is about 0.5 μm (after edge depletion) and $r_c < a$ for $B > 0.2$ T, and we expect the onset of Shubnikov-de Haas oscillations at the larger values of field in the figures. Hence, we should concentrate on the lower field values. It appears that the upper trace in Fig. 7 has significant harmonic content, which is revealed in the upper panel of Fig. 9, and this harmonic content is reduced in the center curve and panel. It should be pointed out, however, that we do not fully understand what these figures are telling us, and our current interpretation is quite tentative. For example, it appears in Fig. 8 that the dominant peak frequency actually splits at higher magnetic field (above 0.3 T), and there is a strange shift in all other peaks to higher frequency. It is possible that this is a further splitting of two different angular momentum states, but this is quite speculative at this point.

Fig. 10. The phase breaking time found in several different dots of varying size as the quantum point contacts are progressively closed (leading to higher resistance). The indicated dot sizes are in microns.

Fig. 9. Variation in the Fourier spectra as the wire bias is made more negative (top to bottom). Each panel corresponds to a trace in Fig. 7. These are discussed in the text.

5. Phase breaking

A crucial point in all of this is the need to be sure that the phase coherence time is sufficiently long that the closed orbits, and the density of states, can be adequately excited. In Fig. 10, we illustrate the phase breaking time computed in several different dots as a function of the opening of the quantum point contacts (actually as a function of the total series resistance) [17]. The phase breaking time itself was determined from studies of the conductance fluctuations in the edge state regime at high magnetic fields [18]. It may be seen that the phase breaking time increases by a significant amount as the leads are closed down to

only a few modes. In this latter case, the carriers are trapped in the dot significantly longer, and this leads to their contributing more to the closed orbits that are excited by the entry collimation. It should be noted that the results found here agree well with simulations that probed the required phase breaking time necessary to see well-defined scars in the wave functions within the dot itself and with temperature dependent studies in which it was found that the dot states become well-resolved as 0D states below a critical temperature [19]. It is assumed that this transition to 0D occurs when the temperature drops below the mean level separation, and this has been confirmed with studies of bias in the dots [20].

6. Discussion

While there have been a great many studies of the transport in quantum dots, particularly in the single-electron regime, it appears that assertions that open dots will be completely chaotic are quite wrong in their interpretations. Indeed, the appearance of discrete states, and their effect on the density of states fully carry over to open dots, when the latter are excited from quantum point contacts supporting only a few modes. Here, the magnetoconductance oscillations, as well as the gate voltage-induced conductance oscillations, are reflections of this underlying structure in the density of states. Indeed, the interpretation that the transport is represented by *simple* trajectories from in-

put to output, which fully sample all the dot phase space, are improper interpretations, with the fundamental properties arising from the *closed* trajectories in the dot. While the studies presented here have focused upon the square dot, they are quite generic and the same behavior has been found in dots with other shapes.

In this sense, then, the measured conductance fluctuations that arise as the gate voltage or the magnetic field are varied, arise from the actual density of states within the dot. The latter is determined by the energy spectrum within the dot and this is not washed out by opening the dot, although it is modified to some degree through the interaction between states within the dot and those in the quasi-two-dimensional electron gas outside the dot. This interaction is itself moderated by the quantum point contacts. In the sense of the Landauer formula, it is therefore quite important to compute the transmission through the entire, real structure which includes considerations of the quasi-two-dimensional gas, the quantum point contacts, *and* the dot states [9,21]. Only when the entire system is properly considered can the real transport properties be determined.

References

[1] J.P. Bird, R. Akis, D.K. Ferry, D.P. Pivin Jr., K.M. Connolly, R.P. Taylor, R. Newbury, D.M. Olatona, A. Micolich, R. Wirtz, Y. Ochiai, Y. Okubo, K. Ishibashi, Y. Aoyagi, T. Sugano, Chaos, Solitons, and Fractals 8 (1997) 1299.

[2] R. Akis, J.P. Bird, D.K. Ferry, J. Phys.: Condens. Matter 8 (1996) L667.

[3] R. Akis, D.K. Ferry, J.P. Bird, Phys. Rev. B 54 (1996) 17705.

[4] P.L. McEuen, E.B. Foxman, U. Meirav, M.A. Kastner, Y. Meir, N.S. Wingreen, S.J. Wind, Phys. Rev. Lett. 66 (1991) 1926.

[5] S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, L.P. Kouwenhoven, Phys. Rev. Lett. 77 (1996) 3613.

[6] R.A. Jalabert, H.U. Baranger, A.D. Stone, Phys. Rev. Lett. 65 (1990) 2442; 70 (1993) 3876.

[7] M. Brack, R.K. Bhaduri, Semiclassical Physics, Section 2.7, Addison-Wesley, Reading, MA, 1997.

[8] J.R. Barker, D.K. Ferry, Solid-State Electron. 23 (1980) 519.

[9] D.K. Ferry, J.P. Bird, R. Akis, D.P. Pivin Jr., K.M. Connolly, K. Ishibashi, Y. Aoyagi, T. Sugano, Y. Ochiai, Jpn. J. Appl. Phys. 36 (1997) 3944.

[10] K.M. Connolly, D.P. Pivin Jr., D.K. Ferry, H.H. Wieder, Superlatt. Microstruc. 20 (1996) 307.

[11] J.P. Bird, R. Akis, D.K. Ferry, J. Cooper, K. Ishibashi, Y. Ochiai, Y. Aoyagi, T. Sugano, Semicond. Sci. Technol. 13 (1998) A4.

[12] D. Ullmo, K. Richter, R.A. Jalabert, Phys. Rev. Lett. 74 (1995) 383.

[13] D.K. Ferry, Quantum Mechanics, Inst. of Phys. Pub., Bristol, 1995, p. 208.

[14] R. Akis, D.K. Ferry, J.P. Bird, Phys. Rev. Lett. 79 (1997) 123.

[15] Y. Okubo, Y. Ochiai, D. Vasileska, R. Akis, D.K. Ferry, J.P. Bird, K. Ishibashi, Y. Aoyagi, T. Sugano, Phys. Lett. A 236 (1997) 120.

[16] D.P. Pivin Jr., J.P. Bird, R. Akis, D.K. Ferry, Y. Aoyagi, T. Sugano, Semicond. Sci. Technol. 13 (1998) A11.

[17] J.P. Bird, A.P. Micolich, H. Linke, D.K. Ferry, R. Akis, Y. Ochiai, Y. Aoyagi, T. Sugano, J. Phys.: Condens. Matter 10 (1998) L55.

[18] D.K. Ferry, G. Edwards, K. Yamamoto, Y. Ochiai, J.P. Bird, K. Ishibashi, Y. Aoyagi, T. Sugano, Jpn. J. Appl. Phys. 34 (1995) 4338.

[19] J.P. Bird, K. Ishibashi, D.K. Ferry, Y. Ochiai, Y. Aoyagi, T. Sugano, Phys. Rev. B 51 (1995) R18037.

[20] J.P. Bird, H. Linke, J. Cooper, A.P. Micolich, D.K. Ferry, R. Akis, Y. Ochiai, R.P. Taylor, R. Newbury, P. Omling, Y. Aoyagi, T. Sugano, Phys. Stat. Sol. (b) 204 (1997) 314.

[21] D.K. Ferry, R. Akis, J.P. Bird, Superlatt. Microstruc. 23 (1998) 611.

Nanoquakes at work: A quantum conveyor belt for photons

Achim Wixforth *

Sektion Physik der Ludwig Maximilians-Univ. München, Geschwister-Scholl-Platz 1, D-80539 München, Germany

Abstract

Storing light in the lateral potential of an intense piezoactive surface acoustic wave propagating on a semiconductor quantum well is demonstrated. We show that the lateral piezoelectric fields of the wave are strong enough to field-ionize and thus dissociate optically generated excitons and to efficiently capture the resulting free electrons and holes. As the wave propagates at the speed of sound across the semiconductor sample, the trapped carriers can be transferred to a location on the sample being different from the one of the optical excitation. Deliberate screening of the lateral electric fields leads to an induced recombination after very long storage times. We explain the physical mechanisms responsible for this remarkable effect and investigate the ionization, trapping, transport, and recombination in detail. © 1998 Elsevier Science B.V. All rights reserved.

The storage of light in some kind of a container and the deliberate release of the radiation at possibly even another location sounds like a very hard task as it is difficult to trap the speedy photons for appreciable long storage times. Especially in semiconductors it seems to be contradicting at first glance. On the one side, there are semiconductors with large absorption coefficients as their direct band structure provides a large probability for a photon to generate an electron–hole (eh) pair. For the same reason, however, the recombination lifetimes are quite short. On the other hand, semiconductors with an indirect bandgap provide longer life-

times but for the same reason the absorption in form of (eh) pairs is very small. Early attempts to overcome this contradiction relied on the fact that a given band structure in k-space may be efficiently modified if it is modulated in real space. For instance, alternating layers of n-type and p-type material result in the so-called nipi superlattices [1,2], where potential minima of both the conduction and the valence band alternate with the period of the superlattice. In these structures, a spatial separation of optically generated electrons and holes has been achieved which resulted in unusual long recombination times. The reason for the increase in recombination lifetime is the reduced wave function overlap between the photo generated carriers being directly linked to the recombination probability.

* Tel.: +49 89 2180 3732; fax: +49 89 2180 3182; e-mail: achim. wixforth@physik.uni-muenchen.de.

Another recent attempt uses a coupled double quantum well system in which the electrons and holes after their generation reside in different parts of the potential [3]. Illumination with a far-infrared source is then used to transfer the electron in real space towards the hole and to induce a photoluminescence recombination. Hence – it seems quite favorable to rely on a spatial separation of photo generated carriers in a direct gap semiconductor to achieve considerably prolonged recombination lifetimes while retaining the large absorption coefficients.

Here, we would like to report on a different technique to achieve this spatial separation of the photo generated electrons and holes in a semiconductor quantum well [4]. For this purpose, we use surface acoustic waves (SAW) propagating on the piezoelectric semiconductor substrate [5]. Surface acoustic waves are modes of elastic energy that propagate at the surface of a semi-infinite medium. The simplest modes of such waves are Rayleigh waves, in which the particle displacement is elliptic in the sagittal plane perpendicular to the direction of propagation. If the medium is piezoelectric, the waves are accompanied by appreciable lateral and vertical electric fields giving rise to both lateral and vertical modulation of the band edges in the semiconductor. If these fields are large enough, exciton ionization or a spatial separation of the photo generated e–h pairs can result [6]. The SAW are quite efficiently excited on a piezoelectric substrate by means of the so-called interdigital transducers (IDT). Those consist of two comb-like metalized structures being connected to a high-frequency signal generator. Mediated by the inverse piezoelectric effect, the application of a voltage to the IDT leads to a deformation of the substrate right below the transducer. If the frequency of the excitation is chosen such that the condition $f = v/p$ is met, a coherent and monochromatic sound wave is launched along the surface of the crystal. Here, v denotes the SAW velocity for the given crystal cut and p the periodicity of the IDT.

In Fig. 1, we depict the typical sample layout as used in our experiments. The undoped quantum well samples are grown by molecular beam epitaxy on a (1 0 0) – GaAs substrate. The well consists of 10 nm pseudomorphic $In_{0.15}Ga_{0.85}As$ grown on a 1 μm thick GaAs buffer and is covered by a 20 nm thick GaAs cap layer. The active area of the sample is etched
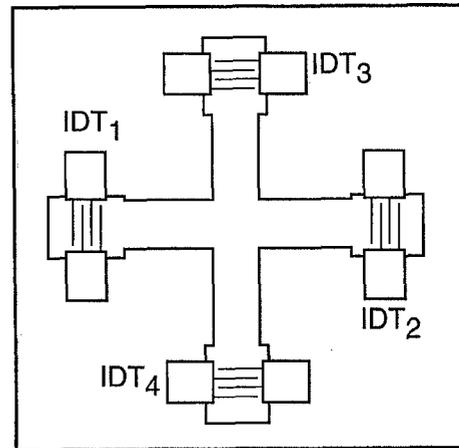
Fig. 1. Sketch of the sample geometry used in our experiments. The active layers of the undoped semiconductor quantum well are etched in the form of a cross-like mesa. On each end of the arms, an IDT is deposited. $IDT_{1,2}$ operate at a frequency of $f = 840$ Mhz, $IDT_{3,4}$ are designed for $f = 420$ Mhz.

into two 2.5 mm long and 0.3 mm wide bars forming a cross-like mesa with an IDT at each end. The SAW velocity for (1 0 0) GaAs and [1 1 0] propagation is $v = 2865$ m/s, and the IDTs are designed to operate at a center frequency $f_{SAW} = 840$ MHz (IDT_1) and $f = 420$ MHz (IDT_3), respectively. They are partially impedance matched to the 50 Ω radio frequency (RF) circuitry using an on-chip matching network thus strongly reducing the insertion loss of each transducer. The sample is mounted in an optical cryostat and the experiments presented here are performed at $T = 4.2$ K. Light from either a pulsed laser diode ($\lambda_{laser} = 780$ nm) or a tuneable titanium–sapphire laser is used for optical interband excitation above band gap and the photoluminescence (PL) of the sample is analyzed in a triple grating spectrometer. Either a photomultiplier or a charged coupled device (CCD) serve as a detector for the PL. Application of a high-frequency signal to one of the IDTs launches a SAW that can be detected at the other IDT after the acoustic delay of the order of 1μs determined by the spacing of the IDTs. Either pulsed or continuous wave (cw) operation of the SAW transducers is possible.

In Fig. 2, we show the observed quantum well PL under the influence of a SAW at $f = 840$ MHz for different SAW amplitudes [4]. The PL is analyzed at the same spot where it is excited, close to the mesa center. As can be seen from the figure, an increase
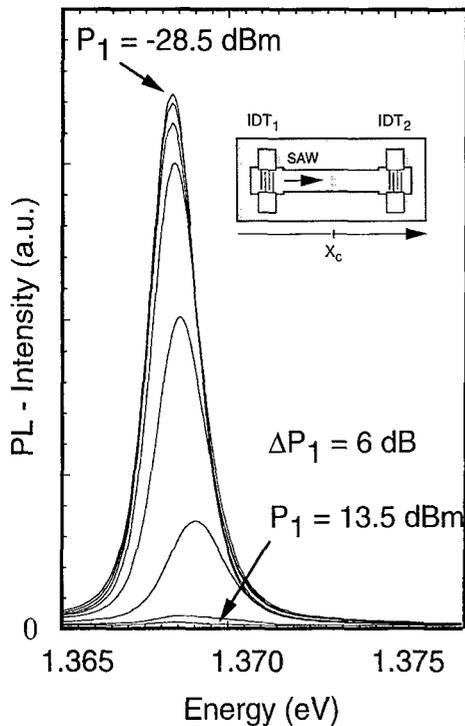
Fig. 2. Photoluminescence (PL) spectra for the quantum well sample under the influence of an intense SAW. With increasing SAW power the PL intensity rapidly decreases whereas its energetic position and line width remain nearly unaffected. At the highest acoustic power show in the figure, the PL is completely quenched [3].


Fig. 3. Lateral and vertical components of the piezoelectric fields in the SAW as a function of the acoustic power and for the given parameters. Also shown (right axis) is the piezoelectric potential associated with the wave.

of the SAW power leads to a decrease of the detected PL intensity, whereas the energetic position and line width remain nearly unaffected. For the highest powers shown ($+13.5\,\text{dBm} = 22.4\,\text{mW}$), the PL becomes completely quenched. This quenching of the PL by the SAW is interpreted by an ionization of the optically generated excitons in the strong lateral piezoelectric fields. To give an impression on the achievable field strengths, we show in Fig. 3 how both the vertical and the lateral fields depend on the acoustic power imposed to our sample. At the same time we plot the magnitude of the corresponding SAW potential. Even for quite moderate SAW powers, very high electric fields of the order of several kV/cm are easily achieved. These piezoelectric fields of the SAW modulate the band edges with respect to the chemical potential similar as in doping superlattices [1,2] or statically imposed laterally periodic electric fields

using an interdigitated gate electrode [7]. In our moving potential superlattice with period $\lambda = 3.4\,\mu\text{m}$ the excitons become polarized predominantly by the lateral electric field until they dissociate at high fields into spatially separated e–h pairs. These are then efficiently stored in the potential minima and maxima of the conduction and the valence band, respectively. A very simple estimate yields that fields as high as $E_{\text{crit}} \approx E_b/(ea_B^*) \sim 10^4\,\text{V/cm}$ will ionize photo generated excitons. $E_b \sim 9\,\text{meV}$ denotes the exciton binding energy, e the elementary charge, and $a_B^* \sim 9\,\text{nm}$ the effective (Bohr) exciton radius of our system. Hence, the lateral field of the SAW is definitely high enough to dissociate the photo generated excitons in our sample. The influence of the vertical fields, however, can be neglected in our case as we are working on a comparatively thin ($d = 10\,\text{nm}$) quantum well.

The observed quenching of the PL under the influence of an intense SAW is already an indication of the increased trapping probability in the moving lateral potential of the SAW. As the potential modulation is moving with the speed of sound, consequently, the spatially separated and trapped e–h pairs are swept away by the SAW and propagate across the sample without recombining. The dramatically prolonged recombination time together with their propagation along the surface of the quantum well enables us to directly follow the transport of photo generated carriers over macroscopic distances that may exceed millimetres in our experiment. For this purpose, we need to cancel out the lateral electric fields that are responsible for the trapping of the charges. On a piezoelectric semiconductor, this is easily achieved by a
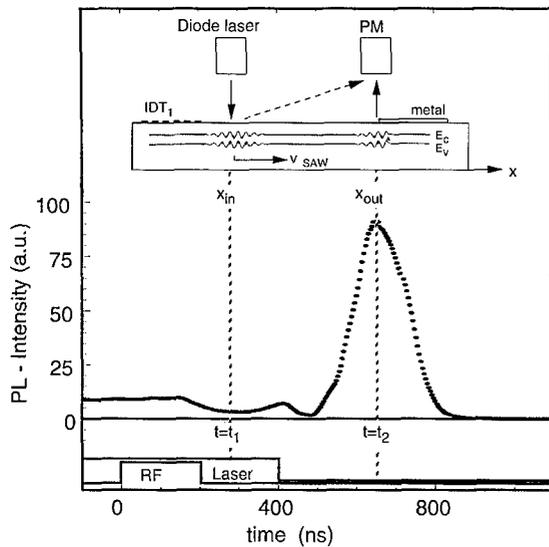
Fig. 4. Experimental evidence for the transport of photo-generated charges in the field of the SAW. A short RF pulse is transformed into a SAW packet at $f = 840$ Mhz. Transversing the excitation region ($x = x_{in}$), this SAW pulse is "loaded" with e–h-pairs which immediately become spatially separated. The spatial separation dramatically decreases the recombination probability which in turn results in a tremendously increased lifetime. After very long storage or delay times the SAW arrives at a metalized region of the sample where the lateral electric fields are screened and the recombination of the photo-generated charges is no longer impossible. The result is a strong PL signal at this location ($x = x_{out}$) [3].

metallization of a part of the samples surface [8]. In this case, the lateral fields are screened and the trapping potentials are flattened out. The carriers are now free again to diffuse, hence considerably increasing their recombination probability. In other words, it should be possible to optically generate excitons at some point of the sample, dissociate them in the field of a SAW and then let them "surf" on the SAW to a different location of the sample. Here, a metallization of the sample surface leads to a re-assembly of the stored bipolar charges into photons which can be detected in the form of a time delayed PL at the remote location.

The corresponding experiment is sketched in Fig. 4. We show the PL being detected at the site x_{out} as a function of time after the optical generation of excitons at site x_{in}. The lower part of the figure indicates the timing sequence of the experiment. A short ($\tau = 200$ ns) SAW pulse at $P_{SAW} = +13$ dBm is generated at $t = t_0$. To avoid any spurious effects, we switch
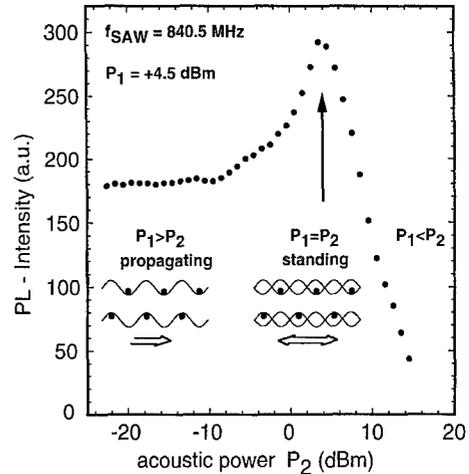
Fig. 5. PL intensity as a function of the acoustic power P_2 at constant power $P_1 = +4.5$ dBm for two counter propagating SAW of the same frequency $f_1 = f_2 = 840$ MHz. Only for $P_1 P_2$ (standing wave condition) the PL recovers, whereas for the other cases the PL is efficiently suppressed [3].

the laser off after the SAW pulse has transversed the optical excitation area at $x = x_{in}$ and $t = t_1$. Some 400 ns later ($t = t_2$), a strong PL signal is detected at $x = x_{out}$, exactly at the time when the "loaded" SAW arrives at the screening electrode. The light has been "stored" in the SAW potential and then transported in the form of a bipolar polarization to this remote location, like on a conveyor belt. The storage time reached in this experiment is mainly given by the transport path on the finite length of our sample ($\Delta x = 1$ mm, in this case). Only a small fraction of the photoexcited carriers is lost along this way [4] and hence much longer storage times can be achieved by a more sophisticated sample design. The fixed time delay between excitation and detection of the PL in the present experiment, however, is given by the fixed distance between excitation area ($x = x_{in}$) and detection area ($x = x_{out}$).

A simple way to also achieve tuneable time delays is presented in Fig. 5. Here, the recombination of the trapped bipolar charges is triggered by a second sound wave propagating in the opposite direction of the first one. If the amplitudes and frequencies of both waves are the same, a standing wave pattern will form, in which the trapping potential is partially lifted and the e–h pairs are free to recombine again. We show the result of an experiment, where two counterpropagating SAWs at $f = 840$ MHz are used to perform this task.
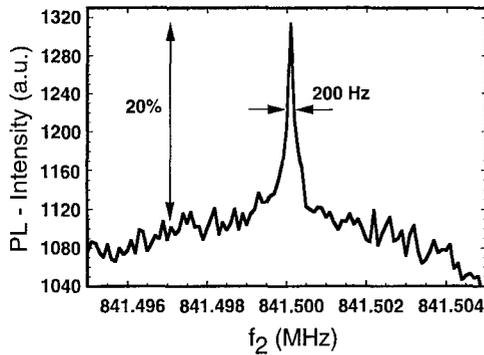
Fig. 6. Corresponding experiment to Fig. 5 in frequency space. Two counter propagating waves of constant and equal power ($P_1 = P_2 = +4.5$ dBm) are launched across the sample. The frequency of the first SAW $f_1 = 841.5$ Mhz, whereas f_2 is tuned. Only for the exact standing wave condition ($f_1 = f_2$) a bright PL is observed. Note that the width of the peak is only some 200 Hz at a center frequency of about 1 Ghz!

We keep the amplitude of the first (pumping) SAW fixed such that the PL is already quenched by about 50%. The amplitude of the second SAW is varied during the experiment. As long as it is small as compared to the one of the "conveying" SAW, the PL intensity remains nearly unaffected. However, as soon as both SAW amplitudes are of the same order, and a standing wave pattern forms, the PL intensity increases dramatically and finally at $P_1 = P_2$ reaches its original value. For higher powers P_2 the situation reverses and now SAW$_2$ starts to efficiently quench the PL.

Even more impressive is the corresponding experiment in frequency space that is shown in Fig. 6. Here, both amplitudes of the SAWs are kept equal and fixed at $P_1 = P_2 = +4.5$ dBm, like in Fig. 5. Now the frequency of SAW$_2$ is varied. As long as both SAW frequencies are different by more than about 1 part in 10^7 (!), the PL remains efficiently suppressed. Only for the exact standing wave geometry the PL recovers. Then and only then the time-averaged wave function overlap of both electrons and holes becomes non-zero which in turn is equivalent to a non-vanishing recombination probability. The extreme sensitivity of this experiment regarding smallest frequency differences might even have some importance for future device applications.

A crucial point for the above experiments is the electric-field-induced ionization process for the exci-

tons. If the ionization in incomplete, only a fraction of the photo-generated charges will be spatially separated and trapped in the wave. This together with a possible escape of some carriers will cause a reduced transport efficiency of the conveyor belt. Above, we gave a simple estimate for $E_{crit} \approx 10^4$ V/cm, the critical field necessary for a spatial separation of the electrons and holes. From our experiments, however, we extract a somewhat lower value – a disagreement that will be addressed in the following: we are making use of the fact that SAW transmission experiments are very sensitive to small areal conductivities of a close-by charge carrier system [5]. The attenuation as a function of the conductivity indicates a maximum at some critical conductivity $\sigma_m \approx 3 \times 10^{-7} \Omega^{-1}$/sq [5], where the impedance of the space-charge layer is matched to the surface SAW admittance of the sample. To directly observe the exciton ionization process in the field of the "pump" SAW, we monitor the transmitted intensity P_4 of a second SAW propagating perpendicularly to the "pump" SAW as a function of the acoustic power P_1 of the latter [9]. On the length scale of the SAW, the optically generated excitons behave like neutral particles as their effective Bohr radius is much smaller than λ_{SAW}. The areal conductivity related to the field-induced polarization of the excitons is negligible, so we do not expect any screening of the piezoelectric potential of the transmitted SAW and hence no attenuation as described in Ref. [5]. However, once the excitons are field-ionized, their "fragments" – electrons and holes – are spatially separated and subsequently confined in the lateral potential wells of the pump SAW. Now the probe SAW encounters many parallel stripes of free electrons and holes, resembling a *lateral* "nipi"-structure with a period $\lambda_{SAW,1}$. Those electrons and holes are free to move along the direction normal to the propagation of the pumping and confining SAW and are therefore expected to interact with the second SAW according to their areal conductivity.

This is exactly what we observe experimentally. The result is shown in Fig. 7, where we plot the transmitted SAW intensity $P_4 = P_3 \exp(-\Gamma d)$ as a function of the power P_1 of the "pumping" SAW and for two different optical excitation energies E_{las}, as will be discussed below. P_3 is the incident power of the "probe" SAW, d the length of the interaction region, which in our case corresponds to the width of the pumping SAW path, namely $d = 300$ µm.
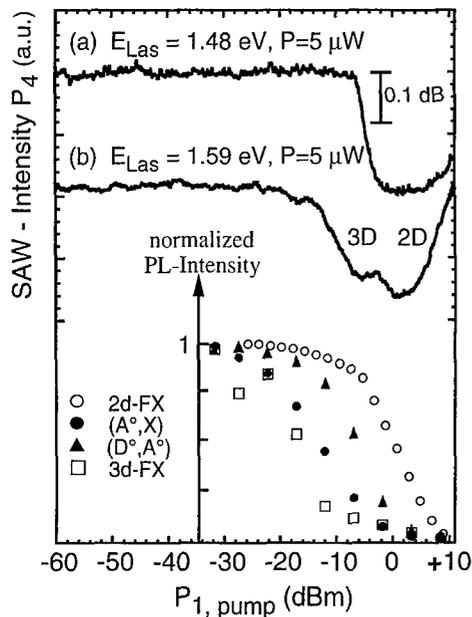
Fig. 7. SAW transmission experiment to study the process of the exciton ionization in the field of a pumping SAW P_1. A second SAW is propagated at low power P_3 perpendicular to the first one and is analyzed in intensity P_4 at IDT$_4$. The attenuation of this "probe" SAW strongly depends on the areal conductivity of the photo-generated charges. As long as the excitons are not yet dissociated by the SAW$_1$, they behave like neutral particles with zero areal conductivity. The sudden inset of SAW absorption indicates the occurrence of a finite areal conductivity caused by the ionization of the excitons. In (a), only the InGaAs quantum well is optically excited, resulting in purely 2d excitons, whereas in (b) also the barriers are excited and 3d excitons are generated simultaneously [8].

The inset of the figure shows the measured intensity of some prominent PL lines as a function of the SAW pumping power P_1. 2d-FX denotes the quantum well (2d) free exciton as already shown in Fig. 2, the other symbols are attributed to the bulk (3d) free exciton and some acceptor (A) or donator (D) bound excitonic states [10]. It is interesting to note that – in agreement to the observations in Ref. [10] – all different PL lines become suppressed under the influence of the SAW. The threshold fields, however, differ slightly which can be attributed to their different binding energies. At very low pumping power levels P_1, the probing SAW is nearly unaffected as indicated by a constant transmitted SAW intensity P_4 over several decades of P_1. However, for pumping power levels greater than approximately $P_1 = -10$ dBm, a strong decrease of

the transmitted SAW intensity P_4 or equivalently an increase of the attenuation coefficient Γ is observed. This decrease of the probe SAW intensity is strongly dependent on the energy E_{las} of the exciting laser, as can be seen in the upper two traces in Fig. 7.

In Fig. 7a, the energy of the incident laser $E_{las} = 1.48$ eV is chosen such that only the InGaAs quantum well is optically excited, resulting in a selective population of the QW by photo generated excitons. For low acoustic pump powers P_1, their quasi-static conductivity in the QW is negligible, indicated by a constant probe intensity P_4.

However, for $P_1 > -6$ dBm, corresponding to a lateral piezoelectric field strength of $E_1 > 600$ V/cm, the probability for the dissociation of the excitons by field-induced tunnelling processes rapidly increases which is marked by a sharp onset of the absorption of the probe SAW. This change in the SAW transmission can be understood by the absorption process as described in Ref. [5]. The point of minimum transmission (or equivalently maximum attenuation) corresponds to the situation, where the averaged conductivity σ of the spatially separated electrons and holes reaches the critical conductivity σ_m as defined above. Further increase of the pumping power P_1 results in an increase of the carrier densities and thus, caused by an increased screening, in a smaller interaction between the probe SAW and the mobile carriers. The result is a significantly reduced attenuation coefficient Γ.

Optical excitation at a laser energy of $E_{las} = 1.59$ eV (see Fig. 7b) being larger than the band gap of the quantum well barriers leads to the generation of both 3D excitons in this GaAs barrier (bulk) material and in the InGaAs quantum well. This simultaneous generation of 2D and 3D excitons is clearly observed in our experiment, as the transmitted SAW intensity exhibits two well-defined minima at about $P_1 = -8$ dBm and $P_1 = 0$ dBm. Those minima correspond to maxima in the SAW attenuation Γ as described above. The onset of the SAW attenuation, however, is now shifted towards lower pumping powers of $P_1 = -13$ dBm, corresponding to a piezoelectric field strength of only $E_1 = 300$ V/cm. This attenuation at smaller pumping powers can be attributed to the dissociation of the 3d excitons in the GaAs barrier material. As the exciton binding energy of 3d excitons is considerably smaller than the one for the 2d excitons in the InGaAs quantum well, the piezoelectric fields necessary for

ionization will also be smaller. Caused by the variation of the SAW – induced fields with depth into the crystal, the dissociation rate of the 3d excitons also varies with depth. Both this field strength variation as well as the presence of different excitonic states with different binding energies in 3d will result in a somewhat smoother onset of SAW attenuation in the 3d case as compared to the 2d case. Note, that the onset of exciton ionization, indicated by the onset in the SAW absorption, agrees very well with the decrease of the PL intensity of the most intense neutral acceptor bound exciton (A^0, X) as shown in the inset of Fig. 7. A theoretical estimate of the critical electric fields necessary for the ionization of either 2d or 3d excitons is given in Ref. [9]. This estimate is based on a calculation of the tunnelling ionization time for excitons in the presence of an electric field [6] and turns out to be of the order of 0.5 ns for an electric field of 1 kV/cm. Time-resolved measurements of the PL on the same samples [1] showed a very good agreement between our calculation, the measured ionization threshold fields and the tunnelling ionization lifetime. It would be very interesting to study the dependence of the lifetime on different external parameters like a magnetic field or the temperature. According to our model the ionization threshold electric field should then be also strongly affected.

In summary, we have demonstrated that it is in fact possible to capture light in a container and then release it at some other time and at some remote location. We showed that the combination of vertical confinement of bipolar charges in a quantum well together with a strong lateral confinement mediated by the lateral piezoelectric fields of a surface acoustic wave can result in what we call a "photon conveyor belt". Combining the large absorption coefficient of direct band gap semiconductors and the long lifetimes for optical excitations in indirect band gap semiconductors yields a new and promising approach also for future optoelectronic device applications.

References

[1] G.H. Döhler, H. Künzel, D. Olego, K. Ploog, P. Ruden, H. Stolz, Phys. Rev. Lett. 47 (1981) 864.

[2] R.A. Street, G.H. Döhler, J.N. Miller, P.P. Ruden, Phys. Rev. B 33 (1986) 7043.

[3] M. Rüfenacht, S. Tsujino, Y. Ohno, H. Sakaki, Appl. Phys. Lett. 70 (1997) 1128.

[4] C. Rocke, S. Zimmermann, A. Wixforth, J.P. Kotthaus, G. Böhm, G. Weimann, Phys. Rev. Lett. 78 (1997) 4099.

[5] A. Wixforth, J. Scriba, M. Wassermeier, J.P. Kotthaus, G. Weimann, W. Schlapp, Phys. Rev. B 40 (1989) 7874.

[6] D.A.B. Miller, D.S. Chemla, T.C. Damen, A.C. Gossard, W. Wiegmann, T.H. Wood, C.A. Burrus, Phys. Rev. B 32 (1985) 1043.

[7] A. Schmeller, W. Hansen, J.P. Kotthaus, G. Tränkle, G. Weimann, Appl. Phys. Lett. 64 (1994) 330.

[8] C. Rocke, S. Manus, A. Wixforth, M. Sundaram, J.H. English, A.C. Gossard, Appl. Phys. Lett. 65 (1994) 2422.

[9] C. Rocke, A.O. Govorov, A. Wixforth, G. Böhm, G. Weimann, Phys. Rev. B 57 (1998) R6968.

[10] K.S. Zhuralev, D.V. Petrov, Yu.B. Bolkhovityanov, N.S. Rudaja, Appl. Phys. Lett. 70 (1997) 3389.

[1] We thank A. Haugeneder, G. von Plessen, and J. Feldmann for the time-resolved PL measurements on our samples.

Ballistic and dissipative electron transport in semiconductor superlattices

G. Strasser *, C. Rauch, K. Unterrainer, W. Boxleitner, E. Gornik

Solid State Electronics, Technical University Vienna, Floragasse 7, A-1040 Vienna, Austria

Abstract

Ballistic electron transport is used to study the transmittance of GaAs/GaAlAs superlattices. In a three terminal device an energy tunable electron beam is injected via a tunneling barrier into an undoped superlattice and the transmitted current is measured as a function of the injector energy. Significant increase of the collector current is observed due to miniband conduction in the GaAs/AlGaAs superlattice. At flat band condition the transfer ratio $\alpha = I_C/I_E$ can be used to probe miniband position and miniband widths. Measuring the transfer ratio of superlattices at various bias conditions, a significant decrease of the miniband transmission with increasing electric field is observed, which can be attributed to the transition between coherent and incoherent transport in these superlattices. The results agree very well with theory. © 1998 Elsevier Science B.V. All rights reserved.

PACS: 73.20.Dx; 73.40.Gk; 72.80.Ey; 73.40.Kp

Keywords: Ballistic electron transport; Semiconductors; Superlattices

1. Introduction

Electron transport in superlattice minibands was first considered by Esaki and Tsu [1]. In their model classical transport determined by the dispersion relation of the miniband is considered. In the quantum mechanical picture a voltage drop over the superlattice causes the break up of the miniband into a Wannier–Stark ladder [2] which requires a different description in terms of tunneling [3,4]. The tunneling current decreases with increasing electric field as the wave functions become localized. In the calculation of the

* Corresponding author. Tel.: +43 1 504 55 25 18; fax: +43 1 504 55 25 9; e-mail: strasser@macmisz.fke.tuwien.ac.at.

current through an infinite superlattice a phenomenological scattering time has to be introduced to reproduce the onset of the negative differential conductivity at $\omega_B \cdot \tau = 1$, in order to mimic realistic current voltage characteristics [1,5,6] and the presence of incoherent transport.

A vast study of electrical transport in superlattices was done in the last decade [7–11]. In short period superlattices formation of the allowed and forbidden bands for resonant tunneling and band filling effects were confirmed experimentally [12,13]. The study of biased superlattices was severely hindered by space charge build up and domain formation [14]. So far there is no conclusive experiment which shows the occurrence of Esaki–Tsu-type negative differential

resistance (RTD) due to Bragg reflection at the zone boundary. Sybille et al. performed the most extensive study in biased, doped superlattices and observed negative differential velocity [15,16]. They were able to fit the drift velocity–voltage curves by using a modified Esaki Tsu drift diffusion model. Optical experiments led to a breakthrough by the observation of Bloch oscillations in the time domain of undoped superlattices [17]. In a previous experiment we have developed a three terminal technique which allowed the study of transport in undoped superlattices [18]. In this paper, we extend this technique of hot electron spectroscopy to biased superlattices [19].

2. Experimental

We measured ballistic transport in GaAs/$Ga_{0.7}Al_{0.3}As$ superlattices, where the influence of electron–electron and electron-impurity scattering can be neglected. Under flat band conditions the eigenstates of the periodic structure are extended over the entire length of the superlattice. To investigate the superlattice properties, a hot electron transistor structure is used [20]. An energy tunable electron beam is generated by a tunneling emitter, passes the superlattice after traversing a thin highly doped GaAs region (base) and an undoped drift region. Electrons passing the superlattice are collected in a doped GaAs layer, reflected electrons are collected by the base contact.

Our samples, grown by molecular beam epitaxy, have the following common features: A highly doped n^+-GaAs collector contact layer is grown on a semi-insulating GaAs substrate. Followed by a superlattice and the drift regions which are slightly n-doped ($\sim 5 \times 10^{14}$ cm^{-3}), in order to avoid undesired band bending. To reduce quantum mechanical confining effects originating from the quantum well formed by the emitter barrier and the superlattice the drift region is chosen to be at least 200 nm in width. This is followed by a highly doped (2×10^{18} cm^{-3}) n^+-GaAs layer (base) of 13 nm width. As found in previous experiments [21], about 75% of the injected electrons traverse the base ballistically. On top of the base layer a 13 nm undoped $Ga_{0.7}Al_{0.3}As$ barrier is grown followed by a spacer and an n^+-GaAs layer, nominally doped to $n = 3 \times 10^{17}$ cm^{-3}, in order to achieve an estimated normal energy distribution of injected elec-
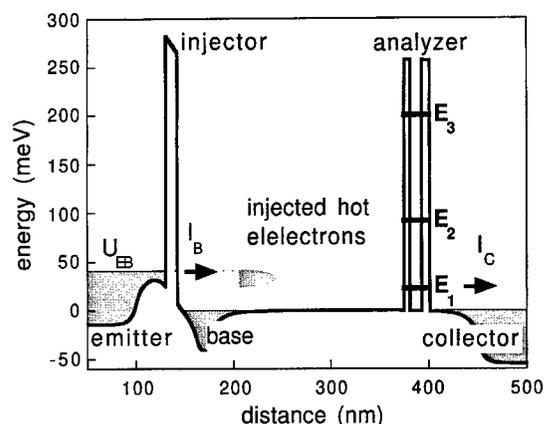
Fig. 1. Schematic band diagram of a three terminal device with a resonant tunneling diode as a filter to measure the injected electron distribution.

trons of about 20 meV [22]. Finally, an n^+-GaAs contact layer ($n = 1 \times 10^{18}$ cm^{-3}) is grown on top of the heterostructure to form the emitter.

Standard photolithographic and wet etching techniques were employed in three terminal device processing. The emitter, base, and collector were contacted from above using a standard AuGe/Ni alloy. An emitter contact pad, which is connected to the $30 \times 30\,\mu m$ emitter mesa, is evaporated on top of a polyimid isolation layer.

3. Results and discussion

Before we start to investigate the transmittance of the superlattice minibands, we want to have an exact knowledge of the energy distribution of the injected electrons. Thus, a three terminal device with a resonant tunneling diode instead of the superlattice was used to determine the shape of the injector. The energy diagram of the conduction band of this device is shown in Fig. 1. A resonant tunneling diode acts as an energy filter of the injected hot electrons. Electrons that have an energy which corresponds to the energy of the first resonant state can pass the energy filter, otherwise they are scattered back to the base and do not contribute to the measured collector current. The RTD (6 nm $Al_{0.3}Ga_{0.7}As$ barrier/12 nm GaAs well/8 nm $Al_{0.3}Ga_{0.7}As$ barrier) is designed in a way that the first state (23 meV) is well below the energy of an
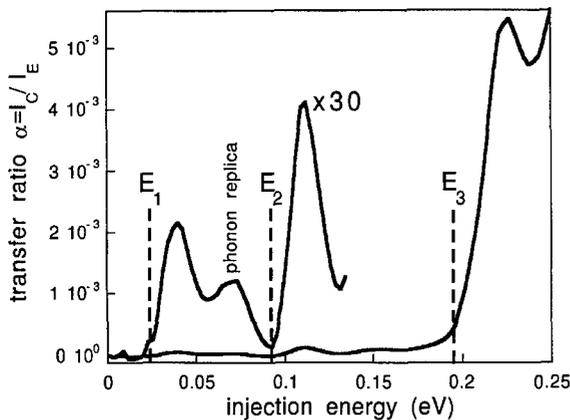
Fig. 2. The transfer ratio $\alpha = I_C/I_E$ versus injection energy of the three terminal device with RTD. For injection energies up to 130 meV the transfer ratio is multiplied by a factor of thirty. The resonant states of the analyzer are indicated by dashed lines. LO-phonon replicas are observed for all levels.

LO-phonon ($\hbar\omega = 36$ meV) in order not to lose electrons that are injected and scattered by LO-phonons. Since the energy level broadening at 4.2 K is less than 1 meV, the measured ballistic current is in the order of 1 μA/cm^2.

The measured transfer ratio is shown in Fig. 2 as a function of the injection energy which is equivalent to the applied negative emitter bias. Below the energy of the first resonant state E_1 of the RTD we observe no collector current, since no electrons are injected that have an energy which is high enough to cross the resonant tunneling barrier. The onset at about 21 meV determines the energy level of the first-resonant state (E_1) of the analyzer RTD. By further increasing the emitter bias, we use the constant energy position of the first resonant state to perform spectroscopy of the injected hot electron distribution. Because the resonant line width of the double barrier structure is negligible compared with the width of the injector distribution, the measured transfer ratio is proportional to the hot electron distribution of the injector. The observed peak at about 70 meV is due to electrons that are injected and have lost the energy of 36 meV due to LO-phonon scattering during transversing the base layer and the drift region. Since the k-vector is conserved for LO-phonon scattering processes, these electrons can be collected as well for higher injection energies. The transfer ratio does not drop to zero in

between these peaks due to the overlap of the injected electrons that traverse the drift region without scattering and those electrons that have lost the LO-phonon energy. Starting at about 90 meV we observe transport through the second-resonant state (E_2). This peak reproduces the shape of the injected electron distribution for higher injection energies. At 187 meV we observe the third quantized level of the RTD. It should be noted that the measured onsets of the transfer ratio fits very well to the calculated positions of the position of the quantized states ($E_{1,\text{calc}} = 23$ meV, $E_{2,\text{calc}} = 87$ meV, and $E_{3,\text{calc}} = 179$ meV). The calculated positions are indicated by dashed lines in Fig. 2.

Since the first peak of the transfer ratio is proportional to the injected hot electron energy distribution we can determine the full-width at half-maximum to be 20 meV. The shape of the distribution is slightly asymmetric with its maximum at the high-energy side with respect to the GaAs conduction band edge. We observe no significant change of the shape of the energy distribution with higher injection energies up 200 meV.

The static transfer ratio $\alpha = I_C/I_E$, of a 8.5 nm/ 2.5 nm GaAs/AlGaAs five period superlattice, plotted in Fig. 3 as a function of the injection energy shows several maxima and a sharp rise at 280 meV. The inset in Fig. 3 shows the measurement circuit used for the determination of the transfer ratio. All measurements are performed in common base configuration at 4.2 K. No current is observed below the energy of the first peak. The position of the first peak coincides very well with the first miniband. Thus, we conclude that the first peak is due to miniband transport through the lowest miniband. For energies higher than the first miniband the transfer ratio drops quite significantly since there is no transport possible through the forbidden minigap of the SL. The second observed peak is shifted 36 meV to higher injection energies and is ascribed to the first LO-phonon emission replica ($\hbar\omega_{LO} = 36$ meV) of the injected electron distribution. The relative position in energy and width are equal to that of the first peak. The energy range of electrons injected at voltages corresponding to these second peak is in the forbidden band and no contribution is expected from electrons which have not lost energy due to optical phonon emission. The peak at 150 meV represents transport through the second SL miniband. For an analysis of the observed features we compare
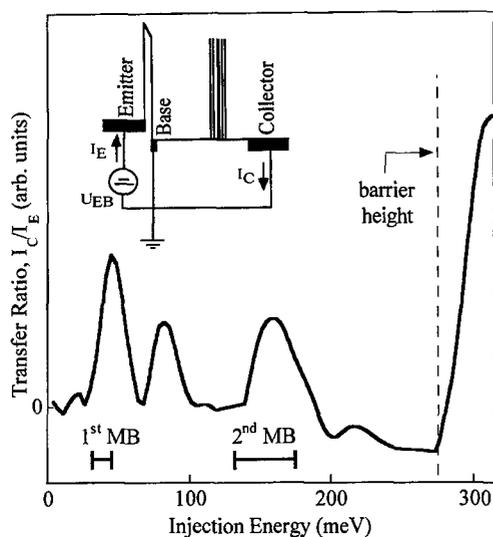
Fig. 3. The transfer ratio $\alpha = I_C/I_E$ versus injection energy ($\approx e.U_{EB}$) of a sample with a 8.5 nm GaAs/2.5 nm AlGaAs superlattice. The calculated miniband positions are indicated by bars (|——|). The inset shows the measurement circuit in the common-base configuration.


Fig. 4. Transfer ratio α versus injection energy at lower injection energies for three samples with different superlattices (|——| indicates the calculated miniband position, |-- --| indicates the broadening due to the energy distribution of the injected electron beam). A double arrow represents the energy of a longitudinal optical phonon ($\hbar\omega_{LO} = 36$ meV).

the experimental data with the theoretically calculated miniband positions. The calculated positions and widths of the first and second miniband are indicated by bars. The sharp rise of the transfer ratio at 280 meV is due to the transition to continuum. This energy, which corresponds to the conduction band offset of the superlattice barriers, gives us a confirmation of the AlAs mole fraction of the AlGaAs compound.

In Fig. 4 we show the transfer ratio α as a function of the injection energy for three different five period superlattice samples with varied well widths at lower injection energies. There is a clear shift of the peaks to higher energies with decreasing superlattice well width. The calculated miniband positions are again indicated by bars as in Fig. 3. Since the calculated miniband width for the widest well is 3.5 meV and the observed peak corresponding to miniband transport through the first miniband has a width of about 24 meV, we can confirm the initial energy distribution of the injector to be about 20 meV wide, this agrees very well with the results from Fig 2.

The measurements on biased superlattices are performed on a superlattice consisting of 10 periods of nominally 2.5 nm thick $Al_{0.3}Ga_{0.7}As$ barriers and 6.5 nm GaAs wells. For these parameters a simple

Fig. 5. Schematic band diagram of a three terminal device with negative bias applied to the superlattice. The miniband positions are indicated by shaded areas. The base is grounded.

Kronig–Penny calculation gives one miniband lying between 46 and 68 meV, and a second one between 182 and 276 meV. The calculated equilibrium Γ-point conduction energy diagram including band bending is shown in Fig. 5 for typical biasing conditions.

Fig. 6 shows a set of measured transfer ratios as a function of electron injection energy at different collector biases. As the injection energy is lower than
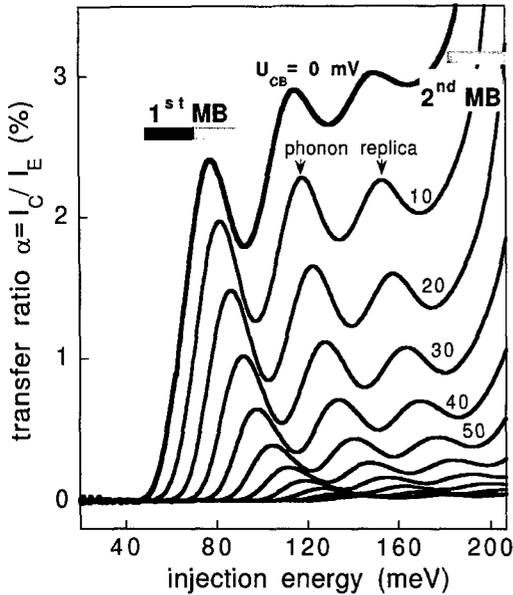
Fig. 6. Transfer ratio α versus injection energy at different collector base voltages. The dark bars indicate the calculated position of the minibands. The gray bars represents the full-width at half-maximum of the injected electron energy distribution.


Fig. 7. Measured total miniband transmission versus applied electric field for a superlattice with five periods and ten periods, respectively. The full (dashed) curve represents the result of a calculation based on a transfer matrix method for five periods (ten periods).

the first miniband no ballistic current is observed. The sharp increase of the transfer ratio at about 45 meV coincides very well with the lower edge of the first miniband. The onset of the transfer ratio (miniband transport) shifts with the applied collector-base bias since the lower edge of the first miniband shifts with the applied bias. The value of the electric field in the superlattice is determined from a comparison with the measured miniband position (using the 3 dB criterion) with the miniband position derived from a self-consistent Schrodinger calculation of the entire structure.

Longitudinal optical phonon replicas, shifted by 36 meV to higher injection energies, are observed at all biases. Due to these phonon replicas the transfer ratio does not vanish between the peaks since the full-width at half-maximum of the injected electron distribution (20 meV) plus the width of the first miniband (22 meV) is greater than the phonon energy. At zero bias the transmission through the superlattice should have a maximum, if we assume the electron wave functions of all superlattice states are extended over the total dimension of the superlattice.

The total miniband transmission, as a measure of the transmission, is defined as the integral over the deconvoluted transfer ratio of the first peak. This represents the transport through the first miniband not taking into account electrons which have lost an LO phonon. Fig. 7 shows the total miniband transmission versus electric field for two identical superlattices with five and ten periods, respectively. It can be seen that the transmission vanishes for an applied electric field exceeding 4 and 8 kV/cm. For low electric fields the superlattice states extend successively one after the other over the total superlattice dimension and become transparent which leads to an increase of current. At zero bias all superlattice states are extended, leading to a maximum of the measured transmission.

Based on a transfer matrix method using an envelope function approximation, we have calculated the transmission times the line width of each superlattice state as a function of the electric field. This value, which is considered to be proportional to the current through the miniband, is multiplied by the slightly asymmetric injected electron distribution. The full curve (five periods) and the dashed curve (ten periods) in Fig. 7 show the result of this calculation. Since no scattering of the injected electrons within

the superlattice is included, the curve acts as an upper limit for coherent transport. It is evident that for the five period superlattice the calculated total miniband transmission, in particular, the quenching of miniband conduction at about 4 kV/cm is in good qualitative agreement with the experiment. Comparing the field dependent transmission of the five and ten period superlattice we find: the transmission for the five period superlattice is symmetric and independent of the bias direction, the ten period superlattice shows a pronounced asymmetry. This asymmetry we assign to the onset of scattering in the ten period superlattice, which is consistent with the assumption of a mean free path in the order of 80 nm, which is just between five and ten periods.

4. Summary

In summary, direct experimental current spectroscopy of minibands in undoped superlattices is demonstrated using the technique of hot electron spectroscopy. Miniband transport through the first and second miniband of superlattices with different well widths as a function of the hot electron injection energy is observed. We measured the injected electron distribution and are able to determine miniband widths and positions for flatband conditions. The obtained miniband widths and positions agree very well with theoretical calculations. The structure described in this paper gives the highest energy resolution reported so far.

Additionally we have shown the controlled decrease of superlattice conduction in a superlattice with bias using the technique of hot electron spectroscopy. For electric fields higher than 4 kV/cm ($\Delta/eFNd = 0.6$) the five period superlattice becomes non transparent. The experimental results are in good agreement with a calculation based on a transfer matrix method assuming coherent transport. The ten period superlattice shows an asymmetric transport behavior. This can be attributed to the transition between coherent and incoherent transport in these superlattices and leads to a mean free path in the order of 80 nm.

References

[1] L. Esaki, R. Tsu, IBM J. Res. Dev. 14 (1970) 61.
[2] G.H. Wannier, Elements of Solid State Theory, Cambridge University Press, London, 1959, pp. 190–193.
[3] R. Tsu, L. Esaki, Appl. Phys. Lett. 22 (1973) 562.
[4] J.B. Krieger, G.J. Iafrate, Phys. Rev. B 33 (1986) 5494.
[5] R.A. Suris, B.S. Shchamkhalova, Sov. Phys. Semicond. 18 (1984) 738.
[6] R. Tsu, G. Döhler, Phys. Rev. B 12 (1975) 680.
[7] L. Esaki, L.L. Chang, Phys. Rev. Lett. 33 (1974) 495.
[8] F. Beltram, F. Capasso, D.L. Sivco, A.L. Hutchinson, S-N.G. Chu, A.Y. Cho, Phys. Rev. Lett. 64 (1990) 3167.
[9] P. Voisin, J. Bleuse, C. Bouche, S. Gaillard, C. Alibert, A. Regreny, Phys. Rev. Lett. 61 (1988) 1639.
[10] J. Bleuse, P. Voisin, M. Allovon, M. Quillec, Appl. Phys. Lett. 53 (1988) 2632.
[11] E.E. Mendez, F. Agulló-Ruada, J.M. Hong, Phys. Rev. Lett. 60 (1988) 2426.
[12] P. England, M. Helm, J.R. Hayes, J.P. Harbison, E. Colas, L.T. Florez, Appl. Phys. Lett. 54 (1989) 647.
[13] G. Brozak, M. Helm, F. DeRosa, C.H. Perry, M. Koza, R. Bhat, S.J. Allen, Phys. Rev. Lett. 64 (1990) 3163.
[14] H.T. Grahn, R.J. Haug, W. Müller, K. Ploog, Phys. Rev. Lett. 67 (1991) 1618.
[15] A. Sibille, J.F. Palmier, H. Wang, F. Mollot, Phys. Rev. Lett. 64 (1990) 52.
[16] A. Sibille, J.F. Palmier, F. Mollot, Appl. Phys. Lett. 60 (1992) 52.
[17] Ch. Waschke, H.G. Roskos, R. Schwedler, K. Leo, H. Kurz, K. Köhler, Phys. Rev. Lett. 70 (1993) 3319.
[18] C. Rauch, G. Strasser, K. Unterrainer, B. Brill, E. Gornik, Appl. Phys. Lett. 70 (1997) 649.
[19] C. Rauch, G. Strasser, K. Unterrainer, W. Boxleitner, E. Gornik, to be published.
[20] M. Heiblum, M.I. Nathan, D.C. Thomas, C.M. Knoedler, Phys. Rev. Lett. 55 (1985) 2200.
[21] B. Brill, 1996, to be published.
[22] S. Bending, A. Peck, J. Leo, K.v. Klitzing, P. Gueret, H.P. Meier, Solid State Electron. 32 (1989) 1161.

Blue emitting heterostructure laser diodes

G. Landwehr*, A. Waag, F. Fischer, H.-J. Lugauer, K. Schüll

Physikalisches Institut der Universität Würzburg Am Hubland, D-97074 Würzburg, Germany

Abstract

After a brief discussion of the present state of development of blue GaN- and green ZnSe lasers an overview is given of efforts to improve the lifetime of ZnSe-based laser diodes by incorporation of beryllium. Be-chalcogenides have a relatively large strength due to their bonding properties. Because BeTe and ZnSe are almost lattice matched to GaAs, the growth of type II superlattices of high perfection is possible. This allows to produce graded BeTe/ZnSe low resistance contacts to ZnSe, which have performed very satisfactorily. It can be anticipated that by making use of the full potential of Be-chalcogenides in the design of II–VI laser diodes emitting in the green spectral range, the lifetime of such devices can be extended significantly. This would allow numerous technical applications. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Blue emitting laser diodes

1. Introduction

The first semiconductor lasers were realized in 1962 [1,2] They were made of GaAs and were simple pn-junction devices. Only pulsed operation at 77 K was possible, with excessive threshold currents. The introduction of the heterostructure principle in laser design by Kroemer [3] marked a significant progress[1]. The first pulsed double heterostructure lasers were demonstrated [4,5] in 1967. It was not until 1980 that continuous operation of GaAs-(GaAl)As lasers at room temperature was achieved. The devices emitted in the infrared spectral range around 900 nm. Subsequently, the dramatic development of opto-electronics began and lasers with long lifetime were developed for glass fibre communication for the wavelengths range around 1.3 and 1.5 μm.

In order to obtain lasers in the visible spectral range, it is necessary to employ materials with a band gap larger than 2 eV. First light emitting diodes (LEDs) emitting in the green were made of GaP and blue emitting devices were made of SiC. However, these materials have an indirect band structure resulting in a rather weak emission. In order to achieve laser action, materials with a direct band gap are necessary. A look at possible candidates shows that there are two promising classes of semiconductors: the III–V nitrides and the II–VI chalcogenides. In Fig. 1 the energy gap of various III–V and II–VI compounds has been plotted as a function of lattice constant.

The efforts to demonstrate blue and green light emitting diodes (LEDs) and lasers started already

---

* Corresponding author. E-mail: landwehr@physik.uni-wueburg.de.

[1] Independently Zh.I. Alferov, R.F. Kazarinov obtained a patent on a heterostructure laser in the former Soviet Union in 1963.
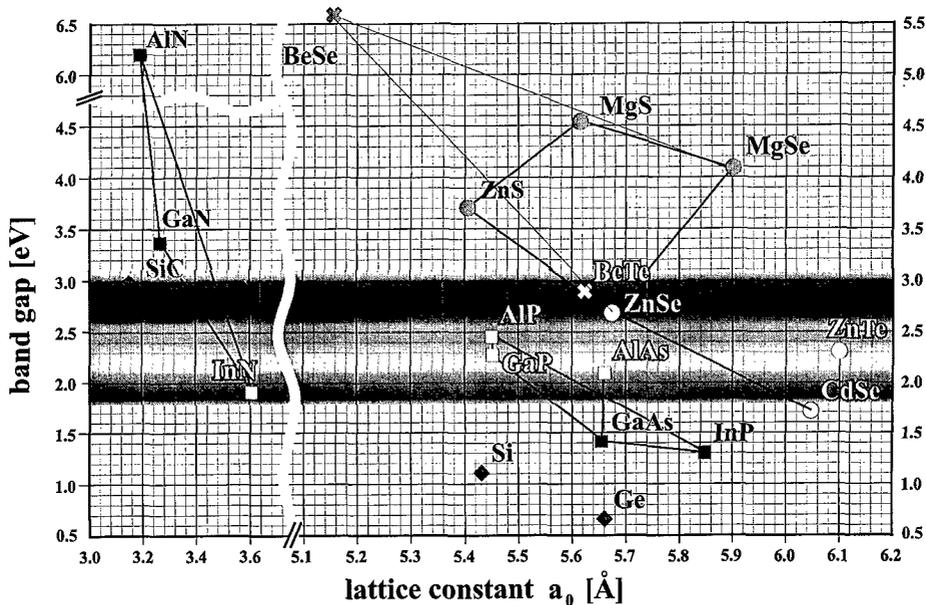
Fig. 1. Band gap versus lattice constant for various wide gap semiconductors.

two decades ago, they centered around GaN and ZnSe. However, the progress was very slow initially. Whereas strong n-type doping was possible for both GaN and ZnSe, the realization of bulk material and thin films with sufficiently high p-type doping caused problems. Since about 1980 it has been tried by many groups to grow p-type ZnSe by MOCVD, but with no success. Then in 1990 it was demonstrated by Park et al. [6] that in ZnSe grown by molecular beam epitaxy (MBE) high hole concentrations could be obtained by using nitrogen as a dopant. Incorporation of the nitrogen at selenium sites could be achieved by using an RF plasma source. The first efforts to produce p-type GaN had started already in the 1960s. However, the early attempts to incorporate reliably group II elements such as Zn, Mg or Be at Ga sites failed. The first p-type GaN was grown in 1989 by Amano and coworkers [7] employing MOCVD and Mg as a dopant. After growth the p-type conductivity was activated by electron irradiation. A breakthrough was achieved by Nakamura and coworkers [8] who in 1992 obtained their first p-type GaN films. Magnesium in GaN-MOCVD films was activated by thermal annealing, rather than by electron irradiation. This was the beginning of a surprising development

which led within 3 yr to commercial, high brightness LEDs based on III–V nitrides. A prerequisite for this success was the capability to fabricate high quality InGaN films, so that heterostructure diodes could be conceived and realized.

2. The evolution of ZnSe- and GaN-based lasers

After the successful p-type doping of ZnSe in 1990 the first laser diode emitting green light was reported already in 1991. The first devices were operated at 77 K in a pulsed mode, they suffered, however, from a short lifetime [9]. The lasers were grown on GaAs substrates, because high quality ZnSe substrates were not available. A look at Fig. 1 shows that the lattice constant of ZnSe and GaAs differ only slightly. Whereas in the first devices the II–VI layers were relaxed, pseudomorphic devices with ternary ZnSSe waveguide layers and quaternary ZnMgSSe claddings were produced [10]. The lifetime of room temperature ZnSe-based lasers, operated in the continuous mode, increased steadily from minutes to more than 100 h in 1996. It was soon recognized that the origin of the observed degradation was non-radiative recombination

at dislocations which had penetrated into the active zone (dark line defects) [11]. The increase in lifetime was achieved by reducing the number of stacking faults and dislocations at the GaAs/ZnSe interface from $10^6/cm^2$ to some $10^3/cm^2$. Systematic studies of the degradation behavior at Sony indicated that not only line defects were responsible for the degradation but also point defects and possibly defect aggregates [12]. An extrapolation of degradation data taken at 80°C to room temperature indicates that the lifetime of the devices should be higher than 10.000 h, which is substantially higher than the reported actual lifetime of the order of 100 h [13]. This is an indication that there must be additional causes for the device failure. One of the reasons for the limitation of the lifetime seems to be the contact on the p-side of the devices. Ohmic metal-semiconductor contacts are usually produced by doping the semiconductor so strongly that tunneling of the free carriers to the metal can occur. This requires carrier concentrations in excess of $10^{19}/cm^{-3}$. It has turned out, however, that p-type doping of ZnSe by plasma activated nitrogen is limited to about $1 \times 10^{18} cm^{-3}$. Additional incorporation of nitrogen leads to an increase in resistance due to defect complexes which show up in the photoluminescence spectra. In order to achieve good ohmic junctions in spite of these difficulties, graded or digital ZnSe/ZnTe contacts have been employed [14,15]. Contrary to ZnSe the compound ZnTe can be heavily p-type doped. There are indications however, that ZnSe/ZnTe contacts degrade with time. In the following we shall present our efforts to solve the problem of the ohmic contact to p-type ZnSe.

In Fig. 1 one can recognize that GaN with a band gap of 3.4 eV has almost the same lattice constant as AlN with a gap of 6.2 eV. Consequently, pseudomorphic epitaxial growth of mixed crystals is feasible provided that suitable lattice matched substrates are available. However, no such substrate material is available and commonly sapphire is used with a lattice mismatch as large as 13.5%. Silicon-carbide with a mismatch of about 3% seems more suitable but unfortunately, high quality substrates are hard to obtain. The large difference in the lattice constant between GaN and sapphire results in a high concentration of misfit dislocations in thin films. Surprisingly, dislocation densities as high as $10^{10} cm^{-2}$ do not seem to degrade the light emission of GaN LEDs

with GaInN quantum wells substantially. Moreover, during operation of the devices the dislocations seem to be stable, which is attributed to the large bonding strength of the nitrides. In the meantime GaN-based LEDs have a lifetime exceeding 10.000 h and recently blue laser diodes (LDs) emitting around 400 nm have achieved a lifetime of several thousand hours. Recently, Nichia reported extrapolated lifetimes of laser diodes exceeding 10.000 h. Two new approaches have been followed in order to solve the substrate problem: epitaxial lateral overgrowth, and the use of quasi-GaN substrates, fabricated by removing the sapphire, leaving a thick MOCVD grown GaN film as a homoepitaxial substrate for the subsequent second epitaxy of the GaN device structure [16]. Laser devices with a threshold current density of $1.5 \, kA/cm^2$ could be fabricated.

The GaN/GaInN laser diodes were developed in only 2 y after the LEDs became available, which is quite a remarkable achievement. For GaAs LDs the development took of the order of 10 yr, with numerous groups working on the problem worldwide.

3. An alternative approach

In Würzburg we have been looking for an alternative approach to solve the existing degradation problems of ZnSe based lasers. It is well known that due to the relatively large ionic part of the chemical bond in II–VI semiconductors, these materials are softer compared to materials with covalent or predominantly covalent bonds. II–VI compounds with a larger part of covalent bonding than ZnSe should lead to a smaller number of dislocations and to a reduced mobility of misfit dislocations and extended defects. Fortunately, such compounds exist. Verié checked the bonding and elastic properties of Be-chalcogenides and concluded that the shear modulus of BeSe is higher than that of GaAs and that the shear modulus of BeTe and BeSe is considerably larger than that of ZnSe [17] Consequently, it was concluded that the Be compounds would be principally suitable materials to improve the performance of laser diodes. In Fig. 1 it can be seen that BeTe has a lattice constant which is only slightly smaller than that of GaAs (0.5%), so that it should be possible to grow pseudomorphic BeTe films. On the other hand the lattice constant of
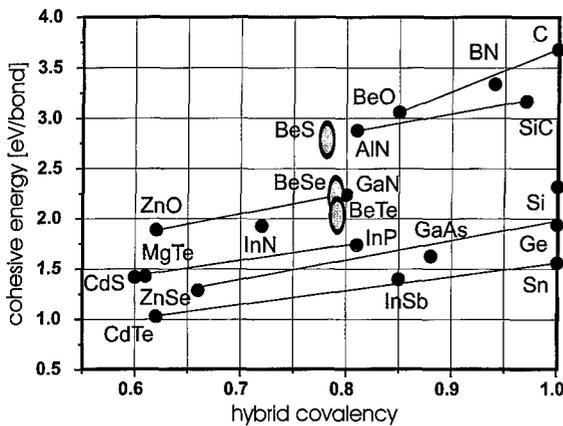
Fig. 2. Bond strength for various II–VI and III–V compound semiconductors.

ZnSe is only 0.27% larger than that of GaAs so that the MBE growth of pseudomorphic BeTe/ZnSe superlattices on GaAs substrates is possible. Especially interesting is BeSe with a lattice constant of 0.51 nm and a band gap of 5.6 eV. This allows the growth of BeMgZnSe films, which are lattice matched to GaAs.

In the following we shall review some of our work on Be-chalcogenides and on Be-containing II–VI mixed crystals as well as device oriented activities based on these new materials.

4. The binding properties of Be-chalcogenides

The bond strength of binary semiconductor compounds is expected to correlate with the degree of covalency. This can be seen in Fig. 2, where the cohesive energy per bond (bond energy) has been plotted versus the hybrid covalency. The latter term is used following Harrison [18]. Plotting the cohesive energy versus the hybrid covalency, one finds that both quantities are linearly related for isoelectronic compounds. This means that binary compounds with the same total number of electrons show a scaling behavior. Therefore, one can use Fig. 2 to estimate the bond energy of Be-chalcogenides by making use of their theoretically calculated covalencies. This was done for BeS, BeSe and BeTe. All other shown values for the cohesive energy per bond were experimentally determined and were taken from Ref. [18]. One can recognize in Fig. 2 that the bond energies of the Be-chalcogenides

are significantly larger than those obtained for the II–VI compounds CdTe and ZnSe. It should be noted that GaN has almost the same hybrid covalency as BeS, BeSe and BeTe and that its bond energy is hardly different from that of the Be compounds. It should also be noted that the bond strength of GaAs is below that of the Be-chalcogenides.

Experimentally, the high bond energy of Be-chalcogenides can be demonstrated by measuring the desorption of the compound in vacuum. The influence of the desorption can be determined by measuring the dependence of the growth rate of a binary compound as a function of substrate temperature [19]. The RHEED oscillation technique has been used to measure the growth rate of BeTe. It stays constant up to temperatures of 550°C, whereas the growth rates of ZnTe, ZnSe and CdTe start to decrease significantly at temperatures below 300°C. The data indicate that desorption of BeTe is not important for substrate temperatures as high as 550°C. Consequently, the MBE growth of BeTe is compatible to that of GaAs, and BeTe–GaAs heterostructures can be fabricated even at the high substrate temperatures typical for the epitaxy of high quality GaAs. In contrast to that, the growth rate of the conventional II–VI compounds decreases due to the desorption at much lower substrate temperatures, demonstrating that the bond energy of BeTe is high compared with that of the conventional II–VI compounds.

5. The properties of BeTe and BeSe

Only very few data on the basic properties of Be-compounds are available in the literature. It was clear from the beginning that BeSe has a large gap, measurements of the optical absorption gave a value of 5.6 eV at 300 K [20]. The large band gap allows to grow quaternary BeMgZnSe alloys with relatively small Be-concentration suitable as barriers in heterostructures with ZnSe quantum wells. With 8% Be and 12% Mg a barrier height of 300 meV can be achieved [21]. Microhardness measurements of ternary and quaternary alloys indicated an increased hardness with respect to ZnMgSSe mixtures for comparable composition [22]. The direct energy gap of BeTe was determined as 4.3 eV and an indirect gap of 2.8 eV. Using angle-resolved synchrotron-radiation photoemission

spectroscopy we have determined the dispersion of the valence band of BeTe along the [1 0 0] direction [23]. The measurements were analyzed using a first principles theoretical calculation of the Be band structure as well as the photoemission peaks, with very good agreement.

6. Ohmic p-contacts and BeTe–ZnSe superlattices

One severe problem of ZnSe-based laser diodes is the fabrication of good ohmic contacts to p-type ZnSe. As mentioned before, ZnTe–ZnSe pseudogradings have been used today in order to tune the valence band from the low lying ZnSe related position up to the ZnTe position, taking additional advantage of the high p-type dopability of ZnTe for a subsequent metal contact. The main disadvantage of this approach is the high lattice mismatch between these two materials, resulting in an extremely high defect density, as well as the absorption of laser light in the ZnTe, the band gap of which is smaller than the one of the active region. High defect densities are known to promote diffusion of both dopant atoms as well as host materials [24].

Replacing ZnTe by BeTe the problem of lattice mismatched graded contacts can be solved. Assuming a high lying valence band for BeTe, the possibility for the fabrication of BeTe–ZnSe graded contacts has already been discussed on a theoretical basis [25]. For the design of a graded band gap structure, the band offset is an important parameter.

Recently, we have determined the valence band offset (VBO) between BeTe and ZnSe experimentally as shown in Fig. 3 by both UV photoelectron spectroscopy [26] as well as photoluminescence (PL) [21]. Due to a type II band offset, this value can directly be measured by PL by utilizing the optical transition between the conduction band of ZnSe and the valence band of BeTe in ZnSe–BeTe quantum well structures. The spatially indirect transition is usually low in intensity, but sharp and detectable. The recombination energy is around 1.8 eV and the energetic difference between the ZnSe band gap photoluminescence and the indirect transition directly gives the BeTe–ZnSe valence band offset. This assumes, however, that the residual doping level in both the ZnSe and BeTe is low enough that band bending effects can be neglected (flat band conditions). Recent data based on photo-
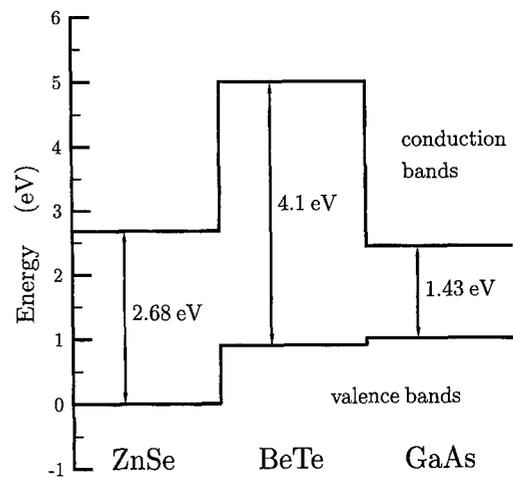
Fig. 3. Band offset of ZnSe, BeTe. The BeTe–ZnSe valence band offset has been determined by photoluminescence as described in the text.
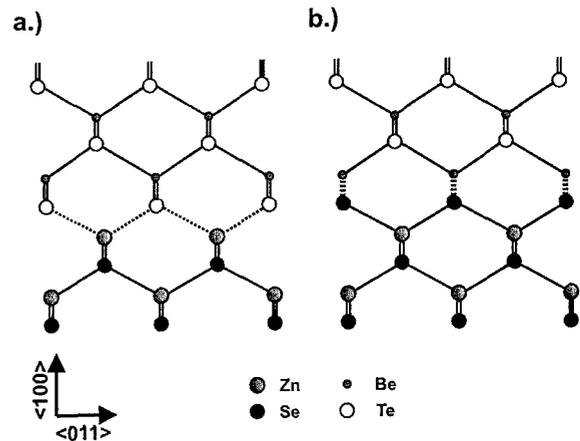
Fig. 4. Two possible configurations at a BeTe–ZnSe interface.

electron spectroscopy indicate, however, that the band offset can vary by as much as 0.8 eV due to the actual configuration of the interface between BeTe and ZnSe [26].

At a BeTe–ZnSe interface, two types of interface configurations can be demonstrated (a) Be–Te–Zn–Se, or (b)Te–Be–Se–Zn, (or, of course, a mixture of both). The two different situations are shown in Fig. 4. In case (a) the intermediate layer consists of ZnTe with a lattice mismatch of + 7%, in case (b) the intermediate layer is BeSe, with a lattice mismatch
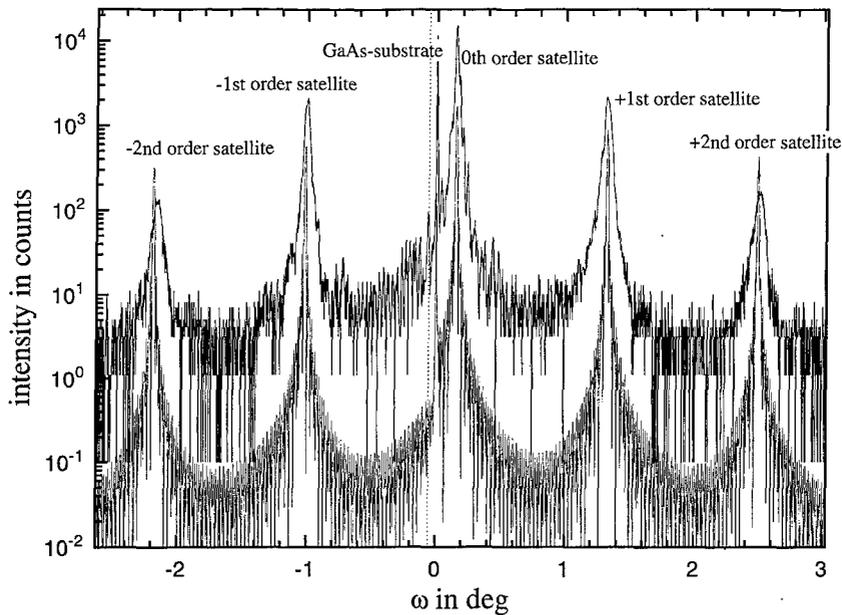
Fig. 5. High-resolution X-ray diffraction (upper curve) and simulation (lower curve) of a BeTe–ZnSe superlattice. The simulation curve has been shifted in y-direction for clarity. The structure used for the simulation was an 80 period stack of the sequence (1.0 Å BeSe, 28.5, Å ZnSe, 1.0 Å BeSe, 8.5 Å BeTe) grown on GaAs substrate.

of -10%. BeTe–ZnSe superlattices with a ZnTe interlayer have in general a high structural quality, whereas in superlattices with a BeSe interlayer, the interfaces are not as abrupt. In Fig. 5 a $(0\,0\,2)$ rocking curve for an MBE grown BeTe–ZnSe superlattice with BeSe interlayers has been plotted. In the figure, the simulation of the X-ray diffraction is also shown. The structure used for the simulation in Fig. 5 was an 80 period stack of the sequence (1.0 Å BeSe, 28.5 Å ZnSe, 1.0 Å BeSe, 8.5 Å BeTe) grown on GaAs substrate. The BeSe interlayer introduces a substantial amount of compressive strain and had to be taken into account in order to match the simulation and the experimental data.

Electron microscopy techniques were used to investigate a BeTe–ZnSe superlattice in detail. Scanning electron microscopy, conventional as well as high-resolution transmission electron microscopy (TEM) were used [27]. High-resolution micrographs were evaluated by correspondence analysis in order to quantify the abruptness of the chemical transition at the interfaces. It turned out that the transition region was of the order of 1 atomic layer. This can be visual-
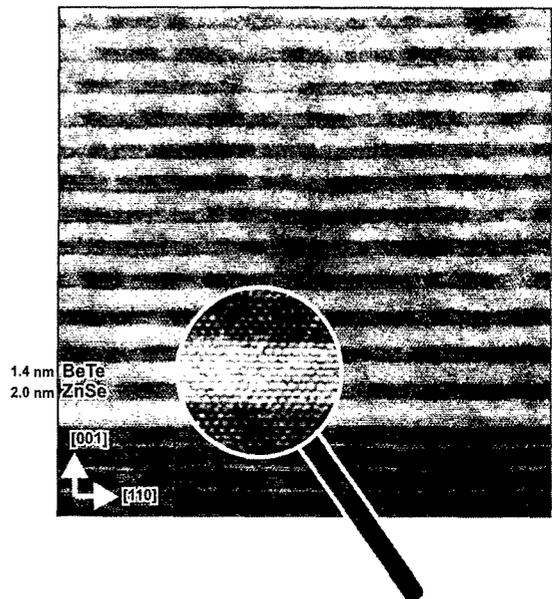
Fig. 6. High-resolution transmission electron microscopy of a BeTe–ZnSe superlattice.
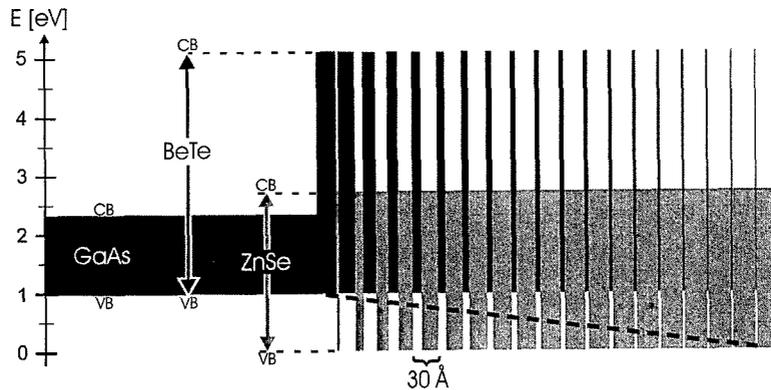
Fig. 7. Schematic representation of a BeTe–ZnSe pseudograding used as a p-type contact.

ized in Fig. 6. No misfit dislocations could be detected in TEM pictures with less resolution. The thickness of the superlattice (300 nm) was much larger than the calculated critical thickness of 60 nm. This indicates that a high barrier for the generation of dislocations must exist. Since it is known that misfit dislocations are easily nucleated in ZnSe/GaAs heterostructures, the observed effect must be attributed to the BeTe.

In addition to the lattice mismatch, ZnTe and BeSe have very different covalencies. This obviously influences the VBO, which as a consequence, the valence band offset also changes significantly with the interface configuration from 0.4 (BeSe interfacial layer) to 1.2 eV (ZnTe interfacial layer). This has unambiguously been determined by ultraviolet photoelectron spectroscopy experiments [26]. The interface termination was adjusted by variation of the MBE growth parameters.

The valence band offset between ZnSe and GaAs is around 1 eV [28] in the case of Zn-rich growth conditions. As a consequence there should be a quite close valence band matching between GaAs and BeTe provided that the transitivity rule for band offsets holds.

The basic principle of a BeTe–ZnSe pseudograding is shown in Fig. 7. The goal is to tune the valence band smoothly from the ZnSe valence band to the BeTe valence band, avoiding barriers in order to get an efficient current injection. For such a pseudograding the BeTe has to be heavily doped p-type. We demonstrated that this is possible by using a plasma
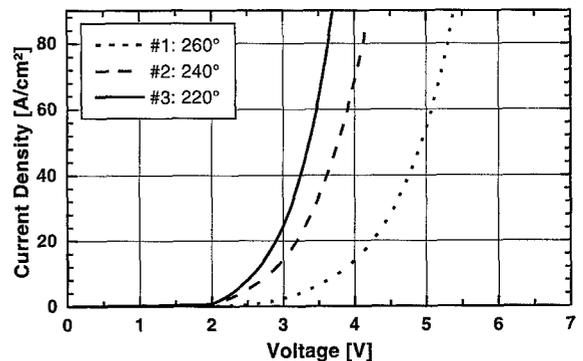
Fig. 8. Current voltage curves for graded band gap contacts grown at different substrate temperatures (dotted: 260°C, dashed: 240°C solid: 220°C).

activated nitrogen source and free hole concentrations up to 1×10^{20} cm^{-3} could be achieved, which was verified by both van der Pauw as well as infrared reflectivity measurements. The BeTe–ZnSe pseudograding concept allows a new laser structure in which the GaAs substrate is p-type. This configuration has the advantage, that the contact to the p-side of a laser device has a large area and moreover, favorable heat transfer properties. Fig. 8 shows current–voltage curves for graded contacts produced under various MBE growth conditions. It turns out that a low growth temperature is a very important parameter, reducing the operating voltage to below 5 V for currents up to 1 kA/cm^2. Recently, a group at 3M has reported stability tests of such BeTe–ZnSe contact structures [29]. The contact structure under study did not show any degradation

when operated at $2\,kA/cm^2$ – a factor 10 higher than the laser threshold – for 1200 h operating time at room temperature. This indicates that the contact problems, which have plagued ZnSe based laser diodes with p-type ZnSe–ZnTe contacts can be solved.

The possibility of a high p-type doping makes BeTe interesting for the p-type cladding, too. In ZnSeTe and ZnMgSeTe, it has been demonstrated that the incorporation of Te enhances the possibility for heavy p-type doping of the material drastically [30,31]. In both cases, however, the lattice matching to GaAs is sacrificed due to the large lattice constant of ZnTe. Using beryllium, the lattice constant can be modified, and lattice matched quaternary BeZnSeTe can be fabricated which is expected to allow heavy p-type doping. This aspect could also be interesting for the growth of laser diodes by metal organic chemical vapor deposition, where the p-type doping of ZnSe is still an unsolved problem. Metal-organic precursors for beryllium are available. One should mention that due to the type II band offset between BeTe and ZnSe, a pronounced band bowing is expected.

BeTe–ZnSe superlattices are also potential candidates for both waveguide and cladding regions of ZnSe-based laser devices. If the period is small enough, minibands are formed allowing an efficient vertical transport and recombination in spite of the type II band alignment between BeTe and ZnSe. However, no results on laser diodes using this approach have been reported so far.

7. Defect reduction

BeTe seems to be an ideal candidate for the growth start on GaAs. As described above, the valence bands can be expected to be nearly aligned. Secondly, a BeTe buffer can serve to passivate the GaAs surface, avoiding a surface reaction of Se and S with GaAs.

To achieve a comparison to reported methods for defect reduction a set of samples has been grown under similar conditions with the growth start being the only difference from run to run. The samples were BeMgZnSe epilayers sufficiently lattice matched to GaAs. The misfit between the BeMgZnSe layer and the GaAs substrate was determined by HRXRD to be about $-0.08\% \pm 0.02\%$. The substrate surfaces were prepared using various techniques with and without

employing a BeTe buffer. It was shown that thin BeTe buffer layers can reduce the density of lattice defects substantially [32]. The density of paired Frank type stacking faults could be reduced significantly to values below $10^2\,cm^{-2}$. It turned out that the Se background pressure played a crucial role for the defect nucleation. Residual defects with a density of about $5 \times 10^3\,cm^{-2}$ could be attributed to the existence of dislocations originating from the GaAs substrate and the GaAs buffer layer. BeTe can form a smooth interface to GaAs and to ZnSe, resulting in pronounced thickness interference effects observed in high-resolution X-ray diffraction experiments on BeTe films.

Though the defect densities reached so far are still somewhat higher than the lowest defect densities reported in the literature, we are convinced that the use of BeTe buffer layers is an alternative to the usual growth start of II–VI films on GaAs substrates.

8. Light emitting devices

The degradation mechanisms of II–VI laser diodes based on sulfur containing II–VI materials have been studied in detail over the last years. It turned out that – depending on defect densities – either extended defects originating mainly at the interface between GaAs and ZnSe [33], or point defects in the active region are responsible for the early degradation of conventional laser diodes. The degradation has been identified to be enhanced by defect recombination of electron hole pairs [34]. In order to strengthen the active ZnCdSe region in laser diodes for an enhancement of the laser lifetimes, the additional incorporation of Be into the quantum well suggests itself resulting in a quaternary BeZnCdSe active region. In addition, the ternary BeZnSe seems to be a natural, easy to grow and lattice matched waveguide material with a suitable band gap. Therefore, first attempts to fabricate Be-containing laser diodes have been based on the structures shown in Fig. 9. It has been briefly mentioned already that, due to the pseudomorphic contact structure, both p-type and n-type GaAs substrates can now be used. For the structure with a p-type substrate, there are some possible potential advantages: a current spreading through the p-type cladding would lead to a lower current density across the contact structure at the bottom of the struc-

CB330

p⁺-BeTe 20nm

BeZnCdSe 7nm

n⁺-ZnSe 50nm

CB335

BeZnCdSe 7nm

p⁺-BeTe 20nm
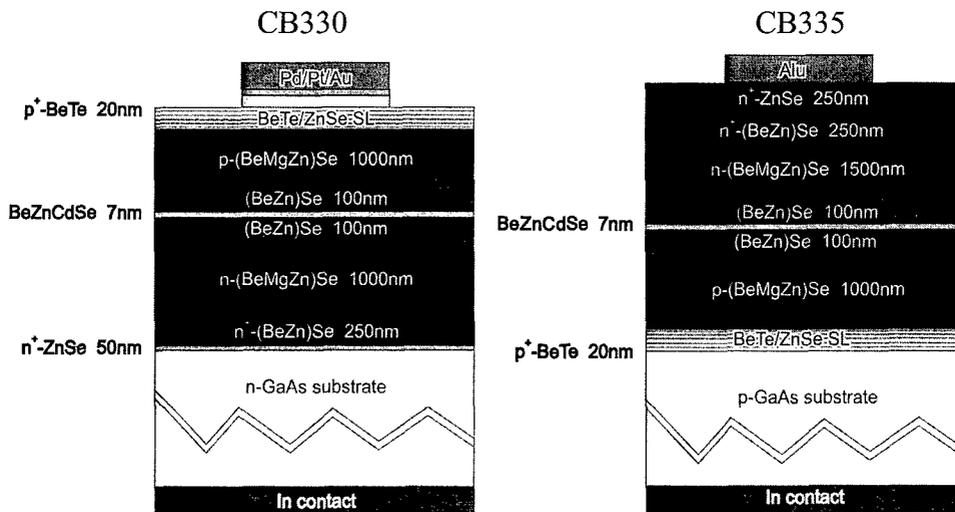
Fig. 9. Schematic structures of the first Be-containing laser diodes, grown on both p-type and n-type GaAs substrates.

ture towards the GaAs substrate. In addition – during processing – an n-type contact can then be fabricated on a degenerately doped n-type top ZnSe layer, which is much easier from a technological point of view.

For the efficient operation of a II–VI laser diode it is essential that the quantum well built into the heterostructure has to have a high quality. The ZnCdSe quantum wells, embedded in BeMgZnSe waveguides which we have used in some of our devices, have been optimized recently. In Fig. 10, the temperature dependence of the photoluminescence intensity has been plotted, indicating a factor of 50 between low temperature and room temperature photoluminescence. It is expected that the incorporation of Be in the quantum well will additionally harden the material without deteriorating the efficiency. The emission of a BeMgZnSe laser diode below, at and above threshold is shown in Fig. 11. It is interesting to note that first laser diodes grown on p-type and n-type GaAs had similar I–V characteristics. These first devices had a still quite high threshold current of about 700 A/cm² at 300 K. Meanwhile, Be-containing laser diodes with a lifetime of 57 h under CW operation at room temperature have been fabricated [29], the lifetime being close to the maximum lifetime of 100 h reported up to now for the best ZnMgSSe devices [12]. In view of the short development time for Be-lasers, this is certainly an exciting result and a proof
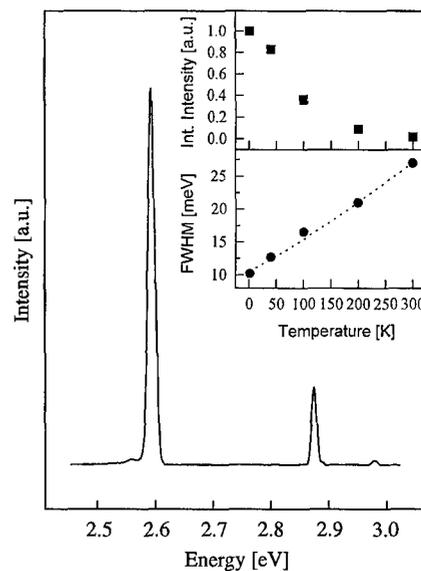
Fig. 10. Photoluminescence of an undoped laser structure with a ZnCdSe quantum well. The PL-line at 2.59 eV originates from the quantum well, the line at 2.88 eV from the BeZnSe waveguide, and the line at 2.98 eV from the BeMgZnSe cladding. The inset shows the temperature dependence of the integrated intensity as well as the line width.

of the feasibility of the beryllium concept. In addition, it should be noted that our Be laser diodes up to now use only ZnCdSe or BeZnCdSe (2.8% Be) quantum
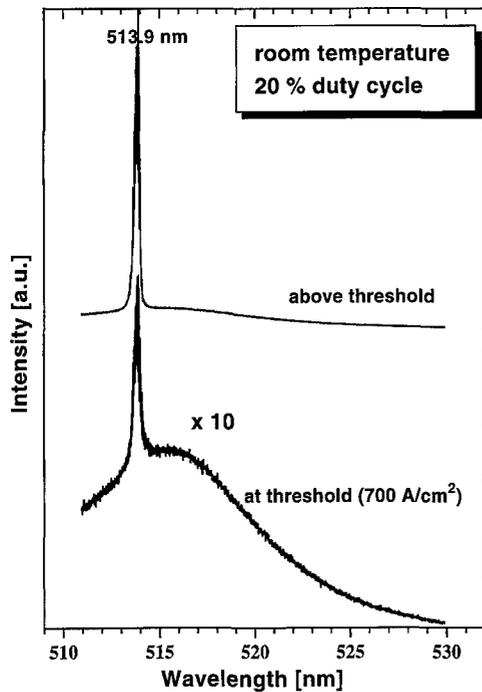
Fig. 11. Emission spectrum of a BeMgZnSe laser diode.

use of BeTe buffer layers can be used to improve the growth start of II–VI films in laser devices substantially. Due to the extraordinary properties of Be-chalcogenides there is hope that the lifetime problems, which have plagued II–VI laser diodes so far, can be solved in the nearest future. Because the ZnSe-based lasers are most efficient in the blue–green spectral range, which has not been covered by nitride-based lasers, it is probable that future applications of these two classes of devices will compliment each other.

wells. This is due to the fact that it is advantageous to keep the temperature of the Be furnace constant throughout the epitaxy of the waveguide and active region, avoiding a growth interruption when sensitive interfaces are formed. This implies that waveguide and active region then have the same low Be concentration of 2.8%, corresponding to a lattice matched ternary BeZnSe. Clearly, the Be concentration of the active region should be increased in the future to bring the larger covalency of Be compounds into play. Up to now only the good reproducibility of the lattice constant as well as the pseudomorphic and non-absorbing contact structure have been exploited.

9. Conclusions

In this review the potential of Be-chalcogenides for the design of blue–green laser diodes and other quantum well structures has been demonstrated. It was shown that by using graded BeTe–ZnSe superlattices there is a good chance to solve the present contact problems of ZnSe-based laser diodes. In addition the

References

[1] R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Solys, R.O. Carlson, Phys. Rev. Lett. 9 (1962) 366.
[2] M.I. Nathan, W.P. Dumke, G. Burns, F.D. Hill Jr., G. Lasher, Appl. Phys. Lett. 1 (1962) 62.
[3] H. Krömer, Proc. IEEE 51 (1963) 1573.
[4] H.S. Rupprecht, I.M. Woodall, G.D. Pettit, Appl. Phys. Lett. 11 (1967) 81.
[5] Zh.I. Alferov, V.M. Andreev, V.I. Korolkov, E.L. Portnoy, D.N. Tret'yakov, Fiz. Tekn. Polupr. 2 (1968) 1545.
[6] R.M. Park, M.B. Troffer, C.M. Rouleau, J.M. dePuydt, M.A. Haase, Appl. Phys. Lett. 57 (1990) 2127.
[7] H. Amano, M. Kito, K. Hiramatsu, I. Akasaki, Jpn. J. Appl. Phys. 28 (1989) L2112.
[8] S. Nakamura, N. Iwasa, M. Senoh, T. Mukai, Jpn. J. Appl. Phys. 31 (1992) 1258.
[9] M.A. Haase, J. Qui, J.M. dePuydt, H. Cheng, Appl. Phys. Lett. 59 (1991) 1272.
[10] H. Okuyama, T. Miyajima, Y. Morinaga, F. Hiei, M. Ozawa, K. Akimoto, Elec. Lett. 28 (1992) 1798.
[11] G.C. Hua, N. Otsuka, D.C. Grillo, Y. Fan, J. Han, M.R. Ringle, R.L. Gunshor, M. Hovinen, A.V. Nurmikko, Appl. Phys. Lett. 65 (1994) 1331.
[12] A. Ishibashi, M. Ukito, S. Tomiya, Proc. Int. Conf. Physics of Semiconductors, Berlin, 1996.
[13] S. Itoh et al., Int. Conf. II–VI Semiconductors, Grenoble, 1997, J. Crystal Growth, in press.

[14] Y. Fan, J. Han, L. He, J. Saraie, R.L. Gunshor, M. Hagerott, H. Jeon, V. Nurmikko, G.C. Hua, N. Otsuka, Appl. Phys. Lett. 61 (1992) 3160.

[15] A. Ishibashi, Y. Mori, J. Crystal Growth 138 (1994) 677.

[16] Nakamura et al., SPIE Photonics West, San Jose, 1998.

[17] C. Verié, in: B. Gil, R.-L. Aulombard (Eds.), Semiconductor Heteroepitaxy, World Scientific, Singapore, 1995, p. 2145.

[18] W.A. Harrison, Electronic Structure and the Properties of Solids, Dover, New York, 1989.

[19] T. Litz, H.-J. Lugauer, F. Fischer, U. Zehnder, U. Lunz, T. Gerhard, H. Ress, A. Waag, G. Landwehr, EMRS Spring Meeting, Strasbourg, 1996 Mater. Sci. Eng. B43 (1997) 83.

[20] V. Wagner, R. Kruse, K. Wilmers, M. Keim, H.-J. Lugauer, F. Fischer, A. Waag, G. Landwehr, J. Geurts, to be published.

[21] A. Waag, F. Fischer, H.J. Lugauer, Th. Litz, J. Laubender, U. Lunz, U. Zehnder, W. Ossau, T. Gerhard, M. Möller, G. Landwehr, J. Appl. Phys. 80 (1996) 792.

[22] A. Waag, F. Fischer, H.-J. Lugauer, Th. Litz, T. Gerhard, J. Nürnberger, U. Lunz, Z. Zehnder, W. Ossau, G. Landwehr, B. Roos, H. Richter, Symp. C, Proc. EMRS Spring Meeting 1996, Strasbourg.

[23] M. Nagelstraßer, H. Dröge, F. Fischer, T. Litz, A. Waag, G. Landwehr, A. Fleszar, W. Hanke, H.-P. Steinrück, Phys. Rev., accepted for publication.

[24] Th. Baron, Thesis, University of Grenoble, 1996.

[25] P.M. Mensz, Appl. Phys. Lett. 64 (1994) 2148.

[26] M. Nagelstraßer, H. Dröge, F. Fischer, T. Litz, A. Waag, G. Landwehr, H.-P. Steinrück, Phys. Rev., submitted.

[27] Th. Walter, A. Rosenauer, D. Gerthsen, F. Fischer, R. Gall, Th. Litz, A. Waag, G. Landwehr, Microscopy of semiconducting materials, Oxford, 7–10 April 1997.

[28] A. Franciosi, L. Vanzetti, L. Sorba, A. Bonnani, R. Cingolani, M. Lomascolo, D. Greco, Mater. Sci. Forum 182–184 (1995) 17.

[29] D. Grillo et al., Proc. SPIE Photonics West, San Jose, 1998.

[30] W. Faschinger, S. Ferreira, H. Sitter, R. Krump, G. Brunthaler, Mater. Sci. Forum 182–184 (1995) 29.

[31] R. Krump, S.O. Ferreira, W. Faschinger, G. Brunthaler, H. Sitter, Mater. Sci. Forum 182–184 (1995) 349.

[32] F. Fischer, M. Keller, T. Gerhard, T. Litz, H.J. Lugauer, M. Keim, G. Reuscher, T. Baron, A. Waag, G. Landwehr, J. Appl. Phys., accepted for publication.

[33] S. Guha, J.M. de Puydt, M.A. Haase, J. Qiu, H. Cheng, Appl. Phys. Lett. 63 (1993) 3107.

[34] A. Ishibashi, M. Ukito, S. Tomiya, in: M. Scheffler, R. Zimmermann (Eds.), Proc. 23rd Int. Conf. Physics of Semiconductors, Berlin, 1996, p. 3155.

Ferromagnetic transition induced by a two-dimensional hole gas in semimagnetic quantum wells

Y. Merle d'Aubigné [a,*], A. Arnoult [a], J. Cibert [a], T. Dietl [a,b], A. Haury [a],
P. Kossacki [a,c], S. Tatarenko [a], A. Wasiela [a]

[a] *Laboratoire de Spectrométrie Physique, Université Joseph Fourier Grenoble 1 - CNRS (UMR 55 88), Boîte Postale 87,
F-38402 Saint Martin d'Hères Cedex, France*
[b] *Institute of Physics, Polish Academy of Sciences, al. Lotników 32/46, PL-02668 Warszawa, Poland*
[c] *Institute of Experimental Physics, Warsaw University, ul. Hoża 69, PL-00681 Warszawa, Poland*

Abstract

A review is given on the effects of the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction between localized spins in various dimensionality systems of doped diluted magnetic semiconductors (DMS). Since this interaction is long-range, its influence on the temperature and magnetic field dependencies of magnetization and spin splitting of the bands is evaluated in the mean-field approximation, but by taking into considerations disorder-modified carrier–carrier interactions. The results show that the hole densities, which can presently be achieved, are sufficiently high to drive a paramagnetic–ferromagnetic phase transition in bulk and modulation-doped structures of II–VI DMS. Our recent photoluminescence studies on MBE-grown samples, containing a single, modulation-doped, 8 nm quantum well of $Cd_{1-x}Mn_xTe/Cd_{1-y-z}Mg_yZn_zTe:N$ are discussed. These studies reveal the presence of a ferromagnetic transition induced by the two-dimensional hole gas. The transition occurs between 1.8 and 2.5 K, depending on the Mn concentration x, in agreement with the theoretical model. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

It has been known for a long time that the compensation of antiferromagnetic interactions between the localized spins by a ferromagnetic coupling would result in a dramatic enhancement of the sensitivity of

diluted magnetic semiconductors (DMS) [1] to the temperature and the magnetic field, particularly in the vicinity of the ferromagnetic phase transition. Search for the ferromagnetic transition has been successful in the case of Mn-based p-type IV–VI [2] and III–V compounds [3], in the latter case a critical temperature as high as 110 K has recently been reported for $Ga_{0.93}Mn_{0.07}As$ [4]. In the case of II–VI DMS, a tight binding model [5] suggests that the superexchange in

* Corresponding author. Fax: +33 476 514544; e-mail: merle@spectro.grenet.fr.

Cr-based DMS might be dominated by a *ferromagnetic* contribution. Accordingly, an attempt has been undertaken [6] to overcome the well-known small solubility of Cr in II–VI compounds by means of MBE growth of $Cd_{1-x}Cr_xTe$.

Recent years have witnessed a rapid progress in doping of II–VI wide gap semiconductors by *substitutional* impurities. For instance, electron and hole concentrations in the excess of 10^{19} cm^{-3} were reported for ZnSe : I [7] and ZnTe : N [8]. At the same time, modulation doping of II–VI quantum wells by either electrons [9] or holes [10] as well as patterning of conducting quantum wires [11] has successfully been performed. Motivated by this progress some of the present authors [12] undertook a theoretical analysis of the nature and strength of the carrier-mediated spin–spin interactions in bulk, layered, and nanostructured II–VI compounds. Results of that paper suggested that even for the highest available *electron* density no transition to the ferromagnetic phase is expected above 1 K. By contrast, such a transition was predicted for p-type materials, either in the bulk or modulation-doped form. In a subsequent work [13,14], we demonstrated that indeed the hole liquid in a single modulation-doped quantum well of $Cd_{1-x}Mn_xTe/Cd_{1-y-z}Mg_yZn_zTe : N$ induces the foreseen transition. Our findings mean, therefore, that the well established methods of modulation of the carrier concentration in semiconductor quantum structures can be applied for tailoring of the magnetic properties. Furthermore, a comparison of the experimental results with the theoretical predictions revealed the quantitative importance of the interactions between the holes as well as demonstrated why the mean field approximation is valid.

This paper is organized as follows. In Section 2 we describe the theoretical model [12]. We discuss two alternative mean-field approaches to the carrier-mediated spin–spin interactions. The first starts from the usual RKKY model, whereas in the second one molecular fields acting on the carrier and the spin subsystems are treated in a self-consistent manner. We then show that the both approaches lead to the same results. The growth procedure and characterization of the structures [10] are presented in Section 3. In Section 4 we discuss experimental evidences for the ferromagnetic transition [13,14]. Principal conclusions of our work are summarized in Section 5.

2. Theoretical model

2.1. Carrier mediated ferromagnetic interaction

Through this section we consider the influence of delocalized or weakly localized carriers on the interaction between magnetic ions. The dimensionality d of the subsystem of the carriers is determined by the shape of the potential $V(\zeta)$ that leads to their confinement. Accordingly, the case $d = 1$ or $d = 2$ corresponds to a two- or one-dimensional potential well, respectively. Because of a short magnetic correlation length [1] the localized magnetic moments $g\mu_B S_i$ are assumed, to form a macroscopic 3D system. Thus, according to experimental studies [1,15] their magnetization in the absence of the carriers can be described by a modified Brillouin function, $M(T,H) = g\mu_B \tilde{x}N_0 S B_S(T + T_{AF}, H)$, where effective spin concentration $\tilde{x}N_0 < xN_0$ and temperature $T + T_{AF} > T$ account for the influence of antiferromagnetic superexchange interactions [1,5,15].

We begin by considering the Ruderman–Kittel–Kasuya–Yosida (RKKY) model, which provides the energy J_{ij} of the exchange coupling, $H_{ij} = -J_{ij}S_iS_j$, between two spins located at R_i and R_j as a function of the density-of-states of the carriers at the Fermi level, $\rho_d(\varepsilon_F) = \pi^{1-d}(2/\pi)^{(d-2)(d-3)/2}m^*k_F^{d-2}/\hbar^2$, and the exchange integral I of their interaction with the spins, $H_i = -Is S_i\delta(R - R_i)$. Following the well-known procedure [16] and adopting the one-band effective-mass approximation we obtain to the second order in I

$$J_{ij} = \frac{\rho_d(\varepsilon_F)k_F^d I^2}{2\pi}\mathscr{F}_d(2k_F|r_i - r_j|)|\varphi_0(\zeta_i)|^2|\varphi_0(\zeta_j)|^2.$$

(1)

Here r is the vector in the d-dimensional space, $\varphi_0(\zeta)$ is the ground-state envelope function of the carriers in the confining potential $V(\zeta)$, and

$$\mathscr{F}_1(y) = -\pi\text{si}(y)/2,$$ (2a)

$$\mathscr{F}_2(y) = \int_1^{-} dt\frac{\mathscr{J}_1(yt)}{yt(t^2 - 1)^{1/2}},$$ (2b)

$$\mathscr{F}_3(y) = [\sin(y) - y\cos(y)]/y^4,$$ (2c)

where si(y) is the sine-integral and $\mathscr{J}_n(y)$ is the Bessel function. The asymptotic behavior of $\mathscr{F}_d(y)$ for large

y is $\pi\cos(y)/2y$, $\sin(y)/y^2$, and $-\cos(y)/y^3$, while for $y \to 0$, $\mathscr{F}_d$ tends to $\pi/4$, $[1/2 - \gamma + \ln(4/y)]/2$, and $3/y$ for $d = 1, 2$, and 3, respectively, where $\gamma = 0.57721\ldots$ is the Euler constant. For the same sequence of d, the first zero of $\mathscr{F}_d(y)$ occurs for $y \approx 1.7, 3.5$, and 4.5, respectively. The formula for $d = 3$ reproduces the result first obtained by Ruderman and Kittel [17], and for $d = 1$ that of Yafet [18], whereas in the case of $d = 2$ only the asymptotic form for $y \to \infty$ has so far been considered in the literature [16].

Knowing the dependence J_{ij} on the distance between the spins r we can calculate the mean-field value of the Curie–Weiss temperature of spins located at ζ_i, $\Theta(\zeta_i) = S(S+1)\sum_j J_{ij}/3k_B$. Since in semiconductors, in contrast to metals, the value of y that corresponds to the distance r between the nearest neighbor spins is much smaller than the period of the oscillatory functions in Eqs. (2a), (2b) and (2c), $y_{nn} = 2k_F r_{nn} \ll 1$, we replace the summation over the ion positions by an integration extending from $y = 0$ to ∞. Under the assumption that the distribution of the magnetic ions is random we obtain Θ, averaged over ζ, in the form

$$\Theta = S(S+1)N_0 I^2 \rho_d(\varepsilon_F) \int d\zeta \tilde{x}(\zeta) |\varphi_0(\zeta)|^4 / 12k_B, \quad (3)$$

which shows that the net RKKY interaction is ferromagnetic, $\Theta > 0$. This is in contrast to the case of metals where a spin glass phase is observed. In semiconductors, however, where the mean distance between magnetic ions $\tilde{r} = (4\pi\tilde{x}N_0/3)^{-1/3}$, is much smaller that the electron wavelength $\lambda_F = 2\pi/k_F$, the spin–spin interaction mediated by the carriers is merely ferromagnetic. Since Θ is proportional to the effective mass, to the degree of confinement as well as to the square of the exchange integral I we expect much greater magnitudes of Θ in the presence of the holes than for the electrons in II–VI DMS.

The above approach neglects intervalley or intersubband virtual transitions. Since those terms in J_{ij}, which result from such transitions contain products of orthogonal envelope functions, $\varphi_\nu^*(\zeta_j)\varphi_{\nu'}(\zeta_j)$, their contribution to Θ vanishes, provided that the concentration of magnetic ions is uniform. Moreover, in such a case each of the occupied valleys or subbands gives an independent contribution to Θ, described by the relevant $\varphi_\nu(\zeta)$ and $k_F^{(\nu)}$.

2.2. Carrier–carrier interaction and disorder effects

So far we have disregarded the influence of the potential scattering and the Coulomb interactions among the carriers upon the magnitude of Θ. The former is known [20] to introduce a random phase shift in the oscillatory functions of Eq. (1). This random phase shift has no effect on the magnitude of the second moment of the distribution of J_{ij} [21] but, after averaging over the disorder, leads to the dumping $\exp(-r/l)$ of the first moment [20,21], where l is the mean free path for elastic collisions. The corresponding reduction factor of Θ to the lowest order in $1/k_F l$ is given by $1 - 1/12(k_F l)^2$, $1 - \pi/8k_F l$, and $1 - \pi/4k_F l$ for $d = 1, 2$, and 3, respectively. Hence, the effect of disorder becomes important on approaching the strongly localized regime, $k_F l \to 1$. In this range, however, the magnetic susceptibility of the carriers becomes substantially enlarged by the disorder-modified electron–electron interactions [19]. The resulting enhancement factor of Θ, originating from the RKKY interactions between the spins located at $r > l$ [22], can be written in the form: $A_F = 1 + FL_s/l$, $1 + 2F\ln(L_s/l)/\pi k_F l$, and $1 + 3F(1 - l/L_s)/(2k_F l)^2$, respectively. Here F is the effective Coulomb amplitude, which becomes greater than 1 for $k_F l \to 1$; $L_s = \sqrt{(Dt_s)} \gg l$ is the spin diffusion length, where $D = k_F l/dm^*$ and t_s is the spin-disorder scattering time [19,23]. It is worth noting that A_F is approximately equal to $1 + g_3$, where g_3 is a parameter that controls the spin-splitting-induced positive magnetoresistance in disordered systems [19]. For example, by taking g_3 determined for bulk n-$Cd_{0.95}Mn_{0.05}Se$ [24] and superlattices of n-$Si/Si_{0.5}Ge_{0.5}$ [25] we obtain $A_F \approx 2.3$ and 2.5, respectively. We conclude that, in fact, the disorder is expected to *enlarge* the temperature of the ferromagnetic interaction in semiconductor structures. This enhancement may also be important in pure 1D systems, where the interaction-driven separation of the charge and spin degrees of freedom modifies $J(r)$ [26]. In the zero-dimensional case such as quantum dots both correlation effects and the fluctuations of magnetization associated with the finite volume visited by the carriers are of the paramount importance. A variant of the theory developed for bound magnetic polarons [27] should be applied for those systems.

2.3. Magnetic susceptibility: an alternative approach

In addition to Θ, it is interesting to determine the temperature and magnetic field dependencies of the magnetization of the localized spins, $M(T,H)$ and the spin-splitting of the relevant band,

$$\Delta(T,H) = I \int d\zeta \, M(T,H,\zeta) |\varphi_0(\zeta)|^2 / g\mu_B + g^* \mu_B H, \tag{4}$$

where g and g^* are the Landé factors of the localized spins and the carrier spins, respectively. In the mean-field approximation [28] M is induced by the external field H and the molecular field produced by the carriers, H^* so that $M(T,H) = g\mu_B \tilde{x} N_0 S B_S(T + T_{AF}, H + H^*)$, where $H^* = I(n_\downarrow - n_\uparrow)/2g\mu_B$. Here $n_{\downarrow,\uparrow}$ is the density of spin down and spin up carriers, respectively, which, at $\varepsilon_F \gg k_B T$, is given by $n_{\uparrow,\downarrow} = 1/2 |\varphi_0(\zeta)|^2 \int_{\pm\Delta/2}^{\varepsilon_F} d\varepsilon \, \rho_d(\varepsilon)$, where the dependence of ε_F on Δ is to be determined from the condition $n_\uparrow + n_\downarrow = n$, with n being the total carrier density. The effects of disorder and electron–electron interactions can also be incorporated into this formalism by taking into account collision broadening of the density of states as well as by determining H^* from $g\mu_B H^* = I\partial\Omega/\partial\Delta$ with the Gibbs free energy of the carriers, Ω containing effects of the disorder-modified electron–electron interactions [19].

The above set of coupled equations makes it possible to determine in a selfconsistent way the mean-field values of Δ as functions of temperature and magnetic field. In particular, in order to evaluate $\Delta(T,H)$ above the phase transition we take into account the terms which are linear in $H + H^*$. Such a procedure leads to,

$$\Delta(T,H)$$
$$= \frac{\Delta_0(T,H)}{[1 - \frac{I^2 A_F \rho_d(\varepsilon_F)}{4g^2 \mu_B^2} \int d\zeta \, \chi_0(T,\zeta) |\varphi_0(\zeta)|^4]}, \tag{5}$$

where $\Delta_0(T,H)$ and $\chi_0(T)$ are the spin-splitting and the magnetic susceptibility in the absence of the free carriers.

We see that for the magnetic susceptibility of the form $\chi_0(T) = C/(T + T_{AF})$, $\Delta(T)$ diverges at $T_c = \Theta - T_{AF}$, where Θ turns out to coincide with the

Curie–Weiss temperature determined by the RKKY interactions, displayed for $A_F = 1$ in Eq. (3). This quantitative agreement between the two approaches constitutes a remarkable result which is valid whatever is the dimensionality (1, 2 or 3D).

2.4. Nature of the phase transition

The long range character of the ferromagnetic interaction, $\lambda_F \gg \bar{r}$, has also important consequences for the nature of the phase transition. It has been shown [29] that as long as $\sigma < d/2$ in the dependence $J(r) \sim 1/r^{d+\sigma}$, the mean-field approach to the long wavelength susceptibility $\chi(T)$ is valid, a conclusion non-affected presumably by disorder in the spin distribution. In contrast, the critical exponents [29] $\eta = 2 - \sigma$ and $\nu = 1/\sigma$ point to much faster decay of $\chi(q)$ with q than that expected from the classical Ornstein–Zernike theory [30]. This means that our mean-field model of the ferromagnetic transition driven by the RKKY interactions should remain valid down to at least $|T - T_c|/T_c \approx (\bar{r}k_F)^2$. At the same time, unlike the case of short-range interactions, the length scale of magnetic correlation is set by λ_F, not by $\bar{r}$. This means that for $\lambda_F \gg \bar{r}$ the critical fluctuations of magnetization will strongly be suppressed, an expectation corroborated by the virtual absence of critical scattering of the carriers by the Mn spins in bulk $Pb_{1-x-y}Mn_xSn_yTe$ [31], where the free holes drive a ferromagnetic transition [2], and its presence in, e.g., EuS : Gd and EuO : Gd [32], in which a short-range ferromagnetic interaction between the Eu spins dominates [33]. At the same time, the suppression of the critical fluctuations is expected to make possible to probe directly the magnetic long-range order by means of inter- or intraband magnetospectroscopy.

2.5. Application to CdMnTe

Fig. 1 shows the magnitude of $\Theta(x)$ for p-type and n-type bulk $Cd_{1-x}Mn_xTe$, calculated by Eq. (3), with the values of $\tilde{x}(x)$ and $T_{AF}(x)$, as determined by Gaj et al. [15] at $T = 1.7$ K as well as by taking $m^* = 0.8$ and $0.1m_0$ as well as $IN_0 \equiv \beta N_0 = -0.88$ and $IN_0 \equiv \alpha N_0 = 0.22$ eV for the holes and the electrons, respectively [15,34]. It was noted that the spin–orbit interaction reduces the spin-splitting of the Γ_8 heavy holes at the Fermi level according to [35]
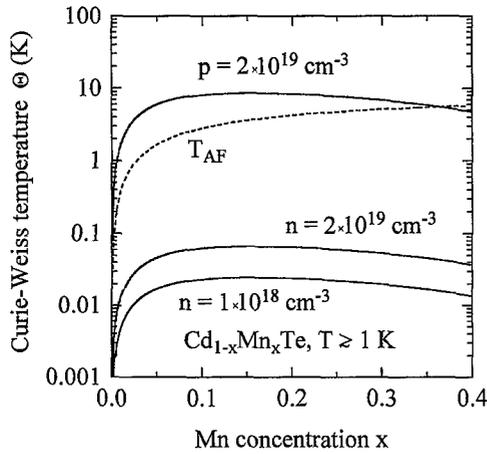
Fig. 1. Mean-field value of the Curie–Weiss temperature Θ calculated from Eq. (3) for p- and n-type $Cd_{1-x}Mn_xTe$, compared to antiferromagnetic temperature $T_{AF}(x)$. Ferromagnetic phase transition occurs at $T_c = \Theta - T_{AF}$. Material parameters as determined at 1.7 K together with the enhancement factor $A_F = 1$ were adopted for the calculation (after Ref. [12]).

$\Delta(k) = I|Mk|/g\mu_B k$, which after angular averaging results in an effective spin density of states $\tilde{\rho}_F = 1/2\rho_F$. We see in Fig. 1 that for sufficiently high hole concentrations so that the holes remain delocalized[1] [36], $\Theta > T_{AF}$ in a wide range of Mn concentrations. Furthermore, since for the above parameters the Kondo temperature $T_K \approx \varepsilon_F \exp[-1/(3|I|\tilde{\rho}_F)] = 1.1$ K, a crossover to the Kondo regime, $T_K > \Theta, T_{AF}$ will take place at relatively low Mn concentrations, $x < 1\%$. These considerations suggest, therefore, that a ferromagnetic phase transition can occur above 1 K in p+ II–VI compounds. This is in contrast to the case of n-type doping, for which no ferromagnetic phase transition is expected above 1 K, as shown by the two bottom curves in Fig. 1. It is worth noting, however, that on lowering temperature T_{AF} decreases, especially for low Mn concentrations [37]. Indeed, for the highest value of χ_0 ever reported for any DMS [37], that is $\chi_0 = 4.8 \times 10^{-3}$ emu/g for $Cd_{0.99}Mn_{0.01}Se$ at 15 mK, a ferromagnetic phase transition is predicted from Eq. (5) for material parameters of n-$Cd_{1-x}Mn_xSe$, [27] provided that $n \geqslant 1.5 \times 10^{19}$ cm^{-3}. Since $I > 0$, there is no Kondo effect for the electrons in DMS.

[1] We note that the formation of bound magnetic polarons shifts the delocalization transition towards higher carrier concentrations; see, e.g., Ref. [36].

Turning to the case of holes in DMS quantum wells we note that the confinement and possibly the biaxial strain lead to a splitting of the heavy and light hole bands [38] as well as to a strong anisotropy of the spin-splitting [39]. Actually, the coupled system of the 2D holes and the Mn spins is Ising-like as the spin-splitting of the ground-state subband undergoes a maximum for the magnetization parallel to the growth axis. By taking parameters suitable for the uppermost heavy-hole subband in a quantum well of $Cd_{0.9}Mn_{0.1}Te$, $m_h^* = 0.25m_0$ and $|\varphi_0(z)|^2 = 1/L_W$, where $L_W = 50$ Å, we predict the ferromagnetic transition to occur at about 2 K for $A_F = 1$, independently of the hole areal concentration p, since in our model the 2D density of states does not vary with ε_F. At the same time, the saturation values of Δ and M do depend directly on p according to $\Delta_s(T) = IM(T, H_s^*)/g\mu_B$, where $H_s^* = Ip/2g\mu_B L_W$, a value of the order of 1 kOe for $p = 2 \times 10^{11}$ cm^{-2}.

Finally, we note that in the case of 1D structures Θ increases with *decreasing* the carrier concentration. This, together with the correlation effects discussed above demonstrate the outstanding properties of such systems.

3. Sample preparation and characterization

Our studies have been carried out on structures grown in a molecular beam epitaxy (MBE) chamber equipped with a home-designed electron cyclotron resonance (ECR) plasma cell as a nitrogen source. Prior to fabrication of the proper samples, doping characteristics of the barrier material $Cd_{1-y-z}Mg_yZn_zTe$ have been determined by means of the Hall effect, capacitance–voltage profiles, cathodoluminescence, and X-ray diffraction. It was found that by lowering the growth temperature down to 220–240°C it becomes possible to reduce the nitrogen induced diffusion of Mg atoms and to obtain hole concentrations up to 5×10^{17} cm^{-3} in $Cd_{1-y-z}Mg_yZn_zTe$ with $z = 0.07$ and y up to 27%.

The studied modulation-doped structures consist of a single 8 nm quantum well (QW) of $Cd_{1-x}Mn_xTe$ embedded in $Cd_{0.66}Mg_{0.27}Zn_{0.07}Te$ barriers grown coherently onto a (0 0 1) $Cd_{0.88}Zn_{0.12}Te$ substrate. Such a layout insures large confinement energies for the holes in the QW, minimizing at the same time the

effects of lattice mismatch. Nitrogen-doped region in the front barrier is at the distance of 20 nm (samples No. 1, 2, and 5) and of 10 nm (sample No. 3) to the QW. Furthermore, in order to reduce depleting effects, two additional nitrogen doped layers reside at the distance of 100 nm from the QW on both sides. The nominal hole concentrations in the doped structures, evaluated from a self consistent solution of the Poisson and Schrödinger equations, are 2×10^{11} and 3×10^{11} cm^{-2} for the two employed values of the spacer width. In the same way we obtain $\tilde{L}_W = 6.3$ nm as an effective width of the hole layer, which is related to the hole ground state envelope function $\varphi_0(z)$ by $\tilde{L}_W = 1/\int dz|\varphi_0(z)|^4$. For a control purpose, an undoped structure (Sample No. 4) was also grown and examined.

The actual samples have been characterized at 2 K and in the magnetic fields up to 110 kOe by photoluminescence (PL), its excitation spectra (PLE), and reflectivity. The latter was taken in the presence of additional white light, so that under stationary conditions the holes are neutralized by electron diffusion [42]. In such a case, we observe free excitons with a well-known [15] giant Zeeman splitting in the $Cd_{1-x}Mn_xTe$ QWs, whose quantitative description gives $x = 0.024 \pm 0.002$ for samples No. 1–4 and $x = 0.037 \pm 0.002$ for sample No. 5.

The presence of the delocalized hole gas in the modulation-doped QWs is assessed by a step-like shape of PLE spectra in the σ^+ Faraday polarization, for which optical transitions involving the majority-spin hole band is probed in DMS [1,15]. This is illustrated in Fig. 2, which also shows how the field-induced redistribution of the holes between $\pm|\frac{3}{2}\rangle$ heavy-hole subbands, and the corresponding phase-space-emptying, leads to the appearance of the *quasi*free σ^- exciton in the minority-spin band. At the same time, the Moss–Burstein shift E_{MB} between the energies of the PLE step and the PL maximum in the σ^+ polarization makes it possible to evaluate the hole concentration p. In the case of the full spin polarization we observe $E_{MB} = 11 \pm 2$ meV for the wider spacer, $L_s = 20$ nm (sample No. 1, No. 2, No. 5), while for $L_s = 10$ nm (sample No. 3) $E_{MB} = 23 \pm 2$ meV. For the bare values of the electron and the heavy hole in-plane effective masses, $m_e^* = 0.1m_0$ and $m_h^* = 0.25m_0$ [38], we obtain $p = (1.6 \pm 0.2) \times 10^{11}$ cm^{-2} and $p = (3.2 \pm 0.2)$
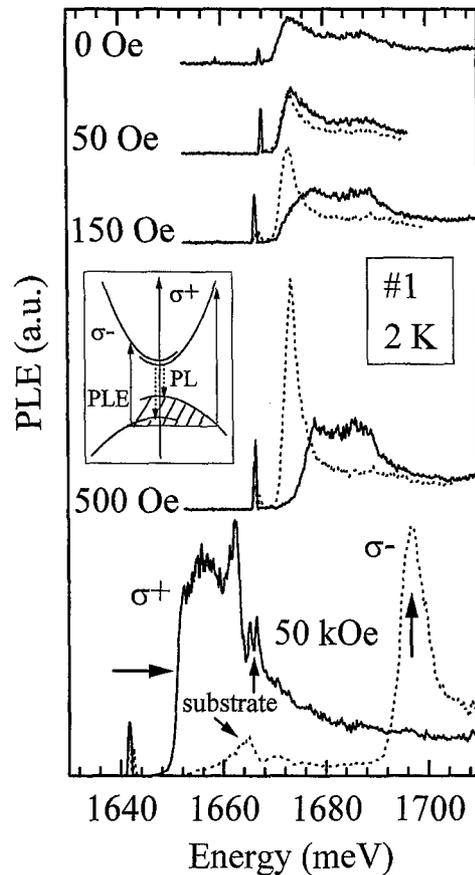
Fig. 2. PLE spectra for selected values of the magnetic fields in a modulation-doped p-type QW of $Cd_{0.976}Mn_{0.024}Te$ (sample No. 1) at 2 K for σ^+ (solid lines) and σ^- (dotted lines) circular polarizations. The PL was collected in σ^+ polarization at energies marked by the narrowest features. The sharp maximum and step-like form (denoted by large arrows in 50 kOe data) correspond to quasi-free exciton and transitions starting at the Fermi level, respectively. Substrate lines (denoted by small arrows) are superimposed on data for 50 kOe. The band arrangement at 150 Oe is sketched in the inset (after Ref. [13]).

$\times 10^{11}$ cm^{-2}, respectively, the values that compare favorably with those expected from the design of the structures. This may point out to a little importance of the renormalization of the Moss–Burstein shift by localization and hole–hole interactions in the case of their full-spin polarization. It is important to note that the formation of magnetic polarons is possible in DMS provided that disorder is sufficiently strong [40]. The absence of the hole localization indicates, therefore, high quality of the structures.
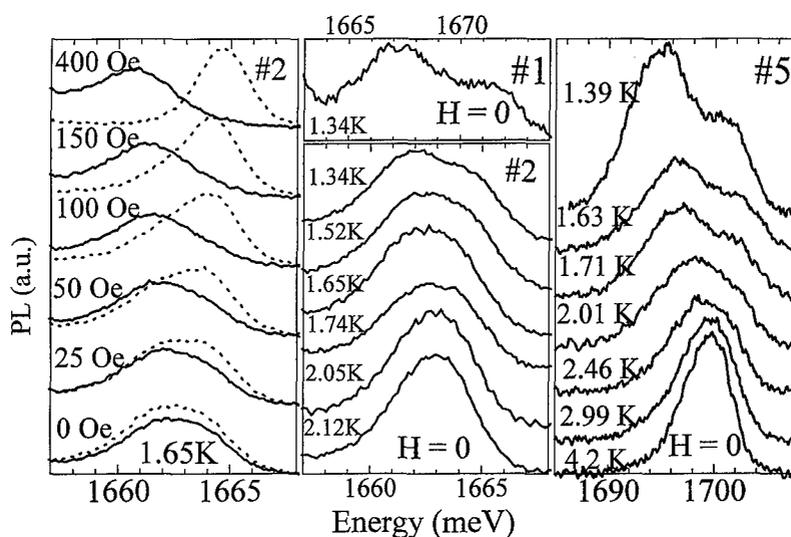
Fig. 3. Photoluminescence intensity in modulation-doped p-type QWs of $Cd_{1-x}Mn_xTe$ for selected values of the magnetic field H at 1.65 K (left panel, sample No. 2) or temperature at $H = 0$ (midle panel, samples No. 1 and 2; right panel No. 5). Solid and dotted lines in left panel correspond to σ^+ and σ^- circular polarizations, respectively. (after Refs. [13,14]).

4. Ferromagnetic transition

A thorough examination of the PL as a function of both temperature and the weak magnetic field has been carried out in a system equipped with copper Helmholtz coils and a rotary pump that allows reducing the temperature of the helium bath down to 1.3 K. Furthermore, in order to minimize heating of the hole and spin subsystems, the incoming photon energy was smaller than the barrier gap (direct excitation of the QW) and the laser power was kept below 50 μW per mm^2 of the illuminated surface. As shown in Figs. 3 and 4, the PL splitting Δ under such conditions is not only exceptionally large, but increases in a dramatic way on lowering temperature. Actually, in all doped samples, we observe a colossal value of $\partial\Delta/\partial H$ above a characteristic temperature T_c, and a zero field splitting below T_c, where $T_c = 1.8$ K for $x = 0.024$ (samples No. 1–3) and 2.5 K for $x = 0.037$ (sample No. 5). The corresponding experimental data are summarized in Fig. 3. The large value of $\partial\Delta/\partial H$ explains also a strong polarization of the hole liquid in very small magnetic fields witnessed by the PLE spectra of Fig. 2. No such effects are visible either in the undoped structure or in the presence of illumination by white light that depletes the QW from the carriers.

The appearance of zero-field splitting is preceded by a critical increase of χ, as shown in Fig. 4b for sample No. 2. We determine the susceptibility in the limit of vanishing fields as

$$\chi(T) = (g\mu_B/|\alpha - \beta|)(\partial\Delta/\partial H). \qquad (6)$$

The transition temperature deduced from the extrapolation of the susceptibility data, as shown in Fig. 4b, agrees well with the temperature deduced from the appearance of the spontaneous magnetization shown in Fig. 3 (middle and right panels). The critical behavior of $\chi(T)$ is in contrast to gradual changes of $\chi(T)$ associated with the formation of magnetic polarons [27,41], for which the finite volume involved precludes the existence of any second-order phase transition [40,27].

These findings are interpreted as a ferromagnetic phase transition driven by the free holes which by means of the RKKY mechanism mediate ferromagnetic exchange interactions between the Mn spins. This conclusion is strongly supported by the theoretical model discussed in Section 2. According to this model, the critical fluctuations of magnetization are of minor importance, so that we may write $\chi(T) = C/(T + T_{AF} - \Theta)$, where the influence of the delocalized carriers is described by Θ. According
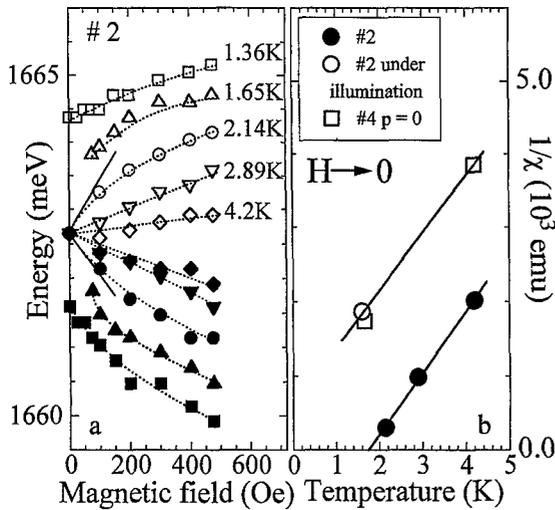
Fig. 4. (a) Energy $E_\pm$ of photoluminescence maxima at σ^+ and σ^- circular polarizations (filled and open symbols, respectively) as a function of the magnetic field H for selected temperatures in a modulation-doped p-type QW of $Cd_{0.976}Mn_{0.024}Te$ (sample No. 2). Dotted lines are guides for the eye; solid lines denote the assumed initial slope of E_- at 2.14 K. (b) Inverse magnetic susceptibility calculated from $d(E_\pm - E_+)/dH$ at $H \rightarrow 0$, as given by data in (a) for p-type QW of $Cd_{0.976}Mn_{0.024}Te$ (full circles). Note the presence of a ferromagnetic transition. Results for empty $Cd_{0.976}Mn_{0.024}Te$ QWs, where antiferromagnetic interactions dominate, are shown by empty symbols (after Ref. [13]).

to the results of Fig. 4b, and in agreement with the model, the same values of the Curie constant C are observed in the presence and the absence of the holes. At the same time, again in accord with the model, T_c does not virtually depend on p (as long as holes are delocalized) but increases with $\tilde{x}$. Having T_c we determine $\Theta = T_c - T_{AF}$, from which we obtain the enhancement factor A_F to be 2.5 ± 0.5, an expected value in view of the discussion presented in Section 2.

In addition to enhancing Θ, static disorder and carrier–carrier interactions impose a strong influence on the optical properties. In particular, they lead to a band-gap narrowing [43] that amounts to 29 meV for all our doped samples, as deduced from the energy difference between the free exciton reflectivity for the empty QW and the PL in the presence of the holes (7 meV at $H = 0$), and from an exciton binding energy of 22 meV, according to an estimate based on the $1s$–$2s$ excitonic energy separation [44] in the undoped sample No. 4. Moreover, these renormalization ef-

fects are expected to vary strongly with the degree of carrier polarization. Accordingly, while the evaluated value of the splitting of the heavy-hole subband corresponding to the full-hole polarization, $\Delta \approx 6$ meV for $p = 2 \times 10^{11}$ cm^{-2} (samples No. 1, 2, and 5), is not inconsistent with the experimental results of Figs. 3 and 4, the achievement of the saturation with the lowering of temperature proceeds slower than predicted by the model. A work is under way aimed at determining the effect of the renormalization on the dependence $\Delta(T, H)$ as well as at examining the domain structure in this novel ferromagnetic system.

5. Conclusions

The findings presented above demonstrate that indeed p-type doping constitutes *the* method for substantial enhancement of magnetic, and then magnetooptical effects in DMS, leading – in appropriately designed structures – to a ferromagnetic phase transformation. Additional enhancement of the tendency towards the ferromagnetic ordering is possible by engineering such a microscopic distribution of the magnetic ions, which would reduce T_{AF} and/or increase Θ. Since the interactions between localized spins mediated by the carriers are long range, a simple mean-field approach gives a correct quantitative description of magnetic properties, even in reduced dimensionality systems. At the same time, the data make possible to evaluate the strength of many body effects for the case of two-dimensional hole gas. Moreover, while our results demonstrate the possibility of changing the magnetic phase by light, other means – such as gates – are expected to provide also a high degree of control over magnetic properties in modulation-doped structures. This opens new perspectives for further studies of coupled carrier liquids and localized spins in novel geometries and material systems.

References

[1] For a review on DMS, see, e.g., T. Dietl, in: T.S. Moss (Ed.), Handbook on Semiconductors, vol. 3b, North-Holland, Amsterdam, 1994, p. 1251; J.K. Furdyna, J. Appl. Phys. 64 (1988) R29.

[2] See, P. Łazarczyk, T. Story, M. Arciszewska, R.R. Gałązka, J. Magn. Magn. Mater. 169 (1997) 151 and references therein.

[3] See, S. Koshihara, A. Oiwa, M. Hirasawa, S. Katsumoto, Y. Iye, C. Urano, H. Takagi, H. Munekata, Phys. Rev. Lett. 78 (1997) 4617; H. Ohno, A. Shen, F. Matsukura, A. Oiwa, A. Endo, S. Katsumoto, Y. Iye, Appl. Phys. Lett. 69 (1996) 363.

[4] H. Ohno, F. Matsukura, A. Shen, Y. Sugawara, A. Oiwa, A. Endo, S. Katsumoto, Y. Iye, in: M. Scheffler, R. Zimmermann (Eds.), Proc. 23rd Int. Conf. on the Physics of Semiconductors, Berlin, 1996, World Scientific, Singapore, 1996, p. 405; F. Matsukura, H. Ohno, A. Shen, Y. Sugawara, Phys. Rev. B 57 (1998) R2037.

[5] J. Blinowski, P. Kacman, J.A. Majewski, Phys. Rev. B 53 (1996) 9524.

[6] T. Wojtowicz, G. Karczewski, J. Kossut, Thin Solid Films 306 (1997) 271.

[7] N. Shibata, A. Ohki, A. Katsui, J. Crystal Growth 93 (1988) 703.

[8] T. Baron, S. Tatarenko, K. Saminadayar, N. Magnea, J. Fontenille, Appl. Phys. Lett. 65 (1994) 1284.

[9] F. Bassani, S. Tatarenko, K. Saminadayar, N. Magnea, R.T Cox, A. Tardot, C. Grattepain, J. Appl. Phys. 72 (1992) 2927; S. Scholl, H. Schäfer, A. Waag, D. Hommel, K. von Schierstedt, B. Kuhn-Heinrich, G. Landwehr, Appl. Phys. Lett. 62 (1993) 3010; I. Smorchkova, N. Samarth, Appl. Phys. Lett. 69 (1996) 1640.

[10] S. Tatarenko, T. Baron, A. Arnoult, J. Cibert, M. Grün, A. Haury, Y. Merle d'Aubigné, A. Wasiela, K. Saminadayar, J. Crystal Growth 175/176 (1997) 682; A. Arnoult, S. Tatarenko, D. Ferrand, J. Cibert, A. Haury, A. Wasiela, Y. Merle d'Aubigné, Proc. Int. Conf. II–VI Compounds, Grenoble 1997, J. Crystal Growth, in press.

[11] J. Jaroszyński, J. Wróbel, M. Sawicki, E. Kamińska, T. Skośkiewicz, G. Karczewski, T. Wojtowicz, A. Piotrowska, J. Kossut, T. Dietl, Phys. Rev. Lett. 75 (1995) 3170.

[12] T. Dietl, A. Haury, Y. Merle d'Aubigné, Phys. Rev. B 55 (1997) R3347.

[13] A. Haury, A. Wasiela, A. Arnoult, J. Cibert, S. Tatarenko, T. Dietl, Y. Merle d'Aubigné, Phys. Rev. Lett. 79 (1997) 511.

[14] J. Cibert, P. Kossacki, A. Haury, A. Wasiela, Y. Merle d'Aubigné, T. Dietl, A. Arnoult, S. Tatarenko, Proc. Int. Conf. II–VI Compounds, Grenoble 1997, J. Crystal Growth, in press.

[15] J.A. Gaj, W. Grieshaber, C. Bodin-Deshayes, J. Cibert, G. Feuillet, Y. Merle d'Aubigné, A. Wasiela, Phys. Rev. B 50 (1994) 5512.

[16] see, C. Kittel, Solid State Physics, vol. 22, F. Seitz, D. Turnbull, H. Ehrenreich (Eds.), Academic Press, New York, 1968, p. 1 and references therein.

[17] M.A. Ruderman, C. Kittel, Phys. Rev. 96 (1954) 99.

[18] Y. Yafet, Phys. Rev. B 36 (1987) 3948.

[19] B.L. Altshuler, A.G. Aronov, in: A.L. Efros, M. Pollak (Eds.), Electron–Electron Interaction in Disordered Systems, North-Holland, Amsterdam, 1985, p. 1; see also, H. Fukuyama, A.L. Efros, M. Pollak (Eds.), Electron–Electron Interaction in Disordered Systems, North-Holland, Amsterdam, 1985, p. 155; P.A. Lee, T.V. Ramakrishman, Rev. Mod. Phys. 57 (1985) 287.

[20] P.G. de Gennes, J. Phys. Radium 23 (1962) 630.

[21] A.Yu. Zyuzin, B.Z. Spivak, Pisma Zh. Eksp. Teor. Fiz. 43 (1986) 185 [JEPT Lett. 43 (1986) 234]; L.N. Bulaevskii, S.V. Panyukov, Pisma Zh. Eksp. Teor. Fiz. 43 (1986) 190 [JEPT Lett. 43 (1986) 240]; A. Jagannathan, E. Abrahams, M.J. Stephen, Phys. Rev. B 37 (1988) 436.

[22] B.L. Altshuler, A.G. Aronov, A. Yu. Zyuzin, Pisma Zh. Eksp. Teor. Fiz. 38 (1983) 128 [JETP Lett. 38 (1983) 153].

[23] see, T. Dietl, in Ref. [1], pp. 1280 and 1284.

[24] M. Sawicki, T. Dietl, J. Kossut, J. Igalson, T. Wojtowicz, W. Plesiewicz, Phys. Rev. Lett. 56 (1986) 508; see, also, Y. Shapira, N.F. Oliveira, Jr., P. Becla, T.Q. Vu, Phys. Rev. B 41 (1990) 5931; J. Jaroszyński, J. Wróbel, M. Sawicki, E. Kamińska, T. Skośkiewicz, G. Karczewski, T. Wojtowicz, A. Piotrowska, J. Kossut, T. Dietl, Phys. Rev. Lett. 75 (1995) 3170; I.P. Smorchkowa, N. Samarth, J.M. Kikkawa, D.D. Awschalom, Phys. Rev. Lett. 78 (1997) 3571.

[25] G. Brunthaler, T. Dietl, J. Jaroszyński, M. Sawicki, G. Stöger, A. Prinz, F. Schäffler, G. Bauer, Semiconductor Sci. Technol. 11 (1996) 1624.

[26] R. Egger, H. Schoeller, Phys. Rev. B 54 (1996) 16 337.

[27] T. Dietl, J. Spałek, Phys. Rev. Lett. 48 (1982) 355; Phys. Rev. B 28 (1983) 1548.

[28] See, e.g., P. Leroux-Hugon, in: P.R. Wallace, R. Harris, M.J. Zuckermann (Eds.), New Developments in Semiconductors, Noordhoff, Leyden, 1973, p. 63; M.A. Krivoglaz, Usp. Fiz. Nauk 111 (1973) 617 [Sov. Phys. Usp. 16 (1974)] 856; E.A. Pashitskii, S.M. Ryabchenko, Fiz. Tver. Tela, 21 (1979) 545 [Sov. Phys. Solid State 21 (1979) 322].

[29] M.E. Fisher, S.-k. Ma, B.G. Nickel, Phys. Rev. Lett. 29 (1972) 917.

[30] See, J.M. Yeomans, Statistical Mechanics of Phase Transitions, Oxford University Press, Oxford, 1993, p. 57.

[31] P. Łazarczyk, W. Dobrowolski, T. Story in: Extended Abstracts of Int. Conf. on Electron Localization and Quantum Transport in Solids, Jaszowiec, Poland, 1996, T. Dietl (Ed.), Institute of Physics, Warsaw, 1996, p. 129.

[32] S. von Molnar, T. Kasuya, Phys. Rev. Lett. 21 (1968) 1757; S. von Molnar, M.W. Shafer, J. Appl. Phys. 41 (1970) 1093.

[33] P. Wachter, CRC Crit. Rev. Solid State Sci. 3 (1972) 189.

[34] Le Si Dang, G. Neu, R. Romestain, Solid State Commun. 44 (1982) 1187.

[35] J.A. Gaj, J. Ginter, R.R. Gałązka, Phys. Stat. Sol. B 89 (1978) 655.

[36] Glód, T. Dietl, M. Sawicki, I. Miotkowski, Physica B 194–196 (1994) 995.

[37] M.A. Novak, O.G. Symko, D.J. Zheng, S. Oseroff, J. Appl. Phys. 57 (1985) 3418.

[38] G. Fishman, Phys. Rev. B 52 (1995) 11 132.

[39] P. Peyla, A. Wasiela, Y. Merle d'Aubigné, D.E. Ashenford, B. Lunn, Phys. Rev. B 47 (1993) 3783; B. Kuhn-Heinrich, W. Ossau, E. Bangert, A. Waag, G. Lanwehr, Solid State Commun. 91 (1994) 413.

[40] C. Benoit à la Guillaume, Phys. Stat. Sol. B 175 (1993) 369.

[41] T. Wojtowicz, S. Koleśnik, I. Miotkowski, J.K. Furdyna, Phys. Rev. Lett. 70 (1993) 2317.

[42] A.J. Shields, J.L. Osborne, M.Y. Simmons, D.A. Ritchie, M. Pepper, Semiconductor Sci. Technol. 11 (1996) 890 and references therein.

[43] S. Schmitt-Rink, D.S. Chemla, D.A.B. Miller, Adv. Phys. 38 (1989) 89 and references therein.

[44] H. Mathieu, P. Lefebvre, P. Christol, Phys. Rev. B 46 (1992) 4092.

List of Contributors

Abrahams, E.   69
Abstreiter, G.   103
Akis, R.A.   137
Aleiner, I.L.   58
Alferov, Zh.I.   129
Altshuler, B.L.   58
Arnoult, A.   169
Ashoori, R.C.   15
Austing, D.G.   112

Badrieh, F.   137
Baldwin, K.W.   38
Barnes, C.H.W.   52
Bauer, G.   79
Bichler, M.   103
Bimberg, D.   129
Bird, J.P.   137
Boxleitner, W.   152
Brunthaler, G.   79
Bunin, G.   47

Chichester, R.J.   8
Cibert, J.   169

de-Picciotto, R.   47
Dietl, T.   169
Dunkleberger, L.N.   8

Ferry, D.K.   137
Fischer, F.   158
Fulton, T.A.   8

Gershenson, M.E.   58
Gornik, E.   152
Grundmann, M.   129

Haury, A.   169
Heiblum, M.   47
Heinrichsdorff, F.   129
Hess, H.F.   8
Hess, K.   1

Hill, N.P.R.   52
Holmberg, N.   137
Honda, T.   112

Janssen, J.W.   112
Jones, G.A.C.   52

Kardynal, B.   52
Kizilyalli, I.C.   1
Klitzing, K.v.   23
Kop'ev, P.S.   129
Kossacki, P.   169
Kouwenhoven, L.P.   112

Landwehr, G.   158
Ledentsov, N.N.   129
Linfield, E.H.   52
Lott, J.A.   129
Lugauer, H.-J.   158
Lyding, J.   1

Mahalu, D.   47
Maximov, M.V.   129
Merle d'Aubigné, Y.   169
Muraki, K.   112

Nicholls, J.T.   52
Nötzel, R.   92

Oosterkamp, T.H.   112
Oswald, J.   30

Pan, W.   38
Pepper, M.   52
Pfeiffer, L.N.   8, 15, 38
Pivin Jr., D.P.   137
Ploog, K.H.   92
Prinz, A.   79
Pudalov, V.M.   79

Rauch, C.   152
Register, L.F.   1
Reznikov, M.   47
Ritchie, D.A.   52

Schedelbeck, G.   103
Schüll, K.   158
Schüller, C.   121
Stormer, H.L.   38
Strasser, G.   152

Tarucha, S.   112
Tatarenko, S.   169
Tersoff, J.   89
Tokura, Y.   112
Tsui, D.C.   38
Turner, N.   52
Tuttle, B.   1

Umansky, V.   47
Unterrainer, K.   152
Ustinov, V.M.   129

Vasileska, D.   137

Waag, A.   158
Wasiela, A.   169
Wegscheider, W.   103
Wei, Y.Y.   23
Weis, J.   23
West, K.W.   8, 15, 38
Willett, R.L.   8
Wixforth, A.   145

Yeh, A.S.   38
Yoo, M.J.   8

Zhitenev, N.B.   15

PHYSICA E

INSTRUCTIONS TO AUTHORS

Submission of papers

Manuscripts (one orginal + two copies) should be sent to one of the following regular editors:

Original material: Submission of a manuscript implies that it is not being simultaneously considered for publication elsewhere and that the authors have obtained the necessary authority for publication.

Types of contributions

Original research papers and letters to the editors are welcomed. Both should contain an Abstract (of up to 200 words) and a Conclusions section, which, particularly in the case of theoretical papers, translates the results into terms readily accessible to most readers.

Manuscript preparation

All manuscripts should be written in proper English. The paper copies of the text should be prepared with double line spacing and wide margins, on numbered sheets.
Structure: Please adhere to the following order of presentation: Article title, Author(s), Affiliation(s), Abstract, PACS codes and keywords, Main text, Acknowledgements, Appendices, References, Figure captions, Tables.
Corresponding author: The name, complete postal address, telephone and fax numbers and the E-mail address of the corresponding author should be given on the first page of the manuscript.
PACS codes/keywords: Please supply one or more relevant PACS-1995 classification codes and 3-4 keywords of your own choice for indexing purposes.
References: References to other work should be consecutively numbered in the text using square brackets and listed by number in the Reference list.

Illustrations

Illustrations should also be submitted in triplicate: one master set and two sets of copies. The *line drawings* in the master set should be original laser printer or plotter output or drawn black india ink, with careful lettering, large enough (3–5 mm) to remain legible after reduction for printing. The *photographs* should be originals, with somewhat more contrast than is required in the printed version. They should be unmounted unless part of a composite figure. Any scale markers should be inserted on the photograph itself, not drawn below it.
Colour plates: Figures may be published in colour, if this is judged essential by the editor. The Publisher and the author will each bear part of the extra costs involved. Further information is available from the publisher.

After acceptance

Notification: You will be notified by the editor of the journal of the acceptance of your article.
Copyright transfer: In the course of the production process you will be asked to transfer the copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.

Electronic manuscripts

The Publisher welcomes the receipt of an electronic version of your accepted manuscript. If there is not already a copy of this (on diskette) with the journal editor at the time the manuscript is being refereed, you are invited to send a file with the text of the accepted manuscript directly to the Publisher on diskette (allowed formats 3.5″ or 5.25″ MS-DOS, or 3.5″ Macintosh) to the address given below. Please note that no deviations from the version accepted by the editor of the journal are permissible without the prior and explicit approval of the editor. Such changes should be clearly indicated on an accompanying printout of the file.

Author benefits

No page charges: Publishing in Physica E is free of charge.
Free offprints: The corresponding author will receive 25 offprints free of charge. An offprint order form will be supplied by the publisher for ordering any additional paid offprints.
Discount: Contributors to Elsevier Science journals are entitled to a 30% discount on all Elsevier Science books.

Further information (after acceptance):