TRANSPORT PROPERTIES OF MODULATION-DOPED Si/SiGe QUANTUM WELL STRUCTURES

CHRISTIAN E. BECKER

A thesis submitted to University of London for the degree of Doctor of Philosophy

October 2002
Abstract

Magnetotransport measurements have been performed at low temperatures and high magnetic fields on a range of high-quality $n$- and $p$-type modulation doped Si/SiGe heterostructures grown by gas source molecular beam epitaxy in order to study the transport properties of the respective two-dimensional electron and hole gas systems. The analysis of the Shubnikov-de Haas effect provided vital information that allowed a structural and electrical characterization of the investigated samples. The observed anomalies in the low-temperature magnetotransport phenomena could be explained by giving a full description of the quantum mechanical states of the system. A self-consistent Poisson-Schrödinger solver has been used to model the band structure of the samples and to engineer the wavefunction. As a result, novel $p$-type structures were suggested in which the wavefunction is pulled away from the inferior upper interface by employing a step-graded Ge profile.

In agreement with theory, the electron effective mass is unaffected by variations in strain or magnetic field, whereas the hole effective mass displays a dependence on both the strain and the magnetic field due to nonparabolicity effects in the valence band. The high-mobility $n$-type samples are dominated by long-range scattering from remote ionized impurities. However, the contribution of short-range scattering from the interface and especially from threading dislocations in the virtual substrate is still believed to be significant, and a substantial increase in scattering from threading dislocations could be seen in the low-mobility $n$-type structures. In the $p$-type samples, the mobility is predominantly limited by interface charge scattering with a growing contribution from interface roughness scattering as the Ge content and thus the sheet density increases.

Representing the dominant scattering mechanism at room temperature, phonon processes have also been studied at low temperatures, whereby the energy relaxation of hot electrons and hot holes occurs by acoustic phonon interactions via unscreened deformation-potential coupling. Values for the deformation potential have been extracted from the temperature dependence of the mobility in the equipartition regime and from the energy-loss rate in the Bloch-Grüneisen regime. In both types of structures, the deformation potential has been found to decrease with increasing strain.
Acknowledgements

First and foremost, I want to express my gratitude towards my supervisor Dr. Jeff Harris, whose guidance and assistance made this work possible. Providing crucial ideas and valuable advice in lengthy discussions, his vast experience and his detailed knowledge of semiconductor physics were pivotal in completing the experiments and provided an almost unlimited source of information I could draw on for the analysis of the results. I am particularly grateful for his patience and understanding during the entirety of my research.

Next, I want to thank Dr. Juan Fernández, who was in charge of growing most of the high-quality $n$- and $p$-type samples, and I would also like to give due credit to Dr. Nick Woods for the growth of the more recent structures as well as for helping me to process a large number of samples. Furthermore, I want to show my appreciation to Prof. J. C. Portal and Dr. Duncan Maude at the High Magnetic Field Laboratory in Grenoble, who were responsible for this successful European collaboration that produced most of the experimental data of my research. In addition, I would like to thank Dr. Kristel Fobelets at Imperial College for giving me access to their clean room and for letting me use the cryomagnetic system in their laboratory. Credits also go to Kevin Lee, who was in charge of processing and bonding of the latest samples.

The efforts of Dr. Till Sander to introduce me to the intricacies of low-temperature experiments are gratefully acknowledged, as are the endless but helpful discussions during the long hours spent in the laboratory. I would further like to extend my appreciation to all of my fellow Ph.D. students in the semiconductor transport group, Dr. Dong-Hoon Shin, Jay Koo, Dr. Kwon Lee and Costas Mavroidis, for stimulating discussions but also for much needed distractions from the exhausting exercise of getting a grip on the problems of quantum mechanics. Additional credits go to all the members of the Diamond group.

Most of all, I want to express my gratitude towards my family for their constant encouragement and their unlimited moral as well as financial support during a rather extended period of study. Finally, I want to thank Bambi for her tolerance and the love and inspiration I received.
Contents

1 INTRODUCTION 1

2 THE Si/SiGe HETEROSYSTEM 7
   2.1 Physical Properties ................................................................. 8
       2.1.1 Material Growth .............................................................. 11
       2.1.2 Band Structure ............................................................... 14
       2.1.3 Strain Effects ................................................................. 21
       2.1.4 Band Alignment .............................................................. 26
   2.2 Physics of Two-Dimensional Systems ....................................... 28
       2.2.1 Modulation Doping and Quantum Wells ............................. 28
       2.2.2 Density of States ............................................................ 34
   2.3 Device Applications ................................................................. 36
       2.3.1 Modulation-Doped Field Effect Transistors ....................... 37
       2.3.2 Heterojunction Bipolar Transistor ................................... 46
   2.4 Growth Techniques ................................................................. 48
       2.4.1 Molecular Beam Epitaxy .................................................. 48
       2.4.2 Chemical Vapour Deposition .......................................... 49
       2.4.3 Structural Characterization .............................................. 50
       2.4.4 Compatibility to Standard VLSI Technology ...................... 51

3 MAGNETOTRANSFER IN 2D SYSTEMS 52
   3.1 Semi-Classical Approach ......................................................... 53
       3.1.1 Hall Effect ................................................................. 55
       3.1.2 Van der Pauw Method .................................................... 59
   3.2 Quantum Mechanical Approach .............................................. 60
       3.2.1 Landau Levels .............................................................. 61
       3.2.2 Shubnikov-de Haas Effect .............................................. 67
       3.2.3 Integer Quantum Hall Effect .......................................... 76
   3.3 Parallel Conduction ............................................................... 78
## Contents

8 **Effective Mass**

8.1 Temperature Dependence of SdH Oscillations ........................................ 164

8.2 Digital Filtering ........................................................................................... 168

8.2.1 Digital Filtering in Frequency Domain ........................................ 170

8.2.2 Digital Filtering in Time Domain .................................................. 171

8.3 Electron Effective Mass .............................................................................. 176

8.4 Hole Effective Mass .................................................................................... 182

8.5 Summary ...................................................................................................... 195

9 **Electronic Transport Properties**

9.1 Fundamental Field ...................................................................................... 197

9.2 Electron Mobility and Sheet Density ........................................................ 199

9.3 Hole Mobility and Sheet Density .............................................................. 209

9.4 Summary ...................................................................................................... 217

10 **Transport and Quantum Relaxation Time**

10.1 Dominant Scattering in 2DEG Samples ................................................. 220

10.2 Dominant Scattering in 2DHG Samples ................................................. 226

10.3 Summary ...................................................................................................... 230

11 **Deformation Potential**

11.1 Equipartition Regime .................................................................................. 234

11.1.1 Reciprocal Electron Mobility ......................................................... 236

11.1.2 Reciprocal Hole Mobility ............................................................... 239

11.2 Bloch-Grüneisen Regime ............................................................................ 242

11.2.1 Energy-Loss Rate of Hot Electrons ............................................... 247

11.2.2 Energy-Loss Rate of Hot Holes ..................................................... 254

11.3 Summary ...................................................................................................... 259

12 **Conclusions**

VI
1 INTRODUCTION

Commanding a share of over 97% in the worldwide market for consumer electronics, silicon is by far the most important and most widely used semiconductor material due to the combination of an easily available semiconductor and its excellent natural oxide which serves as an insulator and also as a protecting passivation layer. Despite the comparatively low electron and hole mobilities in Si, the dominant position of silicon devices in the field of microelectronics can be explained by the unique physical and chemical properties of the Si/SiO₂ system which allows for a monolithic integration of metal-oxide-semiconductor field-effect transistors (MOSFETs) in large quantities. These properties are exploited in the very large-scale integration (VLSI) technology, and in combination with a low density of charged states in the oxide and at the interface to Si, they are the key to the fabrication of several million transistors with identical electrical behaviour as an integrated circuit (IC) on a single chip. Although its unique range of features makes the Si/SiO₂ material system widely applicable for the design of both analogue and digital circuits, it is ideally suited for digital applications with a very high level of complexity. Amongst the most important benefits of the Si CMOS (complementary MOS) technology are low power dissipation, high integration levels, good noise immunity and reliability. However, since the p-channel devices are inferior to their n-channel counterparts in terms of current drive capability and speed performance, they have to be designed about 2 - 3 times larger in order to compensate for these deficiencies, which affects the level of integration and device speed adversely. Despite these drawbacks, the progress in Si CMOS technology promises digital and analogue devices that operate in the GHz regime, ultimately replacing the bipolar circuits currently used in high speed, high gain and low noise applications.

Driven by the demand for ever faster and smaller devices (measured in terms of the gate delay and the gate length, respectively) standard Si technology is about to reach its limits in the near future (around the year 2010) when the size of the device structures will enter the region in which quantum effects have to be taken into account. An important part of this endeavour to increase the speed performance of semiconductor devices is the fast growing market segment representing the area of high-frequency communication which comprises, for example, mobile and cordless phones in the 1 - 3 GHz range, wireless LANs (local area networks) between 2.4 and 5.8 GHz, satellite communication in the 10 - 14 GHz range and wideband communication via optical fibres transmitting at up to 40 Gbit/s which includes the use of optoelectronic devices such as semiconductor lasers, modulators, optical detectors and sensors. Since the required device features appear to be beyond the scope covered by the electronic and optoelectronic properties of Si, these high-frequency analogue and optical applications have previously been exclusively the domain of III-V heterostructures such as the lattice-matched ternary compound semiconductor GaAs/Al₃Ga₉As. The key to the system's superior high-frequency behaviour is the very high electron mobility that can be achieved due to its almost perfect lattice
1 Introduction

match which facilitates the growth of pseudomorphic epitaxial layers with high-quality interfaces, whereas the direct bandgap in III-V compound materials is essential to their advanced optoelectronic functionality. However, despite extensive general research in this field and substantial efforts in the design of FET structures within the last three decades, the fabrication of logic devices based on III-V materials is still prevented by the complete lack of a stable native oxide or other insulator with the quality and versatility required for VLSI technology. Maybe the most important restriction inhibiting the wider use of III-V heterostructures is the very high cost basis compared to standard Si CMOS technology for which the cost was over 200 times cheaper than for GaAs per square mm in 1995.

An elegant solution to overcome the problem of choosing between a good speed performance and a cost-effective technology is the application of the Si$_{1-x}$Ge$_x$ heterosystem which provides the benefits of band structure engineering without sacrificing the mandatory compatibility with standard Si technologies and their capabilities for low-cost mass production. As a result, increasing effort has been put into the research and development of modulation-doped Si/SiGe heterostructures within the last couple of decades to facilitate the fabrication of Si-based heterodevices with superior performance at reasonable costs. The main characteristic of a modulation-doped heterostructure with respect to its use in a lateral transport device, such as an FET, is the formation of a potential well in which electrons or holes can be confined. If the dimension of the confining potential well is of the order of the de Broglie wavelength, the carrier motion in the direction perpendicular to the heterointerface will be quantized while the movement of the carriers along the interface maintains the character of a free electron or hole. Modulation doping of the heterostructure also decreases the scattering rate inside the resulting two-dimensional (2D) electron or hole gas significantly because the mobile carriers are spatially separated from their parent impurities. Hence, the carrier mobility in modulation-doped heterostructures can be increased drastically over that of bulk material or inversion layers.

Initially, the growth of Si$_{1-x}$Ge$_x$ heterostructures was believed to be virtually impossible because of two main reasons. First and foremost, the rather large lattice mismatch of 4.2% between Si and Ge did not allow commensurate layer growth with the available growth techniques at the time, and secondly, the bandgap difference between Si and Ge was considered to be too small for the heterointerface to be able to form a confinement potential for electrons or holes. However, the introduction of a new growth technique, called molecular beam epitaxy (MBE), facilitated the growth of epitaxial SiGe thin films on Si substrates and was instrumental in growing the first Si/SiGe superlattice in 1975 [1]. Yet, progress in material growth and basic understanding of the band alignment still lagged behind similar investigations carried out in GaAs/AlGaAs heterosystems. The quality of the strained material remained very low for years to come, and the measurements of the electronic and optical properties of these heterostructures were dominated by material defects. Advances in the MBE growth technique eventually lead to the successful growth of high-quality pseudomorphic SiGe thin films on Si substrates in 1984 [2, 3]. It soon became evident that the strain has a strong effect on the bandgap and thus determines the band edge line-
up at the heterointerface, allowing electron confinement to form a two-dimensional electron gas (2DEG) in the case of tensile strain [4, 5] and hole confinement to form a two-dimensional hole gas (2DHG) in the case of compressive strain [6]. Strain also lifts degeneracies at the energy band edges, resulting in lighter effective masses in the 2D plane for both electrons and holes. These features were subsequently exploited for the realization of several demonstrative devices including an n-type MODFET (modulation-doped FET) [7], a p-type MODFET [8] and a heterojunction bipolar transistor (HBT) [9]. Continued progress in existing growth techniques and hence in the quality of strained epitaxial layers have lead to the introduction of the first commercial ICs containing SiGe HBTs in 1998.

Improvements in the performance of conventional Si-based electronic devices and the design of novel device concepts are essentially based on band structure engineering and superior materials quality available in the emerging Si/SiGe heterostructure technology. However, the control over certain technological requirements is still deficient in some aspects, which helps to explain the strong demand for further materials based research and development in this area. In particular, the status of the modulation-doped p-type devices is still subject to concern when compared to the performance of current n-type structures. For the ultimate aim of producing symmetric n- and p-channel characteristics, mobility improvements in the p-type structure have to be accomplished, which will yield very significant performance enhancements in the CMOS technology such as better linearity, higher current drive, better noise performance and reductions in the supply voltage. Concerning the tailoring of the band structure, the fundamental properties which influence SiGe device operation are the semiconductor bandgap, the effective mass of the carriers, the scattering processes and the electron-hole recombination processes (significant for HBTs). By varying the Ge composition $x$ in the Si$_{1-x}$Ge$_x$ alloy, it is possible to exploit the various effects of the resulting strain on certain electronic properties of the system. As the most important component, strain changes the energy bandgap (significant for optoelectronic devices), it changes the energy band alignment (required for electron/hole confinement), it reduces the effective mass of the carriers (enhancement of mobility), it modifies the coupling between various energy bands (affects effective mass and scattering) and it changes the crystal symmetry, offering the possibility of enhancing the recombination of electrons and holes (beneficial for HBTs). Other examples of structural parameters that can be varied to tailor the band structure are the width of the spacer layer, the doping concentration and the doping profile to modify the sheet density in the 2D carrier gas and to reduce parallel conduction in the supply layer.

Within the framework of this thesis, the results and subsequent analysis of the low-temperature magnetotransport experiments performed on various Si/Si$_{1-x}$Ge$_x$ 2DEG and 2DHG heterostructures grown by gas source MBE (GSMBE) are presented and possible interpretations are discussed. The emphasis is put on the growth of high-quality heterointerfaces in order to achieve superior transport properties at low temperatures and at room temperature. A good control over the growth parameter and the individual layer thickness in conjunction with the ultra-high vacuum ambient of the MBE guarantees the production of very high quality heterointerfaces.
The crucial advantage of gas source over solid source MBE is its high growth selectivity and the reduced incorporation of impurities from the source. However, being a relatively new growth technique and hence not as perfected as other well-established growth techniques, it still requires extensive feedbacks from structural as well as from electronic characterizations to establish the quality of the individual layers and to verify that the grown structures comply with the initial specifications. Furthermore, it is important to test the stability of the strain and thus the structural integrity over an extended time period (and also after thermal treatment) to ascertain the suitability of GSMBE as a reliable growth technique for Si-based devices in standard Si technologies capable of reproducing morphologically stable heterostructures of very high quality.

Representing a methodical introduction into the physics of the Si/SiGe heterosystem and the practical application of this technology, chapter 2 covers the physical and structural aspects that are relevant for the understanding and the implementation of high-mobility carrier gases in this material system. Accounting for the basic physical properties, the pseudomorphic layer growth of the lattice-mismatched materials as well as the influence of the resulting strain on the energy bandgap and on the alignment of the energy bands are explained first. With regard to the realization of a two-dimensional carrier gas, the formation of a confinement potential by employing the modulation doping technique and by designing a quantum well is described subsequently. To elucidate the various practical uses of the Si/SiGe heterosystem, a concise review of the most important device applications, i.e. MODFET and HBT, and their respective performance figures is also presented. Finally, a brief overview of the two principal growth techniques, their respective position in the commercial mass production of Si-based devices and various methods for the structural characterization of the grown samples is presented.

A comprehensive quantum mechanical description of the electronic states at low temperatures and high magnetic fields is given in chapter 3 to help explain the Shubnikov-de Haas (SdH) and the integer quantum Hall effect (IQHE) in the resulting magnetically quantized system. As the bulk of the magnetotransport experiments are performed in the quantum regime, the physical background provided in this chapter provides the theoretical basis for all analytical methods employed in the analysis and interpretation of the experimental data. In addition, the potentially detrimental influence of parasitic parallel conduction channels on the magnetotransport measurements is considered at low temperatures as well as at room temperature.

A systematic examination of the various scattering processes in a semiconductor is presented in chapter 4. Conventionally, all scattering processes can be divided into time-dependent scattering due to lattice vibrations and time-independent scattering by impurities. Due to changed structural and physical properties, the scattering processes may behave differently in the bulk material of a semiconductor and the quantum well of the corresponding two-dimensional structure, and the scattering mechanisms in both cases are therefore discussed separately. In order to give a general idea of the dominant scattering processes that limit the electron and hole
mobility in $n$- and $p$-type structures, respectively, the strength and influence of each scattering mechanism is investigated individually for both structure types.

In chapter 5, the programme for the self-consistent Poisson-Schrödinger solver is described in sufficient detail to validate its application in the tailoring of the band structure according to a set of predetermined specifications. These band structure calculations represent a vital stage in the feedback process between the growth of the samples and the experimental analysis of the grown structures. Since the emphasis lies on an enhancement of the $p$-type devices, band structure, all calculations are performed exclusively on the valence band of 2DHG structures and are concentrated on improving the low-temperature hole mobility, i.e. reducing the hole effective mass and short-range scattering from the upper interface.

A detailed description of the various cryomagnetic systems used to perform the low-temperature magnetotransport measurements is given in chapter 6. Apart from their working principle and mode of operation, the corresponding electrical set-up and the associated operational limits are also specified for each magnetic system. Since a substantial part of the practical work included the growth and the subsequent photolithographical processing of numerous samples, a brief overview of the complete sample preparation is provided at the end of the chapter.

In chapter 7, a selection of the experimentally acquired Shubnikov-de Hass oscillations and the quantum Hall effect is presented for the investigated $n$- and $p$-type samples. For a better understanding of both effects and their significance in the two-dimensional Si/SiGe heterosystem, a quantum mechanical description of the low-temperature magnetotransport phenomena and potentially correlated anomalies is given for the results of each sample.

From the temperature dependence of the Shubnikov-de Haas oscillations, the electron and hole effective masses are calculated in chapter 8 for the present $n$- and $p$-type samples, respectively. Due to a severe nonparabolicity and a strong coupling of the hole subbands, the determination of the hole effective mass is of higher importance than that of the electron effective mass which remains constant irrespective of the structural properties of the sample. Although the use of a digital filter to remove the low-frequency background of the SdH data is not restricted to this analytical method, a common mathematical description of such a filter and related problems are also included as the digital filter is used in this chapter for the first time.

Being the most important transport properties, the mobility of the carriers and the corresponding sheet density are subsequently calculated in chapter 9 from the SdH oscillations and the quantum Hall effect. The temperature dependence of both properties is discussed in terms of the dominant scattering mechanism with a strong emphasis on acoustic phonon scattering. To demonstrate the influence of electron heating on the transport behaviour of the 2D system at low temperatures, the electric field or electron temperature dependence of the mobility is examined, as well.

However, a proper investigation of the dominant scattering process in each sample is conducted in chapter 10 by evaluating the ratio of transport to quantum relaxation
time. Whereas the transport relaxation time is calculated from the mobility and the effective mass, the quantum relaxation time is extracted from the corresponding Dingle plot which exploits the exponential dampening of the SdH amplitudes with decreasing magnetic field. In addition, the temperature dependence of both relaxation times is investigated at low temperatures to establish whether such a dependence is caused by a change in the dominant scattering mechanism with temperature or by an intrinsic temperature dependence of the dominant scattering process. Alternatively, it is also conceivable that the correlation between the individual scattering processes is altered with varying temperature.

A comprehensive investigation of the acoustic phonon scattering process is conducted in chapter 11 by calculating the deformation potential in two different temperature regimes. In the equipartition regime, the acoustic phonon population increases linearly with temperature, and hence, the deformation potential is derived from the temperature dependence of the reciprocal mobilities. At lower temperatures in the Bloch-Grüneisen regime, on the other hand, acoustic phonon scattering is restricted by the constraints of degeneracy and energy conservation, and consequently, the energy-loss rate is used to determine the deformation potential. The temperature dependence of the momentum and energy relaxation times is also discussed in this temperature regime.

Providing a brief summary of the most important results of this work and a short discussion of their significance, the main achievements and some concluding remarks are presented in chapter 12. In addition, a few suggestions for future work are included at the end.
2 THE Si/SiGe HETEROSYSTEM

Silicon-based heterostructures are ideally suited to introduce the technique of band structure engineering into the VLSI technology of standard IC manufacturing as a powerful tool to enhance the performance level of contemporary Si devices significantly. Until recently, band structure engineering has been solely the domain of III-V heterosystems, but the advent of improved growth techniques enabled the fabrication of Si/SiGe heterostructures with electrical properties superior to those of the traditional Si/SiO₂ material combination by virtue of band structure engineering and higher material quality, thus filling the performance gap to the III-V heterostructures without sacrificing the essential compatibility with standard Si technologies.

The first section of this chapter is intended to serve as a general theoretical basis, explaining the most important physical properties of the Si/SiGe heterosystem that are vital to fully understand the concept of heterosystems. Pseudomorphic growth of silicon-based heterostructures and the growth related restrictions are described first since it illustrates quite clearly the problems of realizing such structures and immediately demonstrates the significance of having good control over epitaxial layer growth. Exact knowledge of the band structure is of utmost importance before the concept of heterosystems can be understood and the tools of band tailoring can be applied effectively. Hence, a detailed description of the energy bands in the conduction and valence band are given for Si, Ge and also for GaAs as the most prominent representative of the III-V semiconductors. Due to its complexity, the effect of strain on the band structure and the means to control the strain in a stack of heteroepitaxial layers, which is the principal tool of band structure tailoring, is explored in detail. Subsequently, the dependency of the band alignment and hence the band offset at the interface of a Si/SiGe heterostructure on the strain is investigated for all feasible variations of the Ge content in the compound to establish under which conditions electron or hole confinement is possible.

Having looked at the fundamental physical background of heterostructures, it is then possible to consider the basic principle of two-dimensional heterosystems which is covered in the following section. The realization of such structures is investigated by looking at the formation of quantum wells and the technique of modulation doping, both capable of producing two-dimensional electron and hole gas structures. As the density of states of an electronic system plays an important role in the profoundly changing behaviour of the electronic states, the different characteristic energy dependencies of the density of states are shown for a three-, two-, one- and zero-dimensional system and the corresponding features are explained.

Regarding the application-driven research into Si/SiGe heterostructures, it is the ultimate aim to develop new device concepts based on the exploitation of the band offset at the heterointerface with superior operating characteristics compared to their standard Si counterparts. Hence, the next section concentrates on the device aspects
of silicon-based heterostructures with the focus being placed on the realization of the most significant device applications which, in particular, are the $n$-type and $p$-type modulation-doped field effect transistors, the combination of the two complementary types and the $npn$ heterojunction bipolar transistor.

The last section is concerned with the various growth techniques capable of pseudomorphic epitaxial layer growth which are therefore used to produce high-quality Si/SiGe heterostructures. Concentrating on the two predominant growth techniques, their most important variations are presented while emphasizing their respective advantages and disadvantages. Physical characterization of the heterostructure during and after growth is vital to determine the quality of the device and confirm the growth parameters like doping concentration and channel thickness. The common tools for such a characterization are described in short. Finally, an assessment of the compatibility of Si/SiGe heterostructures and their growth techniques with standard Si VLSI technology is given.

2.1 Physical Properties

In order to understand the physical properties of a semiconductor material, it is important to comprehend the crystal structure of bulk semiconductors and to gain some insight into the resulting band structure. A fundamental concept in the description of any crystalline solid is that of the Bravais lattice which specifies the periodic array in which the repeated atoms of the crystal are arranged [10-12]. It is important that not only the arrangement but also the orientation must appear the same from every point in the Bravais lattice. Three distinct Bravais lattice types have to be considered for the lattice structure of most crystalline solids: the simple cubic (sc) lattice has an atom located at each corner, the body-centred cubic (bcc) lattice has an additional atom at the centre of the cube, and the face-centred cubic (fcc) lattice has additional atoms on each face plane. A lattice can be described uniquely by defining a unit cell as a basis and a set of lattice translation vectors that produce the entire lattice. The size of a unit cell is specified by the lattice constant $a$. A primitive cell is the smallest unit cell that can be repeated to form the lattice without either overlapping itself or leaving voids, but in many cases it is convenient to use a unit cell that is not a primitive cell, e.g. when the unit cells have orthogonal sides and the required symmetry whereas the corresponding primitive cells have nonorthogonal sides. One special way of choosing a primitive cell is by defining the Wigner-Seitz cell about a lattice point which is the region of space that is closer to that point than to any other lattice point. The significance of such a primitive cell will be explained later.

Many important semiconductors crystallize in the diamond or zinc-blende lattice whose unit cell is shown in Figure 2.1. Both lattice types belong to the tetrahedral phase where each atom is surrounded by four equidistant nearest neighbours which lie at the corner of a tetrahedron. The bond between two nearest neighbours is
formed by two electrons with different spin. The diamond and zinc-blende lattices can be considered as two interpenetrating face-centred cubic Bravais lattices, displaced along the body diagonal of the cubic cell by one quarter of the length of the diagonal. Neither the diamond nor the zinc-blende lattice is a Bravais lattice because the environment of any point differs in orientation from the environments of its nearest neighbours. In a diamond lattice, all atoms are of the same element whereas in a zinc-blende lattice, one sublattice consist of atoms from one element and the other sublattice consists of atoms from another element. In the case of GaAs, for example, both Ga and As sit on their own separate fcc lattice. Having established the difference between the two lattice structures, it is evident that only binary compound materials can have the zinc-blende structure where each atom has four atoms of the other element as nearest neighbours. With the exception of GaN, which has a wurtzite structure, all III-V compound semiconductors crystallize in the zinc-blende lattice. Except for diamond itself, Si, Ge and the random $\text{Si}_{1-x}\text{Ge}_x$ alloy are the only semiconductors to crystallize in the diamond lattice.

For most analytical studies of periodic structures, a reciprocal lattice is introduced which is the set of all wavevectors $\mathbf{K}$ that yield plane waves $e^{i\mathbf{K}\cdot\mathbf{r}}$ with the periodicity of a given Bravais lattice. The reciprocal lattice itself is a Bravais lattice. The Wigner-Seitz cell of the reciprocal lattice is known as the first Brillouin zone. Consequently, a periodic energy potential can be defined in $k$-space to describe the energy-momentum relationship of electrons and holes in the crystal. Assuming these requirements, Bloch's theorem states that the wavefunction only needs to be solved within the first Brillouin zone. All other positions in $k$-space can then be described by
imposing a modulation function with the periodicity of the Bravais lattice on the Brillouin zone solution. Figure 2.2 shows the first Brillouin zone of a diamond or zinc-blende lattice and indicates the most important symmetry points and symmetry lines for Si and Ge where the $\Gamma$-point marks the centre of the zone and the six $X$- and eight $L$-points are located at the zone boundaries.

Knowledge of the energy bands, which are made up of allowed energy states at which electrons and holes can exist, is vital to explain electrical and optical properties of crystals. In all semiconductors, the allowed energy states for electrons and holes form a set of conduction and valence bands, respectively. The energy of the lowest conduction band and that of the highest valence band is separated by the bandgap, a forbidden energy region in which allowed energy states cannot exist.

![Figure 2.2 First Brillouin zone of a face-centred cubic lattice with points and lines of high symmetry.](Image)

Position: $\Gamma$-point: $2\pi/a (0,0,0)$, zone centre, $X$-point: $2\pi/a (0,0,1)$, zone edge along the $\{100\}$ axes ($\Delta$), $L$-point: $2\pi/a (1/2,1/2,1/2)$, zone edge along the $\{111\}$ axes ($\Lambda$).

Depending on the location of the conduction band minimum and the valence band maximum in the Brillouin zone, two bandgap types can be distinguished. For direct bandgap semiconductors, the position of the conduction band minimum and that of the valence band maximum coincide (normally at the $\Gamma$-point) whereas in indirect bandgap materials, the energy band extrema are situated at different $k$-points.

In a first approximation, the conduction and valence bands in a bulk material can be described with a parabolic function in which the curvature is determined by the electron and hole effective mass, respectively. The valence band consists of three separate bands, the heavy hole band (the wider band with smaller curvature), the light hole band (the narrower band with larger curvature) and the split-off hole band. Due to the close proximity of the three subbands, the interaction effects between the corresponding states cause the valence bands to be nonparabolic and often anisotropic.
2.1.1 MATERIAL GROWTH

Despite a lattice mismatch of 4.2% between silicon and germanium, these materials are the only group-IV elements that are completely miscible. They form a series of solid solutions with gradually varying properties over the entire composition range. The lattice constants of bulk Si$_{1-x}$Ge$_x$ alloys and their dependence on the composition $x$ have been established by various experiments, and the results obey Vegard’s Law to a very good approximation [13-15]. According to Vegard’s Law, the values for the lattice constant have to lie on a linear fit when plotted against the alloy concentration, and the lattice constant of the bulk Si$_{1-x}$Ge$_x$ alloy can therefore be represented as

$$a_{\text{Si}_{1-x}\text{Ge}_x} = a_{\text{Si}} + (a_{\text{Si}} - a_{\text{Ge}})x = a(x),$$

ignoring the negligibly small quadratic term in the polynomial which causes a minimal deviation from the linear fit. The value of the misfit parameter associated with a Si$_{1-x}$Ge$_x$ alloy layer and a Si substrate is defined as

$$f_m(x) = \frac{a(x) - a_{\text{Si}}}{a_{\text{Si}}} = 0.042x.$$

Hence, the lower the Ge concentration in the Si$_{1-x}$Ge$_x$ layer, the smaller the lattice mismatch between the alloy layer and the Si substrate. If the mismatch between the substrate and the epilayer is sufficiently small (< 10%), it is energetically desirable for the first atomic layers that are deposited to maintain full atomic bonding to the substrate, forcing the epilayer to be strained such that the in-plane lattice constant of the film matches the substrate lattice parameter as shown in Figure 2.3 [16-20].

![Figure 2.3](image)

*Figure 2.3* Tetragonal strain of a mismatched overlayer by pseudomorphic epitaxy [12]:

a) $a_{\text{overlayer}} > a_{\text{substrate}}$: biaxial compression and uniaxial extension,
b) $a_{\text{overlayer}} < a_{\text{substrate}}$: biaxial extension and uniaxial compression.
Thus, the mismatch is accommodated entirely by the elastic tetragonal strain in the epilayer, and the growth of the epilayer is then said to be coherent, pseudomorphic or commensurate. The substrate lattice remains essentially undistorted due to it being both much thicker and stiffer. Since the lattice constant of Ge and Si$_{1-x}$Ge$_x$ alloys are larger than that of Si, pseudomorphic Si$_{1-x}$Ge$_x$ layers grown on a Si substrate have biaxial in-plane compression and uniaxial extension normal to the interface. The reverse case occurs when Si layers are grown coherently on a Si$_{1-x}$Ge$_x$ buffer, i.e. the epilayer will experience biaxial extension and uniaxial compression.

As the thickness of the pseudomorphic layers increases, so does the integrated strain energy, and at some point, it will become larger than that of an alternative arrangement, the relaxed configuration, where both the substrate and the grown layers remain strain free. In this arrangement, the lattice mismatch is accommodated by distortions near the interface and arrays of incompletely bonded atom rows known as misfit dislocations (Figure 2.4). The dangling bonds of those atoms can become an unwanted trap or leakage site for the carriers in the active region. In principle, ideal layers of Si and Ge could be grown on top of each other by confining the misfit dislocations entirely at the epitaxial interface where their negative effect on the carrier transport would be kept minimal. However, another unfavourable effect occurring during the process of strain relaxation is the nucleation of threading dislocations from the misfit dislocations which propagate through the thickness of the epitaxial layer, influencing the active region and reducing its quality drastically.

For the growth of pseudomorphic layers on relaxed buffers, the most important material parameter is therefore the critical thickness $t_c$ above which strain relaxation by the generation of misfit dislocations commences [18, 19]. Films that are thinner than $t_c$ cannot relax because the elastic energy stored in such a homogeneously strained layer is lower than the energy associated with the local distortion around a
misfit dislocation [18]. Above $t_c$, however, misfit dislocations become energetically favourable and provide partial strain relaxation of the film. Theoretical studies into the nature of pseudomorphic growth show that the critical thickness of the epitaxial layer is strongly dependent on the misfit parameter and thus on the Ge concentration in the epilayer, i.e. an increase of the Ge content in the alloy results in a decrease of the critical thickness [21, 22].

Initial calculations of $t_c$ included a somewhat different approach, assuming that the interfacial misfit dislocations are introduced through the glide of grown-in threading dislocations [19]. The results obtained from both approaches are similar in value, and the small divergence is explained by the difference in the presumed expression for the dislocation energy [23]. Extensive experimental studies of the crystalline quality, strain, dislocations and surface morphology of epitaxial layers to determine the critical thickness show significant deviations from the theoretical values in many cases, suggesting that the sensitivity of the measurement technique is important in determining the correct value of $t_c$ [3, 21-25]. But even with more sensitive methods, experimentally obtained values for $t_c$ are generally larger than the calculated values because in most cases, the layers are in metastable states in which the film is fully strained, but relaxation may ultimately occur upon extended thermal processing [3, 26-28]. In general, the theoretical values indicate the equilibrium critical thickness $t_e$ which depends only on the Ge concentration in the alloy. The maximum thickness of the epitaxial layer, however, is called the metastable critical thickness $t_m$ and depends also on the growth temperature (Figure 2.5). For higher growth temperatures (e.g. 900°C), $t_m$ approaches $t_c$, but a significant difference between both limits at lower growth temperatures (e.g. 500°C) can be exploited to increase the layer thickness and thus the active region of a device considerably above $t_c$.

![Figure 2.5 Kinetically limited critical thickness for strained SiGe layers on Si substrates. The upper curve is the metastable critical thickness $t_m$ for growth at 500°C [29], and the lower curve is the equilibrium critical thickness $t_e$ [19]. Below $t_m$, strained layer growth represents the lowest energy states and the strain is stable. Between $t_c$ and $t_m$, non-equilibrium conditions allow strained layer growth for which the nucleation and propagation of misfit dislocations is kinetically suppressed, but partial strain relaxation may occur upon subsequent heat treatment. Above $t_m$, misfit dislocations form because their associated energies occupy the lowest energy states.](image-url)
Although Figure 2.5 shows the two curves of the critical thickness for strained SiGe layers on Si substrates, the same curves also apply for the case of strained Si layers on SiGe buffers. To achieve such a configuration, a strain-relaxed SiGe buffer layer has to be grown on a Si substrate to allow for a subsequent growth of a strained Si layer. The most straightforward method of implementing a strain-adjusted buffer is the growth of a SiGe layer with constant composition and a thickness exceeding \( t_m \) by a large enough margin to allow for a reasonably high degree of strain relaxation. However, such a buffer layer is generally associated with a high density of threading dislocations penetrating through the buffer and ending at the interface to the strained Si layer. A solution to this problem is the introduction of graded buffer layers in which the Ge content is linearly graded from zero at the substrate interface towards the desired composition in the topmost layer, taking into account that the slowest grading rates achieve the lowest dislocation densities [30-33]. If the linearly graded buffer layer is thick enough to be almost fully relaxed, it is followed by a second buffer layer in which this final composition is kept constant. Thus, the second layer is virtually strain free and provides a fully relaxed buffer, free of misfit dislocations, for the succeeding growth of a strained Si layer. It also serves as a spacer between the active region and the topmost dislocations of the graded buffer layer. Dislocation densities of compositionally graded structures are four to six orders of magnitude lower than those of uniform layers.

### 2.1.2 Band Structure

In semiclassical quantum mechanics, the physical system of electrons is fully described by a set of wavefunctions which determines the energy and momentum of each particle in the system. In order to find the wavefunctions for a quantum mechanical system, the time-dependent Schrödinger equation has to be solved for all electrons. Following this approach, it is possible to obtain a complete description of the band structure in solids. However, solving the Schrödinger equation directly is by far too complicated, but various simplifications and approximations can be introduced to obtain a more manageable system.

**The One-Electron Approximation**

Assuming the potential does not depend on time, the thus time-independent energy levels of the electronic states in a quantum mechanical system (i.e. atoms, molecules or solids) are given by the time-independent Schrödinger equation

\[
H\Psi(r_i) = E\Psi(r_i)
\]  

(2.3)

where \( r_i \) is the three dimensional co-ordinate of the i-th electron, \( H \) is the Hamiltonian and \( E \) the energy eigenvalue. General solutions of this equation are given by the eigenfunction
\[ \Psi(r) = e^{ikr} \]  \hspace{1cm} (2.4)

where \( k \) is the wavevector and \( \Psi(r) \) is known as the wavefunction. The electrons in the completely filled inner shells are expected to be tightly bound to the nucleus and these electrons will therefore be little effected by outside forces. As a result, ions consisting of a nucleus and its surrounding inner shells can be introduced, which limits the many-body problem to only the outer shell electrons. In the adiabatic approximation in which the motion of the nuclei is taken to be much slower than the motion of the electrons (by a factor of about 100), the Schrödinger equation is split into a purely ionic and a purely electronic equation. Making the additional assumption of having a perfect lattice and neglecting any lattice oscillations, the electronic Hamiltonian takes the form [11, 34]

\[ H_e = \sum_k \left( \frac{-\hbar^2}{2m_0} \nabla_k^2 + \sum_{i,l} V(r_i - R_l) + \frac{e^2}{4\pi \varepsilon_0} \sum_{i,j} \frac{1}{|r_i - r_j|} \right) \]  \hspace{1cm} (2.5)

where \( i \) and \( j \) label the electrons and \( R \) runs over all lattice sites \( l \). In this form, the Hamiltonian is still insoluble as it has an explicit dependency on the co-ordinates of all outer shell electrons involved. If the electron-electron interaction is averaged, any deviation from this average can be regarded as a small perturbation. Assuming the perturbations are small enough, they can be neglected and each electron reacts independently with the lattice of ions and the many electron wavefunction can be written as the product of single electron wavefunctions (Hartree approximation)

\[ \Psi(r, R) = \prod_i \psi_i(r_i, R), \]  \hspace{1cm} (2.6)

providing the occupation of the one-electron states is in accordance with the Pauli exclusion principle. Hence, the one-electron Schrödinger equation is given by [11, 34]

\[ -\frac{\hbar^2}{2m_0} \nabla_i^2 + \sum_l V(r_i - R_l) \psi_i(r_i) = E \psi_i(r_i) \]  \hspace{1cm} (2.7)

with each electron having the momentum

\[ p = -i\hbar \nabla = \hbar k \]  \hspace{1cm} (2.8)

and a corresponding energy

\[ E = \frac{\hbar^2 k^2}{2m_0}. \]  \hspace{1cm} (2.9)

A further complication arises from the interaction of the magnetic momentum (spin) of an electron with its own angular momentum giving rise to an energy perturbation. To obtain the correct electron energies, an additional potential term in the Hamiltonian representing the spin-orbit coupling has to be taken into account [11, 12,
34]. A qualitative explanation of the interaction can be given by looking at a classical system in which a spinning electron moves around a charged nucleus. In the electron rest frame, the nucleus is moving around the electron which is thus encircled by a current flow and consequently experiences an external magnetic field.

**THE BLOCH THEOREM**

For a perfectly periodic potential, the eigenfunction of the Schrödinger equation is a Bloch function which is the product of a plane wave $e^{i\mathbf{k}\mathbf{r}}$ and a function $u_n(r)$ which has the same periodicity as the potential of the crystal $U(r)$ so that $u_{nk}(r + R) = u_{nk}(r)$ where $R$ is a vector of the Bravais lattice. The form of the Bloch function is then given by [10-12]

$$\psi_{nk}(r) = u_{nk}(r)e^{i\mathbf{k}\mathbf{r}}$$

where $n$ labels the band and $k$ is the wavevector of the electron in the first Brillouin zone. The other important consequence of the periodic potential is that the Bloch function only needs to be given for the first Brillouin zone to describe the entire lattice [10-12]

$$\psi_{nk}(r + R) = \psi_{nk}e^{i\mathbf{k}\mathbf{r}}.$$  

As an interpretation of eqn. (2.11), the Bloch theorem does not predict the wavefunction to be periodic in space (in fact, it does change from one unit cell to another), but it asserts that the probability density given by $|\psi|^2$ is periodic and follows the underlying Brillouin zones, as should be expected.

Rewriting the one-electron Schrödinger equation in terms of the cell-periodic functions $u_{nk}(r)$, it follows

$$H_k u_{nk}(r) = \left( E_n - \frac{\hbar^2 k^2}{2m_0} \right) u_{nk}(r).$$

It can be seen immediately that the application of the Bloch theorem has an advantageous effect on the solvability of the one-electron Schrödinger equation. It is evident that in eqn. (2.7), $\psi(r)$ has to be solved for the whole crystal whereas in eqn. (2.12), $u_{nk}(r)$ only has to be solved within one Wigner-Seitz cell.

By ignoring the $k$-dependence in $H_k$, an interesting observation can be made without actually solving for the eigenvalues as one finds parabolic energy bands with energy

$$E_n(k) = \frac{\hbar^2 k^2}{2m_0} + E_0^0.$$  

A very important implication of the Bloch theorem is that an electron propagates without scattering in the perfectly periodic background potential that the crystal presents. Each electronic state is an extended wave which occupies the entire crystal
where the quantity $\hbar \mathbf{k}$ of eqn. (2.8) is associated with the momentum of the electron in the crystal. Although $\hbar \mathbf{k}$ contains the effects of the internal crystal potentials and is therefore not the true electron momentum, it responds to the external forces as if it was the effective momentum of the electron. Hence, $\hbar \mathbf{k}$ is often called the crystal momentum and represents a tremendous simplification of the electron problem in crystalline materials. Once the $E$ vs $k$ relation is established, the background potential $U(r)$ does not have to be considered any longer, and the electrons can be treated as if they were free. In scattering processes, the crystal momentum usually behaves as though it obeyed momentum conversation laws.

**ENERGY BANDS**

As indirect bandgap semiconductors, the band structures of both Si and Ge display different $k$-points for the top of the valence band and the bottom of the conduction band. Whereas the valence band maximum is located at the $\Gamma$-point in both semiconductors, the conduction band minimum is positioned at the $L$-point along the $(111)$ direction in Ge (usually referred to as $\Delta$ minimum) and close to the $X$-point along the $(100)$ direction in Si ($\Delta$ minimum). Due to the symmetry of the fcc lattice, there are eight degenerate $L$-points and six degenerate $X$-points, consequently leading to eight conduction band edge valleys in Ge and six in Si. The wavefunction near the band edge is highly anisotropic due to a strong mixture of $|s\rangle$-type and $|p\rangle$-type functions. Thus, the six band edge valleys in Si can be described by six ellipsoids of constant energy with their centres being located about 85% of the way from the $\Gamma$-point to the six corresponding X-points. For Ge, the centres of the resulting eight ellipsoids are fixed at the $L$-points, leading to eight half-ellipsoids within the Brillouin zone boundary. Both configurations of the ellipsoids are shown in Figure 2.6 where for comparison, the spherical surface of constant energy in GaAs is also shown.

![Figure 2.6](image)

**Figure 2.6** The shape of constant-energy surfaces in the conduction band minima in Si, Ge and GaAs. For Si, there are six ellipsoids of revolution along the $(100)$ axes ($\Delta$ minima), for Ge, there are eight half-ellipsoids along the $(111)$ axes ($\Delta$ minima) with the Brillouin zone boundaries being at the centre of the ellipsoids and for GaAs, the constant-energy surface is a sphere at the zone centre ($\Gamma$ minimum) [35].
Since the areas of constant energy in the conduction band minima consist of six (Δ) and eight (L) ellipsoids of revolution for Si and Ge, respectively, the electrons in both materials are completely described by two mass parameters where \( m_l \) is the longitudinal mass along the symmetry axis of the ellipsoid (⟨100⟩ direction in Si and ⟨111⟩ direction in Ge), and \( m_t \) is the transverse mass within the plane normal to the symmetry axis. As the \( E(k) \) relation in eqn. (2.9) has to be fulfilled, each of the energy ellipsoids has the form \[ E(k) = \frac{\hbar^2 k_x^2}{2m_l} + \frac{\hbar^2 (k_y^2 + k_z^2)}{2m_t}, \] (2.14)

consisting of a longitudinal and a transverse part regarding the orientation of the ellipsoids. The index \( x' \) of the wavevector \( k \) corresponds to the longitudinal direction of the ellipsoid and the indices \( y' \) and \( z' \) to its transverse plane. In both materials, the longitudinal mass is significantly larger than the transverse mass. Thus, the two ellipsoids in Si that are orientated along the growth direction have a light in-plane mass, whereas the four ellipsoids within the plane have a large in-plane mass. A similar approach to determine the in-plane masses of the eight ellipsoids in Ge with respect to the growth direction is not possible due to the fact that their orientation is in ⟨111⟩ direction and not along or in the plane normal to the growth direction.

According to the tight binding method, the topmost valence band at the zone centre in tetrahedral semiconductors with a diamond structure such as Si and Ge is three-fold degenerate (Γ₅v), corresponding to the degeneracy of the \( p_x \), \( p_y \), and \( p_z \) state as the central cell part of the wavefunction is primarily \( |p⟩ \)-type. If spin is included into the calculation, this degeneracy increases to being six-fold. The strong interaction of the spin with the orbital motion of the electron results in the so-called spin-orbit coupling which splits the bands by an amount \( \Delta \gamma \). With respect to the total angular momentum at \( k=0 \), the splitting occurs between the \( j=3/2 \) states (Γ₅⁺ symmetric) and the \( j=1/2 \) states (Γ₇ symmetric). The four-fold degenerate states \( j=3/2 \) at the zone centre have different curvatures resulting in two heavy hole (HH) bands with a smaller curvature and two light hole (LH) bands with a larger curvature. The two-fold degenerate states \( j=1/2 \) are called split-off states (SO). With reference to the average band position, the LH and the HH bands are shifted up by an amount \( \Delta \gamma/3 \), while the split-off band is shifted down in energy by \( 2\Delta \gamma/3 \). The nature of the hole states can be described in terms of the angular momentum states \( \phi_{j,m} \) with \( j \) being the total angular momentum and \( m \) the projection of the angular momentum along the \( x \)-axis:

- heavy hole states: \( \phi_{3/2,3/2} \) and \( \phi_{3/2,-3/2} \) or \( |3/2,±3/2⟩ \),
- light hole states: \( \phi_{3/2,1/2} \) and \( \phi_{3/2,-1/2} \) or \( |3/2,±1/2⟩ \),
- split-off hole states: \( \phi_{1/2,1/2} \) and \( \phi_{1/2,-1/2} \) or \( |1/2,±1/2⟩ \).

Figure 2.7 shows the heavy and light hole bands in the \( xy \)-plane which are degenerate at \( k=0 \) and have the form of warped spheres. As a consequence, the effective masses of both HH and LH bands are strongly dependent on the crystal direction. Whereas the effective mass of the HH band is lightest along the main axes and heaviest along the diagonals, the situation is reversed in the LH band where the
The effective mass is lightest along the diagonals and heaviest along the main axes. Since the split-off band is spherical in the \( xy \)-plane, its effective mass is independent of the crystal direction, and the SO band has therefore been omitted from this illustration.

![Figure 2.7](image)

**Figure 2.7** Constant-energy surfaces for valence bands in the \( xy \)-plane. Only HH and LH bands are shown as the SO band is spherical in this plane [36].

This pure description of the hole states is strictly valid only at the \( \Gamma \)-point since the LH and HH states mix strongly at \( k \neq 0 \) and thus become nonparabolic. In Figure 2.8, a schematic description of the conduction band for direct and indirect bandgap semiconductors is shown together with the degeneracy and the spin-orbit coupling of the valence bands. In general, the \( E \) vs \( k \) relation of a semiconductor is shown along the principal axes of high symmetry, i.e. from the \( \Gamma \)-point to the \( X \)-point along the \( (100) \) axis or from the \( \Gamma \)-point to the \( L \)-point along the \( (111) \) axis.

![Figure 2.8](image)

**Figure 2.8** Schematic \( E(k) \) diagram of the central cell states in direct and indirect bandgap materials.

An overview of selected structural, band and electrical parameters for Si and Ge at room temperature is given in Table 2.1. The corresponding values for GaAs serve for comparison with a direct bandgap material. Since the conduction band edge states in GaAs have an \( |s\rangle \)-type symmetry at the \( \Gamma \)-point, the wavefunction is quite isotropic.
leading to a spherical surface of constant energy centred around \( k=0 \). Hence, the energy relation can be derived directly from eqn. (2.9) to be

\[
E(k) = \frac{\hbar^2 k^2}{2m^*}.
\]

where the electrons in the lowest conduction subband have only one effective mass. As for the composition of the valence band, it consists of the standard HH, LH and SO states since the principle of degeneracy and spin-orbit coupling also applies to the valence band of GaAs.

<table>
<thead>
<tr>
<th>lattice type</th>
<th>Silicon</th>
<th>Germanium</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>lattice constant ( a_0 ) (Å)</td>
<td>5.431</td>
<td>5.657</td>
<td>5.653</td>
</tr>
<tr>
<td>direct bandgap ( E_{gd} ) (eV)</td>
<td>3.2</td>
<td>0.8</td>
<td>1.424</td>
</tr>
<tr>
<td>indirect bandgap ( E_{gd} ) (eV)</td>
<td>1.12 (( \Delta ))</td>
<td>0.66 (L)</td>
<td>1.734 (L)</td>
</tr>
<tr>
<td>spin-orbit splitting ( \Delta_0 ) (eV)</td>
<td>-0.044</td>
<td>0.296</td>
<td>-0.34</td>
</tr>
<tr>
<td>electron mass ( m_0 )</td>
<td>( m_t )</td>
<td>0.19</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>( m_l )</td>
<td>0.98</td>
<td>1.64</td>
</tr>
<tr>
<td>hole mass ( m_0 )</td>
<td>( m_{hh} )</td>
<td>0.49</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>( m_{lh} )</td>
<td>0.16</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>( m_{so} )</td>
<td>0.29</td>
<td>0.084</td>
</tr>
<tr>
<td>bulk mobility (cm²/Vs)</td>
<td>electrons</td>
<td>1450</td>
<td>3900</td>
</tr>
<tr>
<td></td>
<td>holes</td>
<td>505</td>
<td>1800</td>
</tr>
</tbody>
</table>

Table 2.1 Properties of unstrained, undoped bulk Si, Ge and GaAs at room temperature. The effective masses are band edge masses, which apply only for low-doped structures at low electric fields; \( m_0 \) is the free electron mass [35].

At room temperature, the movement of the electrons and holes in Si and Ge are strongly effected by the presence of ionized impurities and acoustic phonons which act as scattering centres for the carriers and thus limit their mobilities. The mobility from the ionized impurity scattering \( \mu_i \) and from the acoustic phonon interaction \( \mu_{ac} \) can be approximated by

\[
\mu_i \propto (m^*)^{-1/2} N_i^{-1} T^{3/2} \quad \text{and} \quad \mu_{ac} \propto (m^*)^{-5/2} T^{-3/2}
\]
with $N_i$ being the density of the ionized impurities and $m^*$ the effective mass. The temperature dependence of both mobilities is opposite in character, i.e. the mobility limited by ionized impurities increases with temperature (the impurities become less effective scattering centres), whereas the mobility limited by acoustic phonon scattering decreases with increasing temperature (a growing phonon population results in a higher scattering rate). Taking both scattering mechanisms into account, the combined mobility $\mu$ is then given by Matthiessen's rule as

$$\mu = \left(\mu_i^{-1} + \mu_{ac}^{-1}\right)^{-1}. \quad (2.18)$$

Clearly, scattering from ionized impurities dominates the mobility of electrons and holes at low temperatures where acoustic phonon scattering is very weak. However, acoustic phonon scattering growth stronger as the temperature increases and eventually takes over as the dominant scattering process. Interactions with optical phonons play an additional and important role in the scattering process of polar semiconductors such as GaAs, and the dependence of the mobility on the effective mass and the temperature can be described by

$$\mu_{opt} \propto (m^*)^{-3/2} T^{1/2}. \quad (2.19)$$

The strong influence of the effective mass on the carrier mobility explains the big difference between the high electron mobilities of Ge and the much lower values for Si. Due to the degeneracy of the lowest conduction subband in Si and Ge, their electrons are exposed to an additional scattering process called inter-valley scattering. The absence of this scattering mechanism in GaAs leads to a much higher electron mobility compared to that of Ge although the effective electron masses of both materials are quite similar. The very low effective hole mass in conjunction with a rather large spin-orbit splitting in Ge are responsible for one of the best hole mobilities among all semiconductor materials.

### 2.1.3 STRAIN EFFECTS

In SiGe alloys, the structure of the conduction band stays Si-like for a Ge concentration up to 85%. Only then do the electrons occupy the eight constant energy ellipsoids of the conduction band minima at the L-point. In most Si/SiGe heterostructures, the Ge content does not exceed 85%, in fact, with a common Ge concentration of around 30%, the value stays well below this point. Hence, all further calculations can be done considering the six-fold degenerate conduction band of Si. The only exception to this generalization are hole gas structures which either employ a pure Ge channel or which are grown on a Ge substrate. However, since the purpose of those structures is the confinement of holes, they deal exclusively with the valence and not the conduction band.

In presence of uniaxial stress in growth direction, the six-fold degenerate conduction band ($\Delta_6$) is split into sets of two-fold ($\Delta_2$) and four-fold ($\Delta_4$) valleys, two ellipsoids
along the growth direction with a light in-plane mass and four ellipsoids in the plane normal to the growth direction with a large in-plane mass. These splittings are expressed with respect to the average band position which is unaffected by the uniaxial components of the strain. Under in-plane compressive strain, the two-fold subband is lifted and the four-fold subband is lowered in energy compared to the unsplit conduction band. Under biaxial tensile strain, the direction in which the split subbands are shifted is reversed, i.e. the four-fold minima are lifted and the two-fold valleys are lowered in energy. For lateral transport, it is desirable to have a light in-plane electron and hole effective mass in order to enhance the carrier mobilities. Achieving electron confinement by growing a Si layer under tensile strain, for which the two-fold subband with its light in-plane mass is lowest in energy, is thus favourable to a compressively strained SiGe channel for which the four-fold subband with its large in-plane mass is lowest in energy. Since the conduction band minima of the eight-fold degenerate L-band are orientated in (111) direction and not in growth direction or in the plane normal to it, the L-band remains unaffected by uniaxial strain.

The structure of the valence band is also effected by uniaxial strain, resulting in the splitting of the HH and LH states. Under biaxial compression, the upper valence band, which is lowest in energy, is occupied by the HH states $|3/2, \pm 3/2\rangle$ with the LH states being shifted up in energy. However, under biaxial tensile strain, the LH states $|3/2, \pm 1/2\rangle$ are shifted down in energy below the HH band and become lowest in energy to occupy the band edge states. In addition to splitting the subbands, uniaxial strain produces in-plane mass inversion, thus altering the character of the valence band in lateral direction as shown in Figure 2.9.

![Figure 2.9](image)

**Figure 2.9** Schematic diagram of the valence subbands showing the character of the HH and LH states for the in-plane and out-of-plane direction for unstrained Si (top), a Si layer under tensile strain (bottom left) and a Si layer under compressive strain (bottom right).

Considering lateral transport under compressive strain, the ground state HH subband lies lowest in energy, but due to its larger curvature in (100) direction, its in-plane mass at the zone centre is reduced to that of a LH subband. Concurrently, the
in-plane mass of the energetically higher lying LH subband is augmented to that of a HH subband. The same principle of mass inversion applies to the valence subbands under tensile strain where the LH states are lowest in energy but the in-plane mass is HH in character, and the HH states have a LH in-plane mass. Consequently, compressively strained SiGe channels exhibit lower effective hole masses in lateral direction than Si layers under tensile strain.

The influence of uniaxial stress normal to the plane on the splitting of the conduction and the valence band is shown in Figure 2.10 along with the effect of strain on the position of the HH and LH band with respect to each other. It is important to be aware of the different character the two valence bands possess in growth direction and in the plane normal to it. Because of its simplicity and for a better understanding of the concept of the splitting, only a schematic of the splitting and an idealized $E(k)$ diagram of the valence band states in growth direction are displayed. Any effect of the hydrostatic strain on the band edges has been omitted, since it does not contribute to the lifting of the degeneracy and thus the splitting of the bands but merely shifts the energy of all subbands ($\Delta_e$-band and HH-/LH-/SO-bands, each by an equal amount).

![Diagram showing splitting of the valence and conduction band at the Si/SiGe interface for both materials being under biaxial strain (top). Also shown is the effect of strain on the HH and LH states in the valence band (bottom). Under biaxial tensile strain, the LH states move above the HH states which results in an anticrossing behaviour of the hole bands. HH and LH nomenclatures are chosen for effective masses in growth direction.](image)

**Figure 2.10** Splitting of the valence and conduction band at the Si/SiGe interface for both materials being under biaxial strain (top). Also shown is the effect of strain on the HH and LH states in the valence band (bottom). Under biaxial tensile strain, the LH states move above the HH states which results in an anticrossing behaviour of the hole bands. HH and LH nomenclatures are chosen for effective masses in growth direction.

Tensile strain represents a special case of the valence band splitting for the momentum in growth direction as the LH band moves above the HH band. Initially, a crossing of the LH and HH bands is expected, however, higher-order perturbation terms lead to the anticrossing behaviour of the hole bands as shown in Figure 2.10,
thus increasing the LH band mass away from $k=0$ to that of the HH band and correspondingly decreasing the HH band mass to that of the LH band. This effect causes strong nonparabolicity of the hole bands and can occasionally result in a hole band having a negative curvature and hence a negative effective mass. Equivalently, the in-plane LH band moves above the HH band for compressive strain, thus causing anticrossing of the hole bands in the plane normal to the growth direction.

Whereas the direction of the energy shift of conduction and valence bands depends solely on the type of strain the layer is subjected to, the amount by how much the split bands are separated is dependent on the Ge concentration in the buffer or the strained layer, respectively. Exact knowledge of the energies of each subband with regard to the Ge content is therefore vital for the design of structures for either electron or hole confinement. As illustrated in Figure 2.11, the subbands with the larger in-plane masses (in the conduction as well as in the valence band) do not vary much in energy, neither for the two different types of strain nor with changing Ge concentration. A comparatively bigger change in energy can be seen for the subbands that exhibit a light in-plane mass whereby an increase of the Ge concentration results in an enlargement of the energy shift.

![Figure 2.11](image)

*Figure 2.11 Effect of strain on conduction and valence bands for strained Si and SiGe layers. In both cases, the strain increases with growing Ge concentration. LH and HH nomenclatures are chosen for in-plane effective masses [37].*

In order to design structures with improved electron (hole) mobilities, two conditions have to be satisfied. Firstly, the energy offset has to be large enough to enable the confinement of electrons (holes). For strained SiGe layers, the offset is given by the difference between the lowest conduction (highest valence) band and the crossover of the corresponding split bands in Figure 2.11. The gap between valence and conduction band at the crossover point is equivalent to the bandgap of the relaxed Si buffer. In the case of strained Si layers, the offset cannot be determined equally straightforward from the energy curves since the band edges of the relaxed
SiGe substrate also change with varying Ge concentration. Secondly, the electrons (holes) that are lowest in energy within the confinement should have a light in-plane mass. According to the energy curves, it is evident that the energy offset is only sufficiently large in strained Si (SiGe) layers to permit electron (hole) confinement. Additionally, electrons (holes) with a low in-plane mass are also lowest in energy for strained Si (SiGe). Hence, both conditions are fulfilled for the same structure, but the structures are different for electrons and holes. The influence of the Ge concentration and thus the strain on the conduction and valence band edges has also a direct effect on the bandgap of the strained material as shown in Figure 2.12. The calculated transition energies were corroborated by optical absorption spectroscopy between the lowest conduction band states and the highest valence band states and thus represent the minimum band gaps for both possible cases [38, 39].

**Figure 2.12** Band gap variations of strained Si and SiGe layers against Ge concentration at 90K [38-40]. The shaded area in both curves represents the calculative uncertainty. The top curve gives the band gap energy for unstrained bulk alloys [41], which show a crossover from the Si-like to the Ge-like band structure at \( x = 0.85 \).

Values for the band gap of strained SiGe were calculated between the \( |3/2, \pm 3/2| \) valence band states (in-plane LH band) and the four-fold degenerate conduction band edge, those for strained Si between the \( |3/2, \pm 1/2| \) valence band edge (in-plane HH band) and the two-fold conduction subband. For comparison, the band gap of unstrained bulk SiGe is also shown which falls with increasing Ge concentration and has a kink where the character of the conduction band changes from having a \( \Delta \)-symmetry to having an L-symmetry. Looking at the curvature of the conduction and valence band edges in Figure 2.11, it is obvious that most of the band gap change in the strained SiGe layer stems from the variation of the in-plane LH band, whereas in the strained Si layer, it is predominantly the two-fold degenerate conduction subband contributing to the band gap shift. It is important to note that the band gap for strained Si and strained SiGe is always smaller than the band gap of the corresponding relaxed bulk alloy with the same Ge concentration. This information is essential for the determination of the band alignment at the interface between a relaxed buffer and a strained layer.
2.1.4 BAND ALIGNMENT

So far, the emphasis has been on investigating the effect of uniaxial strain on the conduction and valence band edges and on the band gap to establish which configuration is preferential for enhanced electron and hole mobilities. However, in order to achieve an improvement in their transport properties, the preferred carriers have to be separated and confined in a potential well next to the interface of the heterostructure. Exact knowledge of the band alignment is important for the calculation of the discontinuity of the valence and conduction band edge to determine whether hole or electron confinement at the interface is possible. The configuration of the band alignment of Si/SiGe heterointerfaces depends strongly on the built-in strain and the bandgap difference between Si and SiGe. For simplicity, only one of the two materials of the heterointerface in Figure 2.13 is assumed to be under stress, the other one is presumed to be totally relaxed. The type of strain present in the system determines the configuration of the band alignment which can be either type I or type II in character [24, 39]. For a compressively strained SiGe layer which is grown on top of a relaxed Si buffer, the resulting band alignment follows a type I configuration where the narrower SiGe alloy bandgap falls within the wider Si bandgap. If a Si layer is grown under tensile strain on a relaxed SiGe buffer, the band alignment is type II in character in which both the conduction and the valence band of Si lie below the corresponding SiGe alloy bands. In both cases, the valence band of SiGe remains inside the bandgap of Si, i.e. the valence band of SiGe is always lower in energy than that of Si.

![Diagram of band alignment types](image)

Figure 2.13 The two types of band alignment in the Si/SiGe heterosystem under biaxial strain (qualitative): type I for compressively strained SiGe on relaxed Si and type II for tensile strained Si on relaxed SiGe. The solid lines represent the lowest conduction band and highest valence band in each material, the dashed lines the corresponding split bands. $\Delta_{c}$ and $\Delta_{v}$ denote the band offsets where carrier confinement is feasible. LH and HH nomenclatures are chosen for in-plane effective masses.

It is obvious that improved hole mobilities can only be achieved in a type I band alignment because of the LH band being lowest in energy and due to the fact that hole confinement in a type II configuration is not possible at all. Theoretically,
electrons can be confined in both types, but only in a type II alignment does the subband with the light in-plane mass occupy the band edge states.

Most importantly though, the deciding factor whether or not any carriers can be confined at the interface is the absolute value of the band edge discontinuity which itself is dependent on the Ge content in both materials and the built-in uniaxial and hydrostatic strain [42-45]. Figure 2.14 shows a complete set of calculated energy contour plots for the offsets between the topmost valence bands \( E_{v}^{\text{max}}(x_{i}) - E_{v}^{\text{max}}(x_{s}) \) and the lowest conduction bands \( E_{c}^{\text{min}}(x) - E_{c}^{\text{min}}(x_{s}) \) of a strained \( Si_{1-x}Ge_{x} \) layer on an unstrained \( Si_{1-x}Ge_{x} \) substrate. Despite a good overall agreement with experimental values of the band offset, an uncertainty of about 100 meV has to be allowed for due to the approximations and interpolations involved in the calculations.

Both diagrams can be divided into two zones, separated by the diagonal which runs through the origin and marks the unstrained case in which substrate and layer have the same Ge content \( x = x_{s} \). In the upper left zone \( x > x_{s} \), the layer is grown under compressive strain (type I), in the lower right zone \( x < x_{s} \), it is grown under tensile strain (type II). Looking first at the more simplistic energy contours of the valence band offset, the relevant area is the upper left zone with the in-plane LH band of the strained layer lying above the topmost valence band of the relaxed substrate. Correspondingly, the significant area for the energy contours of the conduction band offset is the lower right zone where the lowest conduction band of the strained layer lies below that of the unstrained substrate. Because of the changeover of the conduction band from a \( \Delta \)- to an L-band above a Ge concentration of 85%, the contour plot of this diagram is more complex than that for the valence band offset. For \( x > x_{c} \) and \( x < 0.85 \) and \( x < 0.5 \), the conduction band offset remains rather constant, but at around 20 meV, it is negligibly small and represents a basically flat conduction band alignment at room temperature where the thermal energy is \( k_{B}T \approx 26\text{meV} \). In general, most heterostructures for electron and hole confinement employ materials
with either the layer composition $x$ or the substrate composition $x_s$ being zero. In these two cases, the determination of the corresponding band offset energy is reduced to the energy values along the ordinate or abscissa, respectively. For strained Si layers on relaxed Si$_{1-x}$Ge$_x$ substrates, the appropriate conduction band offset energies can be read along the $x$-axis and for strained Si$_{1-x}$Ge$_x$ layers on unstrained Si substrates, the relevant valence band offset energies are positioned along the $y$-axis.

### 2.2 Physics of Two-Dimensional Systems

All theory dealt with in the previous section serves as a basis to comprehend the principal idea of enhanced carrier transport in two-dimensional (2D) heterosystems. A fundamental knowledge of the band structure of semiconductors, the effect of strain on the valence and conduction bands and their alignment at the heterointerface is vital to develop the concept of two-dimensional carrier transport and thus to understand how an improvement of the electron and hole mobilities can be achieved by reducing the freedom of the carriers from three to two dimensions. This section covers the physical background of two-dimensional heterosystems, in particular the structure of quantum wells, the concept of modulation doping and the energy dependence of the density of states in ideal and real 2D systems. Although most of the physics is valid in general and can be applied to most two-dimensional heterostructure systems, the reference is always to the Si/SiGe heterosystem.

#### 2.2.1 Modulation Doping and Quantum Wells

Having discussed the different possible types of band alignment for the Si/SiGe heterosystem and their capability to confine either electrons or holes, it is now possible to introduce the idea of improved carrier mobilities in two dimensions and to describe the formation of two-dimensional electronic systems with enhanced transport properties. There are principally two ways of creating such two-dimensional carrier gas (2DCG) systems in heterostructures, both exploiting the now well-established principle of the conduction and valence band offsets at the interface. One way is by trapping mobile carriers in a potential well at the heterointerface through unilateral doping of the unstrained material. Similarly, the other way exploits the energy offset in the band alignment to create two potential barriers in order to confine the carriers inside a specially designed quantum well (QW). Depending on the specific structure of the system, either electrons or holes can be confined in such a potential or quantum well to form a two-dimensional electron gas (2DEG) or a two-dimensional hole gas (2DHG), respectively.
MODULATION DOPING

Potential wells for both electrons and holes can be established at a heterointerface by employing the modulation doping technique which spatially separates the mobile carriers from their parent impurities [47-50]. To realize modulation-doped heterostructures, the unstrained material, for which the band edge for the corresponding carriers is higher in energy, has to be doped accordingly (n-type for 2DEG and p-type for 2DHG structures), while keeping the strained layer free from intentional doping. Some of the carriers introduced by the dopants transfer into the energetically lower-lying conduction band minimum or valence band maximum of the strained and undoped material, leaving behind ionized impurities in the doped material. Due to the charge redistribution, an electrostatic potential is created which forces the bands to bend in the vicinity of the interface until an equilibrium is reached, thus enabling a constant Fermi-level across the whole structure. As illustrated in Figure 2.15, the resulting potential well inside the strained layer has a triangular-like shape, giving rise to quantum effects for the trapped carriers in the potential well when the width of the potential well becomes comparable to the de Broglie wavelength of the confined carriers. The energy spectrum perpendicular to the interface is no longer continuous but rather a series of discrete energy levels, whereas the energy spectrum parallel to the interface remains continuous, enabling carrier transport within the plane parallel to the interface.

![Figure 2.15](image)

The spatial separation of the carriers from their parent impurities results in profoundly reduced scattering of the mobile carriers in the potential well due to a decreasing influence of the Coulomb potential on the carriers with growing distance.
Hence, a further reduction of this type of scattering effect can be achieved by moving the dopants in the unstrained layer away from the interface to increase the distance between the ionized impurities and the carriers. The width of such an undoped spacer layer between the interface and the doped layer has to be considered carefully because although the scattering effect of the Coulomb potential diminishes with increasing width, so does the number of carriers being able to transfer into the potential well.

**Quantum Well**

Another possibility to form a two-dimensional electronic system in a Si/SiGe heterostructures is the design of quantum wells. A quantum well consists of a strained thin layer of that material, which has the energetically lower-lying band edge for the relevant carriers to be confined, sandwiched between relaxed layers of the other semiconductor material as shown in Figure 2.16. In both cases, the strained layer is doped according to the carrier type it is supposed to trap to ensure carrier confinement even at very low temperatures. As for the modulation-doped potential well, the energy spectrum of the confined carriers becomes quantized in growth direction if the well width is narrow enough to be comparable to the de Broglie wavelength of the confined carrier type. Again, these quantum effects have no influence on the free motion of the carriers in the plane of the well.

![Quantum Well Diagram](image)

*Figure 2.16 The two possible types of band alignment of a quantum well structure: type II for a strained Si layer to confine electrons (2DEG) and type I for a strained SiGe layer to confine holes (2DHG). The width $L$ of the quantum well has to be comparable to the de Broglie wavelength to allow for quantized states.*

It is the quantization of the energy spectrum in growth direction that restricts the movement of the mobile carriers along this axis and reduces their degree of freedom from three to two dimensions. In order to realize 2DCG systems by designing either potential or quantum well structures, all requirements necessary to trap mobile carriers inside the chosen type of well and have their dimensionality reduced have to be verified carefully. Since the well width of both types of structures has to be small enough to be comparable to the de Broglie wavelength to guarantee a two-
dimensional confinement of the carriers, it is important to calculate its value. The de Broglie wavelength is a quantity to describe the periodicity of the plane wave $e^{ikr}$ in the crystal by defining the wavelength of its period and is given by

$$\lambda = \frac{2\pi}{k} = \frac{h}{p} \approx \frac{h}{(2m^* E)^{1/2}}.$$  \hspace{1cm} (2.20)

Whereas the energy of electrons and holes increases with increasing temperature, the corresponding de Broglie wavelength decreases according to eqn. (2.20). Hence, the minimum value of the de Broglie wavelength and therefore the maximum value for the well width has to be calculated at room temperature where electrons and holes have an average thermal energy of $E \approx 26$ meV. For electrons, the de Broglie wavelength can be estimated to be $\lambda \approx 175$ Å, whereas for holes, its value ranges between 100 Å and 250 Å for corresponding effective masses between 0.5 $m_0$ and 0.1 $m_0$, respectively.

With regard to the design of a quantum well, the well width $L$ is equal to the thickness of the strained layer which in turn is only limited by growth conditions. As the different estimated values for the de Broglie wavelength each allow for a maximum layer thickness in the regime of metastable strain (above the equilibrium critical thickness but below the metastable critical thickness, see Figure 2.5), growth related restrictions should pose no problems in successfully fabricating 2DCG structures. The width of the potential well on the other hand cannot be determined by simply manipulating the width of the strained layer. Assuming a triangular-shaped potential well, the value of the well width $L$ can be calculated via

$$L = \left( \frac{1}{4} \frac{\hbar^2 \varepsilon_0}{m^* \varepsilon n_s} \right)^{1/3}$$ \hspace{1cm} (2.21)

with $m^*$ being the effective mass of the free carriers in the potential well, $\varepsilon_0$ the absolute permittivity of vacuum, $\varepsilon$ the dielectric constant (relative permittivity), $e$ the elementary charge and $n_s$ the two-dimensional sheet carrier density. Higher values for either the effective mass or the carrier density, which are the only parameters that can be influenced to vary the width of the potential well, result in a reduced well width. Because of the uncertainty of these two variables, it is not sensible to give an accurate estimation of the value for the potential well width, but assuming the values for both parameters fall within commonly used boundaries, the width of the potential well can be approximated to be between 10 Å and 30 Å, which is about an order of magnitude below the minimum value of the de Broglie wavelength. Hence, quantization of electron and hole energies can definitely be achieved in potential wells that are realized by modulation doping.

**Modulation-Doped Quantum Well**

The design of 2DCG structures is restricted by practical limitations. Whereas it is possible to grow single heterostructures in an unstrained system like GaAs/AlGaAs,
the critical thickness requires the strained layer in the Si/SiGe system to be sandwiched between at least two relaxed layers, thus forming a quantum well. Although a quantum well alone can be realized theoretically, doping the strained layer in order to confine carriers would introduce additional scattering centres, and in addition, the carriers would experience scattering from both interfaces because the strained layer would have to be very thin. Consequently, the best approach is a combination of both techniques where the modulation doping technique is applied to a quantum well structure which leads to a modulation-doped quantum well (MODQW) [4-6, 51, 52]. Figure 2.17 shows the potential variations of both types of MODQW structures where a strong shift of the lowest subband towards the upper interface of the well can be seen, caused by the asymmetric location of the doping region. This leads to a very narrow triangular potential shape next to that interface which efficiently confines the ground states almost independently of the well width.

![Diagram of SiGe relaxed and Si strained layers](image)

**Figure 2.17** Realization of 2DEG and 2DHG structures by applying the modulation doping technique on quantum wells. An undoped spacer layer is included in both structures to reduce Coulomb scattering at the ionized impurities in the doping layer.

The most basic design rule requires that under operational conditions all dopant atoms should be ionized, whereas all free carriers should be restricted to the channel region. This condition defines the integral doping concentration for a given layer sequence and a defined surface potential but still leaves ample design freedom for optimizing a MODQW structure for different purposes. High mobilities, for example, require thick spacer layers to reduce Coulomb scattering at the ionized impurities of the doping layer. Because of the linear potential drop across the undoped spacer layer, charge transfer from the supply layer into the channel becomes less and less efficient as the width of the spacer layer is increased. The carrier transfer also depends to some extent on the volume concentration in the doping layer which should be as high as possible for maximum efficiency. The doping level should be designed such that the lowest subband in the doping layer is located just below (above) the Fermi level in order to fulfill the low-temperature condition that free electrons (holes) should only exist in the channel. At room temperature, however, an increase of the density of states of the fully degenerate carriers by roughly an order
of magnitude may lead to a free carrier exchange between channel and doping region where a parasitic conducting channel of low mobility can develop which is limited by ionized impurity scattering. In Figure 2.18, the layer sequence of both types of MODQW structures is illustrated. The doping concentration is reduced towards the surface to avoid a lowering of the Schottky barrier, and a thin undoped Si cap layer is added to protect against in-depth oxidation and to allow for a well defined gate oxide (employed in enhancement transistors) or Schottky gate (used for depletion-type gate control). By reducing (increasing) the surface potential in n-type (p-type) enhancement transistors, there is the danger for the Si cap layer to operate as a parallel 2D channel of low mobility, which has to be avoided strictly.

Figure 2.18 Layer sequence for an n-type MODQW (left) and a p-type MODQW (right) structure.

The correlation between carrier density and mobility requires compromises which are ruled by the chosen application of the MODQW structure. While transport and magneto-transport experiments need mobilities as high as possible, device structures will rather be optimized with regard to channel conductivity, i.e. the product of carrier density and mobility or the saturation drift velocity. Certain modifications of the standard layer sequence discussed so far can be made that might improve the transport properties of the carriers in the channel. Significant examples are a reversal of the layer arrangement with the doping region on the substrate side and the channel on top or a symmetric configuration with spacer and doping layers on either side of the channel. An inverse layer sequence can be advantageous if the quality of the top interface is worse than that of the bottom one. This phenomenon occurs in many heterostructures and results from Ge segregation in the SiGe system. Since SiGe layers tend to a 3D growth, a Si to SiGe interface is usually more abrupt and laterally smoother than the inverse succession. The main disadvantage of a sequence reversal is the increased background doping level in the channel which is caused by dopant segregation or diffusion from the doping layer. Because of the increased distance between gate and doping layer, this type of heterostructure is not well suited for field-effect transistors unless a back gate is used. By way of implementing a symmetric layer geometry, the carrier density in the channel can be doubled, which is useful for device applications that require low channel resistance rather than extremely high mobilities. In addition, a symmetric wavefunction in an almost
rectangular potential well can be achieved, which minimizes the penetration of the wavefunction into the heterobarrier.

### 2.2.2 Density of States

As both electrical and optical properties of any system depend strongly upon the density of states (DOS) of the electronic bands, knowledge of the DOS is a fundamental requirement to determine the behaviour of electron or hole transport in an electronic system. By definition, the density of states denotes the number of available electronic states per unit volume per unit energy around an energy \( E \). Having assigned the expression \( N(E) \) to represent the energy-dependent DOS, the number of states in an energy interval \( \partial E \) around an energy \( E \) is given by \( N(E)\partial E \). It is reasonable to assume parabolic energy bands for a free electron and hole gas. Furthermore, the dimensionality of the system has to be known since the DOS is also dependent on the volume which in turn depends strongly on the degree of freedom of the electrons and holes. In a three-dimensional system, the DOS has a dependency on the square root of the energy

\[
N_{3D}(E)\partial E = \frac{g_s g_v}{4\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} \partial E \tag{2.22}
\]

where \( m^* \) denotes the effective mass of the corresponding carrier type and \( g_s \) and \( g_v \) are the factors for the spin and the valley degeneracy, respectively, which indicate the number of equivalent energy bands. The factor for the spin degeneracy \( g_s \) is always two due to the up and down spin of the fermions. For the electrons in a strained Si layer, the factor for the valley degeneracy \( g_v \) is two according to the fact that the lowest conduction band minimum is two-fold degenerate. Since there is no degeneracy of the highest valence band maximum, \( g_v \) equals one in the calculation for the hole DOS.

Reducing the dimensionality of a system alters the energy dependence of the density of available states profoundly. Removing one degree of freedom from a system results in a quantum well with a two-dimensional DOS that is independent of energy. The DOS of a 2D system is given by

\[
N_{2D}(E)\partial E = \frac{g_s g_v}{4\pi} \frac{2m^*}{\hbar^2} E \partial E \tag{2.23}
\]

Reducing the dimensionality even further to only one degree of freedom leads to a quantum wire with a one-dimensional DOS that is inversely proportional to the square root of the energy

\[
N_{1D}(E)\partial E = \frac{g_s g_v}{4\pi} \left( \frac{2m^*}{\hbar^2} \right)^{1/2} E^{-1/2} \partial E \tag{2.24}
\]
Finally, after having removed all three degrees of freedom, the zero-dimensional DOS of the resulting quantum box is determined simply by the $\delta$-function of the energy. An overview of the energy dependences of the DOS in the four electronic systems with different dimensionalities is given in Figure 2.19.

![Figure 2.19](image)

Figure 2.19 Density of states and their energy dependence in 3D, 2D, 1D and 0D. The DOS of the 3D system is indicated in the DOS of the 2D system to illustrate their relation.

Even though the density in two dimensions is constant, each quantized level contributes the same amount of available states to the system, and hence, the DOS for a quantum well is a step function with steps occurring at the energy of each quantized level. Similarly, the DOS for a quantum wire is inversely proportional to the square root of the energy between the energies of successive quantized levels but uniform directly at the energy of each quantized level. For quantum boxes, the DOS is a $\delta$-function with a constant magnitude at the energy of each quantized level. These variations in the energy dependence of the DOS are responsible for substantial changes in the properties of electronic systems and thus the driving force in lowering the dimensionality of semiconductor structures.

To emphasize the importance of the energy independence and thus the step function of the density in a 2D system, the characteristic density curves of both a 3D and a 2D system are shown in Figure 2.20 quantitatively. Evidently, the energies of the quantized levels are a quadratic integral multiple of the energy at ground level, corresponding to the density which is an integral multiple of $m^*/\pi\hbar^2$. However, the presence of disorder and imperfections in a real 2D system gives rise to states that can trap electrons which are referred to as localized electrons. The division into
localized and extended states results in a band tail in the step function rather than the sharp transition obtained for the ideal 2D system.

![Diagram showing density of states in two and three dimensions](image)

Figure 2.20 Density of states in two and three dimensions. Comparison of the energy dependent DOS of an ideal 3D system and the energy independent DOS of an ideal 2D system (top). In real 2D systems, the effect of disorder on the density creates a band tail with localized states (bottom).

Electrons in localized states are bound to a finite region in the crystal and hence do not participate in the carrier transport in general, whereas electrons in extended states can move freely throughout the crystal, thus contributing to conduction. The energy which separates the regions of localized and extended states is called the mobility edge $E_m$. Electrons or holes in localized states can only move by being thermally activated to energy states above the mobility edge or by phonon-assisted hopping between localized states at low temperatures. Hence, hopping conduction is essential for transport in localized states.

### 2.3 Device Applications

Apart from providing interesting and important information on the basic physics of quantum mechanics in general and of semiconductors in particular, the main use of heterostructures lies in their device applications. Basically, all devices can be roughly categorized into two separate groups according to their transport behaviour, i.e. electronic devices based on parallel and on perpendicular transport. Devices of the first group are based on parallel transport and comprise all variations of field effect transistors, whereas bipolar transistors, optoelectronic and special microwave devices are all governed by perpendicular transport and thus belong to the second
group. All types of devices and the technology in which they can be realized are listed in Table 2.2 to give an overview of the abilities of heterostructure systems.

<table>
<thead>
<tr>
<th>device types</th>
<th>technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>field effect transistors</td>
<td>Si, SiGe, III-V</td>
</tr>
<tr>
<td>bipolar transistors</td>
<td>Si, SiGe, III-V</td>
</tr>
<tr>
<td>optoelectronic devices:</td>
<td></td>
</tr>
<tr>
<td>optical detectors and sensors</td>
<td>SiGe, III-V</td>
</tr>
<tr>
<td>lasers, LED's and modulators</td>
<td>III-V</td>
</tr>
<tr>
<td>special microwave devices</td>
<td>Si, SiGe, III-V</td>
</tr>
</tbody>
</table>

Table 2.2  Overview of the most important device types and the technology they can be realized in.

Although the fabrication of lasers and LED’s as well as modulators is still restricted to III/IV heterosystems, it is a remarkable and important feature of the SiGe system to provide a means to realize photo detectors in standard Si technology. Out of the above mentioned devices, the optoelectronic devices are the only ones which cannot be accomplished in a pure Si system. Due to their unique property of a negative differential resistance (NDR), special microwave devices based on the tunneling effect (tunnel diode), impact ionization (IMPATT diode) or the transferred-electron effect (TED) are of great importance for specialized systems, mainly in the very high frequency regime [35]. Again, the performance of III-IV systems in this area is by far superior to that of Si and SiGe systems. However, the most commonly and widely employed structure in VLSI (very large scale integration) designs is the field effect transistor (FET) which is largely used for logic circuits. Its counterpart in the analogue world is the bipolar transistor (BJT), but since the market for analogue circuits is relatively small compared to the dominating position of the FET driven logic circuits, the BJT as a device in the production line is slightly less important although it has found its way into the logic design due to its advantageous speed. Because of their importance in the semiconductor industry, the next two sections deal with the field effect and the bipolar transistor and their realization in the SiGe heterosystem.

### 2.3.1 MODULATION-DOPED FIELD EFFECT TRANSISTORS

In general, a field effect transistor is a device in which a current that is running across two terminals (the source and the drain) can be controlled by applying a voltage to a third terminal (the gate) which is electronically insulated from the conducting channel. Hence, FETs are characterized by a high input impedance. There are a
variety of different FET structures, distinguished essentially by the way the gate is isolated from the channel:

**JFET:** Junction FET, gate is isolated via a reverse-biased \( p-n \) junction

**MESFET:** Metal-Semiconductor FET, gate is isolation via a Schottky barrier

**MISFET:** Metal-Insulator FET, gate is isolated via an insulating layer

**MOSFET:** Metal-Oxide FET, subtype of MISFET implemented on Si in which the gate is separated from the active channel by the material’s native oxide.

By far the most important representative of the family of field effect transistors is the Si-based MOSFET which is the backbone of the digital microelectronic systems. A much less significant role in VLSI circuit designs is granted to the other members of the family, which in particular are the JFET and the MESFET. Both types, JFET and MESFET, behave according to the same principle of majority-carrier transport and can basically be seen as a voltage controlled resistor with the JFET having a \( p-n \) junction as the gate and the MESFET a Schottky contact. In a MOSFET structure, minority carriers are transported in a 2D inversion layer whose conductance can be modified by varying the gate voltage. A thin layer of oxide is grown on top of the Si to insulate the metal gate from the inversion layer which builds up directly underneath the oxide at the surface of the Si. Ideally, the MOSEFET is a transadmittance amplifier with an infinite input resistance and a current generator at the output.

**MOSFET**

In \( n \)-channel MOSFETs, the substrate is lightly \( p' \)-doped, whereas the source and drain regions are heavily \( n'' \)-doped. The opposite device layout applies for the structure of a \( p \)-channel MOSFET. Two different operational modes can be defined for \( n \) and \( p \)-channel MOSFETs, depending on the way the channel forms between the source and drain regions. If at zero gate bias the channel conductance is very low in an \( n \)-channel MOSFET, a positive voltage has to be applied to the gate to form the electron channel, and hence, it is called the normally-off or enhancement \( n \)-channel MOSFET. In case the electron channel already exists at zero bias, a negative gate voltage will deplete the electrons in the channel to reduce its conductance. This type is called the normally-on or depletion \( n \)-channel MOSFET. Similarly, there exists the concept of the \( p \)-channel enhancement and depletion MOSFET, resulting in a total number of four different MOSFET structures. The basic device parameters to describe a MOSFET are the channel length, the channel width, the insulator (oxide) thickness, the junction depth and the substrate doping. Its performance, on the other hand, is characterized by the values of the transconductance, the maximum unity current gain cut-off frequency, the maximum frequency of oscillation and the gate voltage swing (logic swing) which defines the abruptness between the on and off state when used as a switch in digital logic. Other parameters like leakage current, subthreshold characteristics or “off” current are also very important to be able to fully characterize a MOSFET device.

In principle, the gate in a MOSFET can be pictured as a parallel capacitor which can either accumulate or deplete carriers in the underlying semiconductor by means of
varying the voltage and consequently raise or lower the conductivity. The current from source to drain through the channel is then given by

\[ I_{DS} = C_G \frac{V_D}{L_G} (V_G - V_T) \]  \hspace{1cm} (2.25)

where \( v_d \) is the transport or drift velocity of the carriers, \( C_G \) is the gate capacitance, \( L_G \) the gate length, \( V_G \) the gate voltage and \( V_T \) the threshold voltage. At low electric fields, the drift velocity of a semiconductor is proportional to the applied electric field, and the proportionality constant is the low field mobility \( \mu \) that is independent of the electric field

\[ v_d = \mu E. \]  \hspace{1cm} (2.26)

However, when the fields are sufficiently large, the mobility becomes non-linear and above a critical field \( E_c \), saturation of the drift velocity occurs. Because the carriers acquire energy from the electric field, they have to lose it to the lattice by net emission of phonons. Above \( E_c \), any attempt to increase the velocity further by raising the electric field is countered by an increased phonon emission (Si), intervalley scattering (n-type GaAs) or impact ionization (p-type Si/GaAs). As the field increases, the average energy of the carriers also increases and they acquire an effective temperature \( T_e \) which is higher than the lattice temperature \( T \). At fields above \( E_c \), the drift velocity has to be corrected by a term that balances the rate at which energy is transferred from the field to the carriers with an equal rate of loss of energy to the lattice. Table 2.3 lists the saturation drift velocities for Si, Ge and GaAs as well as the corresponding critical electric fields.

<table>
<thead>
<tr>
<th></th>
<th>saturation drift velocity ( v_d ) (cm/s)</th>
<th>critical electric field ( E_c ) (V/cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>electrons</td>
<td>holes</td>
</tr>
<tr>
<td>Si</td>
<td>(~10^7)</td>
<td>(~3 \times 10^4)</td>
</tr>
<tr>
<td>Ge</td>
<td>(~6 \times 10^8)</td>
<td>(~2 \times 10^5)</td>
</tr>
<tr>
<td>GaAs</td>
<td>(10^7 \rightarrow ~6 \times 10^6)</td>
<td>(~6 \times 10^6)</td>
</tr>
</tbody>
</table>

Table 2.3 Saturation drift velocities and corresponding critical electrical fields for Si, Ge and GaAs. Note the wide range of differential negative mobility for electrons in GaAs.

Derived from the dc characteristic of a transistor, the transconductance \( g_m \) is an important device parameter because it defines the intrinsic current generator and it stands in direct relation to the high-frequency behaviour of the transistor. The transconductance is defined per unit gate width and is given by

\[ g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{E=\text{const}} = C_G \frac{v_d}{L_G}. \]  \hspace{1cm} (2.27)
Under high-frequency operation, two factors limit the frequency response of a FET, the transit time and the RC time constant. The transit time $\tau = L/v_d$ is derived from the finite time required for carriers to travel from source to drain and is normally small compared to the RC time constant resulting from the input gate capacitance and the transconductance. As an important figure of merit, the maximum unity current gain cut-off or transit frequency $f_T$ is defined as the frequency at which the current through the gate/source capacitance $C_G$ is equal to the current generator $g_m V_G$ in the intrinsic field effect transistor according to

$$f_T = \frac{1}{2\pi \tau} = \frac{g_m}{2\pi C_G} = \frac{v_d}{2\pi L C_G}.$$  (2.28)

In bipolar transistors, the cut-off frequency has the same dependence on the transit time but is defined as the frequency at which the common emitter short-circuit gain $h_{fe}(=\partial I_C/\partial I_E)$ decreases to unity and the transit time is given by the emitter-to-collector delay time.

In this approximation, both the transconductance and the cut-off frequency are directly proportional to the drift velocity and inversely proportional to the gate length. Furthermore, the cut-off frequency is independent of the gate capacitance. Hence, either the gate length has to be reduced or the drift velocity has to be increased in order to raise $g_m$ and $f_T$ and thus enhance the performance of the transistor. However, as the gate length is pushed to smaller dimensions, velocity saturation in the gate region gains importance, leading to the replacement of the low field mobility by a more realistic expression that takes into account the actual velocity profile in the gated area of the device. Trying to increase $v_d$ by raising the electric field has only a limited effect since $v_d$ can only be increased until its value reaches $v_s$. Consequently, the maximum transconductance and cut-off frequency and thus the maximum performance of a MOSFET of fixed dimensions is given by the special case of eqns (2.27) and (2.28) where $v_d$ is replaced by $v_s$.

**MODFET**

Another way of improving the drift velocity is by increasing the mobility of the electrons or holes in the channel. A high carrier mobility combined with a high carrier density that can be achieved in two-dimensional systems suggest that a significant improvement of the performance can be achieved in the application of a modulation-doped FET (MODFET). In a MODFET, the carriers are separated from their parent impurities, they move in a buried QW instead of an inversion layer underneath the surface, and they have a reduced effective mass. Basically, these are the main three factors contributing to a substantially increased carrier mobility which is essential for an enhanced performance, especially in the high-frequency regime. In general, SiGe MODFET devices offer the possibilities of increased speed, reduced high-and low-frequency noise, low operating voltage, low power consumption and as such have potential for both analogue and digital applications. Designing power FETs, on the other hand, requires the transistor to conduct high levels of current which is given by the product of carrier density and mobility. Higher currents can
easily be achieved in a MODFET by maximizing the carrier density without seriously compromising the mobility. Hence, the output of such a modified MODFET can deliver large amounts of power necessary for its application as for instance a power amplifier.

The introduction of advanced growth technologies made the development of MODFET devices available for the design of FETs with a greatly improved performance concerning the high-frequency operation and the power handling capability. Other acronyms that are being used for this type of device include HFET (Heterostructure FET) and HEMT (High Electron Mobility Transistor) which is often used in connection with GaAs-based FETs. Figure 2.21 shows a typical conduction and valence band edge profile of an n-type and p-type MODFET structure. An important factor is the doping concentration because if the doping level in the supply layer is too high, the conduction band edge of the n-MODFET falls below the Fermi level resulting in the confinement of electrons and thus in a low-mobility parallel conduction channel. Obviously, the same considerations have to be applied to the valence band edge in a p-MODFET.

![Figure 2.21](image)

It is worth pointing out that in general, the investigation of both types of MODFET devices has been concentrating on the most important parameters such as the transconductance, the transit frequency and the maximum frequency of oscillation since these are the main figures of merit of an FET. To realize the different modes of an FET (depletion/enhancement, n-type/p-type), many different structural designs have been studied which were also used to look at certain aspects of the MODFET performance. In depletion MODFETs, the channel can be buried and the gate can be grown normally on top of the device as a Schottky contact or with an oxide in between. For an enhancement MODFET, either the gate has to be recessed deep into the doping layer, thus requiring a Schottky contact, or alternatively, a surface channel device has to be used which allows the gate to be grown on top of the MODFET structure. The layer design for the fabrication of both n- and p-type devices is somewhat diverse and there are a number of device structures that have a variety of very different concepts. Looking ultimately at the combination of both types
within a single device as is necessary for complementary MOSFET (CMOS) circuit design, the structural design of such a device becomes very complex.

**N-TYPE MODFET**

Early work on n-MODFET structures used a uniform-composition buffer where the Ge content was kept constant. Due to the resulting high density of threading dislocations, their performance has been comparatively poor. The first n-channel MODFET had a transconductance of only 40 mS/mm and a transit frequency of 2.2GHz. More recent work employed compositionally graded buffers where the Ge content is graded either linearly or in steps to minimize dislocation faults. By additionally having optimized the layer design for high transconductance values by minimizing the distance between the 2DEG and the gate, the performance of these structures showed a dramatic improvement of the transconductance over that of the uniform buffers by roughly a factor of ten. In general, device fabrication steps require low-temperature processing to avoid degradation of the abruptness of the heterointerfaces. Values for the transconductance of various n-MODFET structures are listed in Table 2.4 where the values for the graded buffers are comparable with those reported for HEMT structures fabricated in the GaAs system.

<table>
<thead>
<tr>
<th>type of SiGe buffer</th>
<th>gate length (µm)</th>
<th>transconductance $g_m$ (mS/mm)</th>
<th>mobility $\mu$ (cm²/Vs)</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>300K</td>
<td>77K</td>
<td>300K</td>
</tr>
<tr>
<td>uniform composition</td>
<td>1.6</td>
<td>40 (e)</td>
<td>—</td>
<td>1550</td>
</tr>
<tr>
<td>uniform composition</td>
<td>1.4</td>
<td>80 (e)</td>
<td>88 (i)</td>
<td>1090</td>
</tr>
<tr>
<td>compositionally graded</td>
<td>1.4</td>
<td>340 (e)</td>
<td>380 (i)</td>
<td>670 (e)</td>
</tr>
<tr>
<td>step graded</td>
<td>0.25</td>
<td>330 (e)</td>
<td>600 (e)</td>
<td>1500</td>
</tr>
<tr>
<td>step graded</td>
<td>0.5</td>
<td>390 (e)</td>
<td>520 (e)</td>
<td>2600</td>
</tr>
</tbody>
</table>

$^1$ low field mobility, (e): extrinsic, (i): intrinsic

Table 2.4 Performance of selected n-MODFET devices [59].

In addition, a very low leakage current at high temperature was measured in the device structure of [58]. Measurements of the high-frequency behaviour has only been carried out on a few n-MODFET structures but reports of a transit frequency of 40 GHz and a maximum frequency of oscillation $f_{max}$ of 56 GHz in a Schottky gated MODFET with a gate length of 0.4 µm [60] is comparable to the performance of GaAs/AlGaAs HEMTs. An improvement on these figures was realized for a device
with a 0.18 \( \mu m \) gate length, reaching values for \( f_T \) of 46 GHz and for \( f_{max} \) of 81 GHz [61-63]. However, the best high-frequency performance to date has been achieved in a more recent n-MODFET device with 0.25 \( \mu m \) gate length, displaying an \( f_T \) of 42 GHz and an astonishing value for \( f_{max} \) of 120 GHz [64].

\textbf{P-TYPE MODFET}

Unlike in the design of n-MODFETs in which the QW always consists of a strained Si channel, there is a wide variety of compositions to choose from for the strained SiGe channel in a \( p \)-MODFET. Basically, three different categories can be defined depending on the Ge concentration in the strained layer and in the buffer layer. The first group comprises all pseudomorphic \( p \)-MODFETs that have a Si buffer layer, and consequently, the Ge content in the QW cannot exceed 40 - 50\% without introducing dislocation faults in the channel. All structures that have a SiGe channel with Ge concentrations above 50\% can be put into a second group. Inevitably, these devices must have a relaxed SiGe buffer layer that can vary in composition but must guarantee a fully strained SiGe layer. The third group consists of all structures with a pure Ge channel and which therefore have to employ SiGe buffer layers with a correspondingly high Ge content.

The earliest \( p \)-channel MODFETs were fabricated at the same time as their first \( n \)-type counterpart [7] and were realized in depletion as well as in enhancement mode [8]. The SiGe channel had a Ge concentration of 20\% and transconductance values of 2.5 and 3.2 mS/mm were measured for the enhancement and depletion mode, respectively. A more promising approach to achieve better performance is given by the use of \( p \)-MODFETs on strain-relaxed buffer layers due to a drastic improvement of the hole mobilities in a SiGe layer with Ge contents well beyond 50\%. Critical thickness limitations make it necessary to use a virtual substrate which rather serves as a means of strain limitation than for band structure engineering. Initially, pure Ge channels on a virtual substrate with an effective Ge concentration of 70 - 80\% were considered as the best possible choice in terms of high-frequency performance because of the small hole mass and the absence of alloy scattering. Although this approach seemed to yield devices with a good performance, it is not well suited for a combination with \( n \)-MODFETs employing a pure Si channel in CMOS design. Furthermore, relaxed buffer layers with compositions beyond 50\% are much more complicated to grow and cause morphological problems.

Having found that neither pseudomorphic nor pure Ge channel MODFETs are suitable for high-performance device fabrication, most research now concentrates on structures employing a SiGe alloy channel with a Ge concentration significantly larger than 50\% where hole mobilities are drastically increased. Although a virtual substrate is required, growth and morphology concerns are relaxed because its effective composition can be kept below 50\%. An overview of the transconductance \( g_m \), the maximum unity current gain cut-off frequency \( f_t \), and the maximum frequency of oscillation \( f_{max} \) is given in Table 2.5 for a selection of \( p \)-MODFET devices.
The Si/SiGe Heterosystem Device Applications

Table 2.5 Room temperature performance of selected p-MODFET devices.

<table>
<thead>
<tr>
<th>Channel composition $x$ (Si$_{1-x}$Ge$_x$)</th>
<th>Gate length (μm)</th>
<th>$g_m$ (mS/mm)</th>
<th>Transit frequency $f_T$ (GHz)</th>
<th>Max. frequency of oscillation $f_{max}$ (GHz)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7 - 0.55</td>
<td>0.7</td>
<td>105 (e)</td>
<td>9.5</td>
<td>–</td>
<td>[65]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>230 (e)</td>
<td>24</td>
<td>37</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>280 (i)</td>
<td>30</td>
<td>45</td>
<td>[60]</td>
</tr>
<tr>
<td>0.8 - 0.7</td>
<td>0.1</td>
<td>90 (e)</td>
<td>38</td>
<td>64</td>
<td>[67]</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>258 (e)</td>
<td>70</td>
<td>55</td>
<td>[68]</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>160 (e)</td>
<td>32</td>
<td>85</td>
<td>[69]</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>190 (e)</td>
<td>59</td>
<td>135</td>
<td>[70]</td>
</tr>
</tbody>
</table>

(e): extrinsic, (i): intrinsic

With respect to their combined use in CMOS applications, an approach has been made to confine both the electrons and holes in strained Si channels with enhanced carrier mobilities in each case. Hence, a major focus has been to achieve enhanced performance in p-MODFETs employing strained Si channels [59, 71, 72]. Although early measurements suggest an improved performance compared to their p-MOSFET counterparts, it is still well below that of conventional p-MODFETs with strained SiGe channels.

**COMPLEMENTARY MODFET**

The ultimate aim in terms of logic circuits is the design of complementary MODFETs in analogy to the CMOS architecture in Si which is the most widespread device type in digital microelectronics. Due to higher carrier mobilities, higher saturation drift velocities and higher sheet densities, CMODFETs are expected to display higher transconductances, higher speed, lower gate delay, lower noise and lower power consumption. Additionally, higher packing densities can be achieved by designing equally sized $p$- and $n$-MODFETs, which is possible because of the enhanced performance of $p$-MODFETs. A variety of different layer designs have been considered which basically consist of stacking the two devices on top of each other in a single epitaxial run. Two different approaches have been suggested, either by incorporating the complete MODFET layer sequence including the doping layers or by having stacked MOSFET arrangements (HCMOS) which exploit the strained channels to separate the mobile carriers from the interface with the SiO$_2$ but
otherwise rely on carrier injection from the source/drain contacts rather than on modulation doping (for a detailed review see [28, 46, 59, 63, 73]).

For devices that are made up of two complete MODFET layers, access to the lower-lying transistor requires local removal of the upper device level. Initial investigations into the characteristics of structures in which the n-MODFET was grown on top of the p-MODFET showed a much inferior performance of the p-device compared to the n-device due to the downward-penetrating dislocations from the adjacent strain relaxed SiGe buffer layer of the n-MODFET. A model of the reverse sequence, i.e. the p-MODFET being placed on top of the n-MODFET, with both devices utilizing the same virtual substrate showed some promising results, but devices in accordance with the model have not been fabricated yet [28].

Subsequent efforts have therefore concentrated on the design of HCMOS structures, which have a CMOS compatible layer sequence, and both types of devices utilize a common strain-adjusted buffer layer [74]. The channels for electrons and holes are vertically stacked next to each other with the carriers being injected from the source/drain contacts, i.e. both devices operate in enhancement mode. Because both devices can be realized in the same layer sequence by choosing the respective doping type for the source and drain contacts and by proper adjustment of the threshold voltage, no local removal of layers is required. The planar design can be seen in Figure 2.22 in which a schematic cross section of such a device is shown together with its layer sequence.

Figure 2.22 Cross section of a complementary HMOSFET (top). Layer sequence with a Si channel for the n-type and a SiGe channel for the p-type device (bottom). Conduction and valence band edge profiles for a positive and a negative gate bias just above $V_T$ are also shown [75].
Shallow trench isolation (STI) is used to avoid prolonged high-temperature processing which could give rise to strain relaxation and diffusion. Problems with this design arise from the restrictions imposed by the critical thickness, limiting the channel thickness of electron and hole channel according to the composition in the virtual substrate. Measurements of HCMOS devices found a much improved power delay product for the loaded and unloaded case and reduced minimum delay times. While in standard CMOS gate lengths below 0.2 \( \mu \text{m} \) are required to achieve transconductances of 400 mS/mm at room temperature, the same values for \( g_m \) can be found for gate lengths of 1.2 - 1.4 \( \mu \text{m} \) in HCMOS technology [59], and in the sub 0.2 \( \mu \text{m} \) domain, \( g_m \) is expected to yield more than 800 mS/mm [74].

Ideally, the best combination for an HCMOS design would employ an n-FET with a strained Si channel and a p-FET with a strained Ge channel in which case the complete mismatch of 4.2% between Si and Ge would have to be bridged by the relaxed buffer. Symmetric strain distribution could be achieved by employing a Si\(_{0.8}\)Ge\(_{0.2}\) substrate, but the equilibrium critical thickness would be a mere 30 Å. Such thin channels are conceivable but not ideal since the carrier wavefunctions probe both interfaces leading to enhanced interface scattering and thus a reduction of the mobilities. Another approach that has to be explored further is the confinement of both electron and holes in a strained Si channel which would make the fabrication process much simpler.

### 2.3.2 HETEROJUNCTION BIPOLAR TRANSISTOR

Both \( npn \) and \( pnp \) heterojunction bipolar transistors (HBT) have been fabricated successfully, the most recent ones exhibiting transit frequencies around 120 GHz and maximum oscillation frequencies of 120 GHz, as well [76-79]. However, nearly all of the technological and scientific interest is currently directed at the \( npn \) devices. How the use of heterostructures can improve the device performance of the bipolar transistor is best illustrated by looking at the basic device principle which is explained in short.

In all \( npn \) bipolar transistors, the collector current relies on the injection of electrons from the emitter into the base. In the base, the electrons are minority carriers and move by a combination of diffusion and drift to the reverse-biased collector-base junction where they are swept into the collector region. Current amplification in a BJT is mainly determined by the ratio between the density of electrons injected from the emitter into the base and the density of holes re-injected from the base into the emitter:

\[
\beta = \frac{I_C}{I_B} = \frac{(N_n w_n D_n)}{(P_p w_p D_p)}
\]

(2.29)

where \( N_n \) and \( P_p \) denote the emitter and base doping concentrations, \( w_n \) and \( w_p \) are the respective thicknesses, and \( D_n \) and \( D_p \) are the diffusion constants of electrons in the base and of holes in the emitter, respectively. Since the base current is inversely proportional to the current gain, it has to be kept low. The two main sources for the
base current are the reverse injection of holes from the base into the emitter and the recombination of excess electrons in the base with holes. Due to high minority carrier lifetimes and relatively short base transit times, recombination in the base can be neglected in modern Si-base devices.

The basic idea of the HBT is the introduction of a heterojunction at the emitter/base interface such that the bandgap of the base layer is smaller than that of the emitter. This leads to an enhancement of $\beta$ by an exponential factor

$$\beta_{\text{HBT}} = \beta_{\text{BJT}} \cdot e^{\Delta E_g / kT}$$

which depends only on the bandgap difference $\Delta E_g$ and the thermal energy $kT$. The npn HBT, having a Si emitter and collector and a SiGe base as shown in Figure 2.23, has two principal advantages over its Si homojunction counterpart. For a given base-emitter bias, the conduction band offset between base and emitter is much smaller than the valence band offset. This imbalance of the band offsets reduces the reverse injection of holes into the emitter, which increases the emitter efficiency and reduces the base current, thus leading to an enhanced current gain.

![Figure 2.23 Schematic band diagram of an npn-HBT. The potential wall between the base and emitter is much smaller for the electrons than for the holes. Additionally, acceleration of the electrons in the base can be achieved by a built-in drift field.](image)

This improvement in itself is of marginal benefit for the performance of most circuits since an increase of $\beta$ is hardly needed in most applications, but the enhanced gain can be traded for a higher doping level in the base. With a higher doping, the base can be made much narrower, further reducing the transit time of the electrons and increasing the transit frequency. At the same time, the low sheet resistance has the beneficial effect of reduced noise in microwave applications. The other advantage of HBTs is the possibility of grading the Ge content across the base, thereby providing a built-in field to enhance the speed at which the minority carriers cross the base, which improves the frequency response of the device even further.


2.4 GROWTH TECHNIQUES

The epitaxial layer growth of SiGe heterostructures is a very demanding process due to the need for abrupt interface layers and a good control of the matrix and the dopant concentrations to achieve heterostructures of high quality. A potential growth system has to provide two closely controlled sources of Si and Ge to ensure integrity of the alloy composition and guarantee abrupt interfaces, which in turn requires the possibility of monolayer growth. Good control of the dopant sources is vital if exact doping profiles are to be obtained, and monitoring of the layer growth is essential.

There are basically two different growth techniques which meet the requirements for the fabrication of high-quality heterostructures. In the next two sections, their mode of operation is explained and their most important advantages and disadvantages are emphasized. Subsequently, an overview of the most commonly used characterization tools to determine the physical properties of the grown structures is presented, each with a short description of the specific kind of information they provide. At the end of the chapter, a discussion about the compatibility of heterostructures and their growth techniques with existing standard Si VLSI technology is presented.

2.4.1 MOLECULAR BEAM EPITAXY

Solid source molecular beam epitaxy (SSMBE), which involves electron beam evaporation of solid sources of the matrix and dopant materials, is the epitaxial growth technique that has been at the forefront of SiGe heterostructure research since the very early days, pioneering virtually all the critical developments. Its ultra-high vacuum (UHV) ambient is the basis of providing high-quality heterostructures. Si and Ge fluxes are generated by electron beam evaporation of solid charges of the respective elements which are directed towards the heated Si substrate. Rotation of the substrate ensures excellent lateral uniformity in deposition across the entire wafer. However, spitting defects that stem from the electron-gun evaporator of the Si source used to be a common occurrence. These characteristic surface defects cause stacking faults even if the Si molecular beam is not contaminated. In recent growth machines, this effect has been reduced significantly. Hence, doping has proven to be the single most difficult problem in SSMBE due to the fact that the potential dopants have vapour pressures which are either too high (high background doping levels) or too low (contamination by impurities effusing out of the steel walls), or they have unfavourable tendencies towards surface segregation. The main p-type dopant is elemental boron which is evaporated from a pyrolytic graphite crucible due to its extremely high melting point of 2300°C. Profile control is limited by surface segregation effects which are evident at higher doping levels (> $10^{19}$cm$^{-3}$) and higher growth temperatures (> 650°C). Antimony and phosphorous are the principal n-type dopants with Sb having the tendency to segregate to the growing surface, but
applying a bias voltage to the Si wafer enhances the incorporation of Sb considerably. Because of the high vapour pressure of P, either heavily P-doped Si or GaP is used as the dopant source, but strong memory doping effects restrict the use of phosphorous as a dopant.

The most important benefit of SSMBE is its flexibility which allows independent control of growth parameters, like Si, Ge and doping fluxes, over a wide range of temperatures (room temperature to 1000°C) and the ability to control the individual layer thickness down to one monolayer. Typical growth rates are 1 - 3 Å/s, which is suitable for growing the active region of the structures (<100 nm), but relaxed buffer layers (up to 5 µm) require long growth periods. A negative feature is the necessity to open the UHV growth chamber each time a matrix or dopant material has to be replaced, which interrupts and thus lowers the throughput.

An alternative growth process is the gas source MBE (GSMBE) where the matrix and the dopant materials are supplied by gas source reservoirs external to the growth chamber which can therefore be operated at room temperature. Typically, the gas sources are disilane (Si₂H₆) and germane (GeH₄) for the Si and SiGe layers and arsine (AsH₃) or phosphine (PH₃) for n-type and B₂F₆ for p-type doping. Their molecules are directed at the heated substrate held in a UVH environment where they thermally decompose to produce layer growth.

The fact that the sources do not have to be heated internally and can be recharged quite easily without opening the growth chamber is the most obvious advantage over SSMBE. Since the sources can be operated at room temperature, the incorporation of background impurities into the layers is much lower. The formation of atomic hydrogen at the surface of the wafer as a reaction product suppresses the segregation of Ge, thus enabling more abrupt interfaces. The growth rate of Si and SiGe is strongly dependent on the substrate temperature and can be divided into two distinct regimes. Above the transition temperature, the growth is controlled by the flux (arrival rate) of the hydrides, below the transition temperature, the growth is governed by the desorption of the newly formed hydrogen and independent of the flux. Growth rates of up to 1 Å/s can be achieved at low temperatures. Other advantages are a lower substrate temperature, higher growth selectivity, the availability of in-situ monitoring techniques and the possibility of growing structures with uniform thickness and composition without sample rotation.

2.4.2 CHEMICAL VAPOUR DEPOSITION

Gas sources are also used for the chemical vapour deposition (CVD) technique where decomposition of silicon tetrachloride (SiCl₄), silane (SiH₄) or germane (GeH₄) as the matrix molecules takes place on the surface of the heated substrate. B₃H₆ and PH₃ are used as the p- and n-type dopants, respectively. The rate of decomposition and doping is governed by the substrate temperature and generally increases with increasing temperature. As for GSMBE, there is a complex interdependence of matrix and doping levels.
The adoption of an UHV environment enables the fabrication of structures of higher quality due to reduced gaseous impurity uptake at lower temperatures. Ramping the wafer temperature rapidly up and down allows the growth of very thin layers (down to 5 nm) in the rapid thermal processing CVD (RTPCVD). Even thinner high-quality layers (down to 1.5 nm) can be achieved with UHVCVD where the system is not only constructed according to UHV standards but also operates at low pressures. Selective epitaxy, i.e. restricted deposition to selected areas of the wafer, is possible. A big disadvantage is the memory effect, which has also been observed for phosphine in GSMBE, making it difficult to maintain well controlled doping levels at low concentrations after high levels of doping have taken place or when the dopant material is changed.

2.4.3 STRUCTURAL CHARACTERIZATION

As a powerful tool to characterize the periodicity of a given crystal, X-ray diffraction measurements (XRD) provide useful information about the crystal structure of semiconductors and its quality. Since wave diffraction can be used at fairly high temperatures, it is possible to monitor the quality of a crystal while it is being grown by employing a technique that uses the diffraction of electrons. This particular type of wave diffraction is known as reflection high-energy electron diffraction (RHEED) which is one of the most widely used *in-situ* techniques to study surface quality during growth of epitaxial layers. It has become a standard feature in most MBE growth systems and is used for studying crystal orientation, growth rate calibration, growth quality check and surface lattice parameters such as the surface reconstruction. RHEED involves a high energy electron beam (5 - 50 keV) which is directed at the surface of the substrate at a grazing angle of incidence of 1° to 3°. Due to a penetration depth of the electron waves of only ~0.1 Å, the resulting pattern is formed by the elastic scattering of the electron waves of the incident beam from the surface structure of the crystal. In high-quality growth, the deposition of the monolayers can be counted directly by looking at the pulsation in the RHEED pattern.

Structural characterization and determination of film quality of strained layers after growth are usually carried out by transmission electron microscopy (TEM), both plan view and cross section. With it, defects and dislocations can be determined within the layers and the thickness of the layers can be measured. Energy-dispersive spectroscopy with a scanning TEM is used to measure the film composition. Rutherford backscattering spectroscopy (RBS) yields similar information, and in conjunction with high-resolution XRD, it can be used to determine the strain. An alternative technique that measures the strain in the film directly is Raman spectroscopy. Sputter depth profiling by secondary ion mass spectroscopy (SIMS) is used to study the chemical composition, such as the Ge concentration or the doping profile. However, this method has a limited depth resolution of about 3 nm and cannot give a reasonable profile for ultrathin films.
2.4.4 COMPATIBILITY TO STANDARD VLSI TECHNOLOGY

Although silicon epitaxy is a relatively expensive process, it has established a respectable position in the production of bipolar and submicron CMOS ICs. Throughout the industry, the CVD technique is the standard process used for the epitaxial layer growth of Si. The reactors of the CVD system are capable of handling a large number of wafers simultaneously, whereas an MBE system can only handle one wafer at a time. Hence, MBE seems to be ideally suited to research laboratories for developing new device designs and where there is no demand for high throughputs. However, due to improved deposition uniformity on larger wafers, single wafer epitaxy processing has become more relevant in recent years.

This leads to a direct competitive comparison between CVD and MBE where the key attribute of CVD and GSMBE is the selective growth capability and that of SSMBE the flexibility in matrix and dopant profile control. All growth techniques have difficulties in producing high-quality structures with Ge concentrations bigger than 30% (when the Ge tends to segregate) producing smeared interfaces, and surface rippling occurs due to minimization of the strain energy in the quantum well layer.

Another important factor in the integrated production of Si and SiGe is the incorporation of strained layers which have to survive intense thermal processing required in Si IC manufacturing. So far, surprisingly good stability of the strained layers under restricted thermal processing has been demonstrated, but long-term reliability still remains to be investigated. To give an example, for a 2DEG structure using Sb doping spikes, a maximum thermal budget of 800°C for 60 minutes has been determined which is limited by the bulk diffusion of the Sb doping spike. These values have to be compared to the growth temperatures of the heterostructures which lie between 500 and 700°C.

More problems arise from the threading dislocation density of $10^6 \text{ cm}^{-2}$ which is currently feasible but still seems to be too high for VLSI applications (recent research has demonstrated dislocation densities around $10^3 \text{ cm}^{-2}$ is step-graded buffer layers and are expected to fall to $100 \text{ cm}^{-2}$ and lower). The aim is to have a dislocation density as low as possible to avoid variations of the threshold voltage that might occur if a dislocation happens to be under a gate. Other concerns include the increased power consumption which results from source/drain-to-substrate leakage, for example. An increased source/drain leakage current due to the narrower band gap of the SiGe material has negative effects on the device performance, as well.

SiGe heterostructure technology is believed to have its biggest potential in selected VLSI applications where speed performance is the most important figure of merit. Furthermore, the reduced FET feature size in the next generations of VLSI technology results in the increased importance of high field transport where SiGe MODFETs are expected to have their maximum advantage.
Although the exposure of semiconductors to external magnetic fields is of little technological importance with respect to their use in device applications, it is regarded as the most powerful tool to evaluate the physics of semiconductors due to its analytical capabilities when combined with an optical or electrical field in a two-dimensional system. Hence, magnetic fields have become a crucial element in a variety of characterization techniques used to investigate and identify physical relations in general and for the characterization of a given semiconductor in particular. Two of the most important characterization techniques are based on magneto-optic and on magnetotransport phenomena, each giving a different insight into the semiconductor physics. Both types of phenomena can be explained by the greatly altered nature of the electronic states in presence of a strong magnetic field. All electronic states can be divided into the free electron or Bloch states and the bound states such as impurities and excitons (electron-hole coupled states). Magneto-optic phenomena can be seen in all electronic states, whereas magnetotransport phenomena can only occur in Bloch states since they are the only states that contribute to the electronic transport. Because of the nature of the band structure of the SiGe heterosystem, magneto-optic effects in general are too weak to provide any physical aspect about the investigated system, and most experiments are therefore based on magnetotransport phenomena. Hence, this chapter concentrates exclusively on the magnetotransport effects that are most commonly used to characterize a semiconductor.

The carrier concentration and the carrier mobility are the two most important transport parameters to characterize a semiconductor, and the best method to determine both quantities for a given semiconductor is to examine the electric field response of the carriers under the influence of a magnetic field. The resulting physical phenomena can vary qualitatively depending upon the strength of the applied magnetic field and the temperature. Accordingly, the study of the behaviour of the carriers in the presence of a magnetic field can be divided into two different types of possible effects, the magnetotransport phenomena in weak magnetic fields which can be described by the semi-classical approach and the magnetotransport phenomena in high magnetic fields which require the quantum mechanical approach. Experiments such as the Hall effect or the van der Pauw method employ weak magnetic fields, allowing the semi-classical approach in which the relation between $E$ and $k$ as given by the tight binding method is still valid, and consequently, the concept of the band structure is retained.

In strong magnetic fields, for which the condition $\mu B > 1$ has to be fulfilled, the free electron states are profoundly modified by the onset of quantum effects and can no longer be considered to be Bloch states, which has important consequences for the magnetotransport phenomena. In a 2DCG in a high magnetic field normal to its plane, the carriers are forced to move in cyclotron orbits parallel to its plane. Provided the carrier mobility is high enough and the temperature sufficiently low,
the quantizing magnetic field will cause the 2D DOS to become a series of δ-functions called Landau levels. Depending on the filling of these discrete states, a number of highly significant effects can be seen such as the Shubnikov-de Haas effect which produces oscillations in the longitudinal magnetoresistance and the quantum Hall effect which produces plateaux in the Hall voltage corresponding to the minima in the longitudinal magnetoresistance. For these effects to be observable, several conditions have to be satisfied in addition to the high-field requirement $\mu B > 1$ (which is equal to the condition $\omega_T > 1$). Furthermore, the system has to be degenerate ($E_F > k_B T$), the energy separation between successive Landau levels has to be greater than the thermal energy ($\hbar \omega_c > k_B T$), and the energy separation between successive Landau levels also has to be greater than the width of the Landau levels ($\hbar \omega_c > \Gamma$). A more detailed explanation of these conditions under which the quantum mechanical approach has to be employed will be given later.

Because of the quantum nature of the effects that can be seen in 2D systems under these conditions, the analysis of quantum mechanical magnetotransport phenomena yields considerably more and better information about a semiconductor than the semi-classical approach. Hence, the emphasis of this chapter is placed on the description and explanation of the Shubnikov-de Haas and the quantum Hall effect. Transport studies in weak magnetic fields, on the other hand, do not need a complex experimental set-up and thus are ideal for initial measurements as an indicator of the sample quality. Considering the rigorous experimental restrictions, quantum mechanical effects can generally not be observed except in the very lowest temperature regime, which leaves semi-classical magnetotransport measurements as the only alternative to determine the transport parameters at room temperature and all intermediate temperatures. On account of a much simpler physical concept, the semi-classical approach of the Hall effect and the van der Pauw method are described first. Based on these classical principles and those of quantum mechanics, it is possible to explain the behaviour of a 2D system in the completely confined quantum limit and subsequently the Shubnikov-de Haas and the quantum Hall effect. The last section of this chapter deals with the problem of parallel conduction that can occur within the channel when more than one subband is occupied, in the supply layer or, in the case of a gated structure, at the surface underneath the gate. Due to the lower mobilities of the carriers in such a parallel conducting channel, its contribution to the measured total mobility has a detrimental effect on the analysis of the magnetotransport phenomena and should therefore be avoided if possible.

3.1 SEMI-CLASSICAL APPROACH

Considering an electron travelling in the plane of a 2D system, its motion is characterized by its acceleration, the scattering experienced by the electron drifting through the material and the external transverse electric field. By additionally applying a magnetic field perpendicular to the plane, the electron is forced to move
in cyclotron orbits parallel to the plane, and the semi-classical equation for the motion of an electron in an electric field has therefore to be modified by the Lorentz force $\mathbf{v} \times \mathbf{B}$ to

$$m^* \frac{d\mathbf{v}}{dt} + \frac{m^* \mathbf{v}}{\tau} = -e(E + \mathbf{v} \times \mathbf{B})$$

(3.1)

where $\mathbf{v}$ is the drift velocity. The first term on the left represents the acceleration and the second term all scattering of the electron. Since the electric field has only components along the plane, i.e. $\mathbf{E} = (E_x, E_y, 0)$, and the magnetic field has only one component perpendicular to the plane, i.e. $\mathbf{B} = (0, 0, B_z)$, the motion of the electron within the plane is described by

$$v_x = -\mu E_x - \omega_c \mathbf{v}_y$$

(3.2)

$$v_y = -\mu E_y + \omega_c \mathbf{v}_x$$

(3.3)

where $\mu$ is the electron mobility and $\omega_c$ is the cyclotron resonance frequency which represents the circular motion of the electron perpendicular to the magnetic field. The electron mobility is given by the solution of the Boltzmann equation in the relaxation time approximation

$$\mu = \frac{e\tau}{m^*}$$

(3.4)

and the cyclotron resonance frequency is defined as

$$\omega_c = \frac{eB}{m^*}$$

(3.5)

Hence, it is possible to determine the effective mass of a 2D system directly by measuring the cyclotron resonance frequency. In general, measurements of the cyclotron resonance frequency are performed by observing the frequency of resonant absorption of radio frequencies or infra-red radiation.

It is this kind of behaviour that an electron displays in the presence of an electromagnetic field that can be exploited in the analysis of the Hall effect. In this section, the analysis of the Hall effect in a 2D system is explained using the standard Hall bar configuration in which only two different voltages have to be measured in order to determine the two main transport parameters. A single measurement of the transverse voltage yields the carrier concentration, but an additional measurement of the longitudinal voltage is necessary to determine the carrier mobility, as well. In samples that do not have the Hall bar configuration, first the sheet resistivity has to be determined using the van der Pauw method which in itself consists of two sets of measurements. Afterwards, the Hall voltage can be measured, and with the results from both sets of measurements, it is then possible to calculate both the carrier concentration and mobility.
3.1.1 Hall Effect

Standard magnetotransport measurements are performed on homogeneous and isotropic samples in the Hall bar configuration. A common Hall bar shape is depicted in Figure 3.1, essentially being a rectangle with a current contact at either end and potential probes along both sides. To minimize any misalignment voltages that can influence the measurement of the Hall voltage, the potential probes have to be positioned perfectly opposite each other. Misalignment voltages can be of great significance if the longitudinal voltages are much bigger than the transverse voltages. Usually, the structure of the Hall bar is etched and the electrical contact to the 2DCG is made by alloying a metal into the contact regions. If a current is injected in the x-direction of the Hall bar and a magnetic field is applied along the z-axis, an electric field will build up in the y-direction due to the Lorentz force exerted on the charge carriers. This Hall voltage exactly balances the Lorentz force and can be measured externally at the potential probes of the Hall bar. For simplification of the calculations, the electric field lines are assumed to be uniform along the length of the bar and errors due to geometry are expected to be negligible.

![Figure 3.1](image)

**Figure 3.1**  Hall bar geometry generally used in magnetotransport measurements. Positions 1 and 5 are the current contacts, the other positions are the potential probes [80].

The electrical transport in a two-dimensional layer in the presence of a magnetic field is described by the sheet resistivity tensor \( \rho \) which relates the sheet current density to the electric field through the relation

\[
E = \rho J.
\]  (3.6)

Accordingly, a sheet conductivity tensor \( \sigma \) can be defined to put the electric field into relation with the sheet current density via
Since the electric field and the sheet current density are two-dimensional vectors, the resistivity and conductivity tensors are 2x2 matrices and eqns (3.6) and (3.7) can be written as

\[
\begin{pmatrix}
E_x \\
E_y
\end{pmatrix} =
\begin{pmatrix}
\rho_{xx} & \rho_{xy} \\
\rho_{yx} & \rho_{yy}
\end{pmatrix}
\begin{pmatrix}
J_x \\
J_y
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
J_x \\
J_y
\end{pmatrix} =
\begin{pmatrix}
\sigma_{xx} & \sigma_{xy} \\
\sigma_{yx} & \sigma_{yy}
\end{pmatrix}
\begin{pmatrix}
E_x \\
E_y
\end{pmatrix}.
\]

Assuming isotropic material, certain relations between the components of the tensor matrix can be adopted, i.e. \( \rho_{xx} = \rho_{yy} \) and \( \rho_{xx} = -\rho_{yy} \). Furthermore, the resistivity tensor is the inverse of the conductivity tensor, and hence, the matrix elements of the conductivity tensor are connected to the matrix elements of the resistivity tensor through

\[
\sigma_{xx} = \frac{\rho_{xx}}{(\rho_{xx}^2 + \rho_{yy}^2)} \quad \text{and} \quad \sigma_{xy} = \frac{\rho_{xy}}{(\rho_{xx}^2 + \rho_{yy}^2)}.
\]

In order to completely describe the transport behaviour of a two-dimensional system in an electromagnetic field, each of the four matrix elements of the conductivity tensor has to be expressed by the transport parameter. The conductivity tensor is given by the general relation

\[
\sigma = n_e e \mu
\]

where \( \mu \) is not simply the electron mobility as given in eqn. (3.4) but rather a mobility tensor with its matrix elements defined by the equation for the drift velocity

\[
v = \mu E
\]

as established in eqns (3.2) and (3.3). Extracting the mobility tensor from eqns (3.2) and (3.3) and replacing it into eqn. (3.12) leads to an expression for the conductivity tensor

\[
\sigma = \frac{n_e e^2}{m^*}
\begin{pmatrix}
\tau & -\omega_c \tau^2 \\
\frac{1 + \omega_c^2 \tau^2}{1 + \omega_c^2 \tau^2} & \frac{1 + \omega_c^2 \tau^2}{1 + \omega_c^2 \tau^2}
\end{pmatrix}.
\]
Accordingly, the resistivity tensor can be described as

\[
\rho = \begin{pmatrix}
\rho_0 & \frac{B}{n_e} \\
-\frac{B}{n_e} & \rho_0
\end{pmatrix}
\]

with \( \rho_0 = \frac{m^*}{n_e e^2 \tau} \) \( \text{(3.15)} \).

The analysis of the Hall effect is based on the calculation of the longitudinal and the transverse magnetoresistance by measuring the respective voltages for a given current. A subsequent transformation into the corresponding resistivities enables the determination of the carrier concentration and the mobility through eqn. (3.15). First, the total current flow through the sample in \( x \)-direction is given by

\[
I_x = \int j_x(x,y) dy
\]

which is independent of \( x \) because of current conservation. The transverse and the longitudinal magnetoresistance, \( R_{xy} \) and \( R_{xx} \), are defined as the relation of the transverse and the longitudinal voltages, \( V_y \) and \( V_x \), to the current \( I_x \).

\[
R_{xy} = \frac{V_y}{I_x} = \frac{1}{I_x} \int E_y(x,y) dy = \frac{1}{I_x} \int \left( \rho_{xy}(x,y) j_x(x,y) - \rho_{xx}(x,y) j_y(x,y) \right) dy \tag{3.17}
\]

\[
R_{xx} = \frac{V_x}{I_x} = \frac{1}{I_x} \int E_x(x,y) dx = \frac{1}{I_x} \int \left( \rho_{xx}(x,y) j_x(x,y) - \rho_{xy}(x,y) j_y(x,y) \right) dx \tag{3.18}
\]

Assuming further a homogeneous resistivity tensor over the sample, i.e. the resistivity is independent of both \( x \) and \( y \), eqns (3.17) and (3.18) can be reduced to

\[
R_{xy} = \rho_{xy} \left( 1 - \cot \theta_H \int \frac{j_y(x,y) dy}{j_x(x,y)} \right) \tag{3.19}
\]

\[
R_{xx} = \rho_{xx} \frac{\int j_x(x,y) dx + \tan \theta_H \int j_y(x,y) dx}{\int j(x,y) dy} \tag{3.20}
\]

where \( \theta_H = \tan^{-1}(\rho_{xy}/\rho_{xx}) \) denotes the Hall angle between the directions of the electric field and the current. A geometry with a long Hall bar guarantees a homogeneous current flow with \( j_y(x,y) = 0 \), which reduced eqns (3.19) and (3.20) further to

\[
R_{xy} = \rho_{xy} \quad \text{and} \quad \tag{3.21}
\]

\[
R_{xx} = \rho_{xx} \frac{d_2 A}{d_3 \gamma} \quad \tag{3.22}
\]
with $d_{ij}$ being the sample dimensions between the contacts $i$ and $j$ of the Hall bar in Figure 3.1. Finally, the two-dimensional carrier density $n_s$ is given by

$$\rho_{xy} = -R_H B = \frac{r_H B}{n_s e} \Rightarrow n_s = \frac{r_H B}{e \rho_{xy}} \quad (3.23)$$

where $R_H$ is the Hall coefficient, $r_H$ the Hall scattering factor and $B$ the magnetic field. The Hall factor accounts for a certain distribution of the carrier drift velocities and the resulting imbalance between the velocity dependent Lorentz force and the velocity independent Hall field for those carriers that deviate from the average drift velocity. Its value ranges between one and two depending on the Fermi energy, the strength of the magnetic field and the scattering mechanisms present in the material and is given by

$$r_H = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle} \quad (3.24)$$

with $\langle \tau \rangle$ being the average of all relaxation times for each carrier in the crystal, which represents the mean free time between carrier collisions. Accordingly, the Hall mobility can be calculated as

$$\rho_{xx} = \frac{\rho_{xy}}{\mu_H B} = \frac{r_H}{\mu_H n_s e} \Rightarrow \mu_H = \frac{r_H}{n_s e \rho_{xx}} \quad (3.25)$$

Since the carrier drift mobility is defined by the relation

$$\mu = \frac{1}{n_s e \rho_{xx}} \quad (3.26)$$

the measured Hall mobility $\mu_H$ is related to the actual drift mobility $\mu$ via the Hall factor leading to

$$\mu_H = r_H \mu \quad (3.27)$$

Assuming the Hall factor to be one, which is in reasonable good agreement in the case of a degenerate system where the energy averages in eqn. (3.24) are replaced by the value of the relaxation time at the Fermi energy, the carrier drift mobility is exactly equal to the measured Hall mobility. The carrier sheet density of eqn. (3.23) then becomes

$$n_s = \frac{B}{e \rho_{xy}} \quad (3.28)$$

Hence, it is possible to determine the carrier sheet density by simply measuring the transverse voltage and the carrier mobility by additionally measuring the longitudinal voltage of the Hall bar.
3.1.2 **VAN DER PAUW METHOD**

It is not absolutely necessary to use samples with a Hall bar structure to determine the carrier mobility and the sheet density experimentally. By using the van der Pauw (VdP) method, it is possible to find the transport parameters for samples with an arbitrary shape. As illustrated in Figure 3.2, four ohmic contacts to the 2DCG have to be established by alloying, evaporating or simply soldering of a suitable metal. The contacts should be well-separated and positioned as symmetrical as possible near the edge of the sample, and the resistance between any pair of contacts should be less than $1\text{M}\Omega$. Although it is possible to apply the van der Pauw method to samples of arbitrary shape, sample structures with a four-fold symmetry are preferable since they produce smaller misalignment voltages. There is a variety of commonly used patterns, however, the standard configurations are the simple square geometry and the clover-leaf geometry.

![Figure 3.2 Magnetotransport measurements using the van der Pauw method. Shown are the simple square geometry (left) and the clover-leaf geometry (right).](image)

To measure the sheet resistivity, a current has to be injected using two adjacent contacts, and the voltage drop is measured between the opposite two contacts as shown in Figure 3.2. Subsequently, another set of measurements is taken with different permutations obtained by rotation of the contacts. Having measured the two values for the resistance given by

$$R_1 = \frac{V_{43}}{I_{12}} \quad \text{and} \quad R_2 = \frac{V_{32}}{I_{41}},$$

the sheet resistivity can then be determined using the relation

$$\rho_s = \frac{\pi}{2\ln 2} (R_1 + R_2) F$$

where $F$ is a correction factor for geometrical asymmetry as indicated by the two measured resistances $R_1$ and $R_2$ (and not for anisotropy or inhomogeneity of the material). The correction factor is given by
and becomes one for $R_1 = R_2$.

In order to measure the Hall effect, the current has to be injected using opposite contacts such that the Hall voltage can be measured across the current flow at the two remaining opposite contacts with the magnetic field being applied perpendicular to the sample. Having measured the sample current $I_s$ and the corresponding Hall voltage $V_H$ for a given magnetic field $B$, the Hall coefficient $R_H$ is then determined via

$$R_H = \frac{V_H}{I_s B} = -\frac{1}{n_s e}$$

(3.32)

where the Hall factor is still assumed to be one as discussed in the previous section. Ignoring the sign in the relation of the Hall coefficient, the carrier sheet density and mobility can be calculated according to

$$n_s = \frac{1}{e R_H} \quad \text{and} \quad \mu = \mu_H = \frac{R_H}{\rho_s}$$

(3.33)

(3.34)

with $\rho_s$ being the sheet resistivity as determined in the van der Pauw measurement. Although a bit more complicated concerning the experimental technique and less accurate, the van der Pauw method yields the same information about a semiconductor as the alternative assessment that employs the Hall bar configuration. The biggest advantage of the van der Pauw method over the Hall bar approach is its simplicity in the sample preparation which does not necessarily need a photo-lithographic processing step.

### 3.2 Quantum Mechanical Approach

In the previous section, it has been established that in magnetic fields, electrons and holes are forced to move in orbits perpendicular to the applied magnetic field, whereas their motion parallel to the field is unaffected by it. If the system is in the quantum regime, where $\omega_c \tau >> 1$ (or equivalently $\mu B >> 1$), the cyclotron motion of the carriers in the plane is quantized and the system can no longer be treated with the semi-classical approach. The kinetic energy of the carriers is modified into Landau levels due to the quantization of the cyclotron motion. An experimental set-up to put a system into the quantum regime requires low temperatures and high...
magnetic fields. Under these conditions, magnetotransport measurements provide a vast amount of information about the properties and parameters of the 2D system.

At first, the effect of strong magnetic fields on 2D systems and the formation of Landau levels will be described, to facilitate the interpretation of the Shubnikov-de Haas effect as well as the integer quantum Hall effect which will be explained subsequently. Possible complications in the interpretation of these magnetotransport phenomena due to parallel conduction will be dealt with in the last section.

### 3.2.1 Landau Levels

At low temperatures and sufficiently high magnetic fields, the system enters the quantum regime where the energy of the carriers becomes quantized in the plane perpendicular to the applied magnetic field. In order to describe the quantized energy states of the carriers in a three-dimensional system, it is necessary to employ the quantum mechanical approach which requires the time-independent Schrödinger equation to be solved

\[
H\Psi(r) = E\Psi(r).
\] (3.35)

In presence of an electromagnetic field defined by the vector potential \( A \) and the scalar potential \( \phi \), the Hamiltonian of the Schrödinger equation has to be modified accordingly to include the magnetic vector potential. Hence, the momentum operator \( p = \hbar k \) in the general Hamiltonian has to be replaced by \( p - eA \), leading to the revised expression for the Hamiltonian \([12,36]\)

\[
H = \frac{(p - eA)^2}{2m^*} + e\phi
\] (3.36)

where \( A \) is the magnetic vector potential, \( \phi \) is the scalar potential of the electromagnetic field and \( e \) is the elementary charge containing the sign for either electrons or holes. Adopting all other simplifications and approximations undertaken in section 2.1.2, the resulting Schrödinger equation can then be written as \([12,36]\)

\[
\left[ \frac{1}{2m^*}(-i\hbar \nabla - eA)^2 + e\phi \right]\psi(r) = E\psi(r).
\] (3.37)

Assuming the magnetic field to be in \( z \)-direction, i.e. \( \mathbf{B} = (0,0,B_z) \), and thus perpendicular to the \( xy \)-plane, the magnetic vector potential can be chosen to be the Landau gauge, i.e. \( A = (0,xB_y,0) \) \([12,36,81]\). Following this approach, the discrete values for the energy of the electrons and holes are specified by the eigenvalues of the Schrödinger equation

\[
E_n = \pm \left( n + \frac{1}{2} \right) \hbar \omega_c \pm \frac{\hbar^2 k_z^2}{2m^*} \quad \text{with} \quad n = 0,1,2,...
\] (3.38)
where $\omega_c$ denotes the cyclotron resonance frequency as defined in eqn. (3.5). Here, the plus sign represents the electron energies which are given with respect to the bottom of the conduction. Correspondingly, the energy values of the holes are determined by the minus sign with respect to the top of the valence band. The orbital motion of both electrons and holes is characterized by the magnetic length $l_B$ or the cyclotron radius $r_c$ of the lowest oscillator orbit ($n=0$) which is independent of material parameters

$$l_B = r_c = \left( \frac{\hbar}{eB} \right)^{1/2}.$$  \hspace{1cm} (3.39)

Unlike the cyclotron frequency which in general differs for electron and holes due to a different effective mass, the cyclotron radius is identical for both carrier types and depends only on the strength of the magnetic field. Furthermore, the cyclotron radius decreases with increasing magnetic field, whereas the cyclotron frequency increases.

The discrete energy levels in eqn. (3.38) are referred to as Landau levels and are characterized by the circular motion of the carriers in the $xy$-plane as described by the first term. It is important to note that as an electron or hole describes an orbit in $k$ space, it also describes a similar orbit in real space. The second term in eqn. (3.38) represents the continuous energy of the freely travelling wave parallel to the magnetic field in $z$-direction. In a two-dimensional system, the energy levels of the carriers are also quantized in $z$-direction due to the confining electrostatic potential, and therefore, the second term has to be replaced by the energy level of the appropriate subband. Hence, the Landau levels of a 2D system are given by

$$E_{i,n} = \pm E_i \pm \left( n + \frac{1}{2} \right) \hbar \omega_c \quad \text{with} \quad n = 0, 1, 2, \ldots$$ \hspace{1cm} (3.40)

where $E_i$ represents the lowest energy of the $i$-th subband. Each Landau level can be viewed as a constant number of electrons or holes ($e/\hbar$ per tesla, excluding spin and valley degeneracy) where the quantized energy of the circular movement perpendicular to the magnetic field as shown in Figure 3.3 (a) is that of a classical harmonic oscillator. Although the term describing this set of orbits does not contain the wavevector components $k_x$ and $k_y$ explicitly, it has to fulfil the relation

$$\left( n + \frac{1}{2} \right) \hbar \omega_c = \frac{\hbar^2 (k_x^2 + k_y^2)}{2m^*}$$ \hspace{1cm} (3.41)

of the circular movement in the $xy$-plane, which illustrates the quantization of the continuous energy into Landau levels quite clearly as shown in Figure 3.3 (b). The formation of Landau levels is a direct result of the energy becoming quantized in units of the cyclotron energy, and the energy of a 2D system is thus completely defined by the subband index $i$ and the Landau level index $n$. 

62
Magnetic quantization of the energy levels in a 2D system. The energy of each orbit described by electrons and holes in the xy-plane is given by the corresponding Landau level (a). Diagram of the $E$ vs $k$ relation of the Landau levels in which the energy distribution condenses from a continuum of states at zero magnetic field into states of equally spaced quantized energy levels (b).

In the definition of the Landau levels in eqns (3.38) and (3.40), the interaction of the electron and hole spin with the magnetic field has been ignored. Hence, it is necessary to include an additional energy component, the spin magnetic energy

$$E_s = s g \mu_B B$$

where $\mu_B$ is the free-electron Bohr magneton, $s = \pm 1/2$ is the spin quantum number and $g$ is the Landé $g$-factor. The free electron Bohr magneton is defined as

$$\mu_B = \frac{e h}{2 m^*}$$

and the magnitude of the Landé $g$-factor is 2 for free electrons but can be much larger in semiconductors due to strong spin-orbit coupling

$$g = 2 \left(1 + \frac{m^* - m}{m^*} \frac{\Delta_0}{3E_g + 2\Delta_0} \right)$$

where $\Delta_0$ is the spin-orbit splitting at the zone centre and $E_g$ is the band gap. As can be seen in Figure 3.4, the inclusion of the spin magnetic energy explains the splitting of each Landau level into two energy states of opposite spin which is called the Zeeman effect. Further splitting of each spin-split state into a (+) and a (-) valley occurs only for electrons in the SiGe system due to the existence of two identical energy valleys in the ground state.

Adding the spin magnetic energy to the energy eigenvalues of eqns (3.38) and (3.40), the total energy of the Landau levels for electrons and holes is given by
\[ E_{n,s} = \pm \left( n + \frac{1}{2} \right) \hbar \omega_c \pm \frac{\hbar^2 k_z^2}{2m^*} \pm s \mu_B B \quad \text{for a 3D system} \quad (3.45) \]

\[ E_{i,n,s} = \pm E_i \pm \left( n + \frac{1}{2} \right) \hbar \omega_c \pm s \mu_B B \quad \text{for a 2D system.} \quad (3.46) \]

At relatively low magnetic fields, the spin magnetic energy is small compared with the other components, and thus, spin splitting only occurs at higher magnetic fields with valley splitting only setting in at even higher values of the magnetic field.

The magnetic quantization also has an effect on the density of states which in the case of a three-dimensional system changes from being proportional to \( E^{1/2} \) to the one-dimensional variation proportional to \( E^{1/2} \) with its singularities evenly spaced in units of the cyclotron energy. A description of the DOS for the quasi-1D system can be given via

\[ N_{1D}(E) = \frac{eB}{4\pi^2} \frac{(2m^*)^{1/2}}{\hbar^2} E_{tot}^{1/2} \quad (3.47) \]

where the energy \( E_{tot} \) has to be summed up over all Landau levels

\[ E_{tot} = \sum_n \left[ E - \left( n + \frac{1}{2} \right) \hbar \omega_c \right] \quad (3.48) \]
which themselves are quantized in the \( xy \)-plane. The changeover of the DOS from that of a 3D system at zero magnetic field to that of a quasi-1D system is illustrated in Figure 3.5 (a). Similarly, the effect of magnetic quantization on a 2D system, where the energy in the \( z \)-direction is already quantized, leads to a quasi zero-dimensional system in which the energy is quantized in all three directions. Accordingly, the DOS is characterized by a series of \( \delta \)-functions defined by

\[
N_{0D}(E) = \sum_{i,n,s,v} n_i \delta(E - E_{i,n,s,v})
\]  

(3.49)

with \( n_i \) representing the degeneracy of the Landau levels. Here, the \( \delta \)-functions of the energy not only have to be summed up over all Landau levels \( n \) like for the quasi-1D system but also over the spin \( s \) and for electrons in the SiGe system over the valleys \( v \) until spin and valley splitting sets in, respectively. Figure 3.5 (b) shows the transformation of the 2D DOS into a quasi-0D DOS for magnetic fields below the onset of spin and valley splitting.

Since the average DOS is unaffected by the magnetic quantization, the number of states in zero field contained within the energy separation between two successive Landau states is exactly equal to the number of states per Landau level. Instead of having a 2D continuum of states, these states are all collapsed in a single degenerate Landau state. Hence, the degeneracy of the Landau levels is identical to the number of states per Landau level, spin, valley and unit area which is given by

\[
n_i = \frac{m^* \hbar \omega_c}{2\pi^2 \hbar^2} = \frac{m^* \omega_c}{2\pi \hbar} = \frac{eB_z}{\hbar}.
\]  

(3.50)

As a useful quantity to describe the sequential filling of the energy levels, the filling factor \( v \) is introduced which is defined as the ratio of the carrier sheet density \( n_s \) and the available states per Landau level, spin and valley \( n_L \).
\[ \nu = \frac{n_L}{n_t} = \frac{n_L h}{e B_z} \]

(3.51)

Taking spin and valley degeneracy into account, \( n_t \) has to be multiplied by \( g_s \) and \( g_v \) to obtain the number of states per Landau level for magnetic fields below the onset of spin and valley splitting, consequently modifying the relation for the filling factor accordingly.

Figure 3.6 shows the filling of successive Landau levels starting with the energetically lowest level and ending with the energetically highest level which may remain partially unoccupied since the carrier density \( n_s \) is generally not an integral multiple of the degeneracy \( n_t \). As the Fermi energy is the energy of the last electron (or hole) accommodated in the system at absolute zero temperature, it may be regarded as the energy that divides the filled and the empty levels of the system. In distributing the electrons over the levels, an abrupt break in energy occurs whenever one Landau level is completely filled due to the fact that an additional amount \( h \omega_c \) of energy is required to accommodate each electron in the next higher level.

In real quasi-0D systems, disorder due to scattering at imperfections such as phonons and impurities broaden the \( \delta \)-functions of the ideal DOS (collision broadening) and thermal smearing occurs for temperatures higher than 0 K (thermal broadening). The effect of both thermal and collision broadening is shown in Figure 3.7. All states in the tails of the broadened Landau levels are localized in space and separated in energy from the extended states by the mobility edge. Whether or not two successive broadened Landau levels overlap depends only on the magnitude of the magnetic field, assuming the thermal energy and relaxation time and thus the broadening of the Landau levels to be constant. As the cyclotron frequency increases with increasing magnetic field, so does the cyclotron energy and hence the separation of the broadened Landau levels.
3 Magnetotransport in 2D Systems Quantum Mechanical Approach

Since every state is occupied up to the Fermi energy, a small change of the magnetic field or the carrier density forces a smooth variation of the Fermi level until the corresponding Landau level is completely filled. Any additional electrons (or holes) have to lie in the next Landau level, and if the cyclotron energy is big enough for the broadened Landau levels not to overlap, the Fermi level is forced to jump to the lowest state of the next Landau level. At lower magnetic fields, where the cyclotron energy is smaller, the broadened Landau levels overlap and the variation of the Fermi level remains smooth.

### 3.2.2 Shubnikov-de Haas Effect

The oscillations of the longitudinal resistivity of a 2D system as a function of the magnetic field at low temperatures is known as the Shubnikov-de Haas (SdH) effect. This magnetotransport phenomenon is caused by the changing occupation of the Landau levels in the vicinity of the Fermi level as the Landau levels increase linearly in energy with increasing magnetic field due to the cyclotron energy. With increasing magnetic field, the Landau levels move up in energy and successively pass through the Fermi level as illustrated in Figure 3.8. However, because electronic states can only be occupied up to the Fermi energy, each Landau level is continuously emptied as it passes through the Fermi level. Since only the DOS close to the Fermi energy is of interest for the conductivity and hence the resistivity, both quantities have a maximum whenever the peak of a Landau level passes through the Fermi level.

As the Landau levels move further upwards in energy, the DOS around the Fermi energy gets smaller and so does the conductivity and resistivity until both reach a minimum when the Fermi energy lies exactly between two adjacent Landau levels. While a Landau level passes through the Fermi level, the occupied states close to the

---

**Figure 3.7** DOS of an ideal and a real quasi-0D system. The broadening of the \( \mathcal{\delta} \)-functions creates band tails with localized states, and the mobility edge is the energy that separates the localized and the extended states. The mobility gap represents the energy between the mobility edges of two consecutive Landau levels.
Fermi energy are either extended or localized, and since only the extended states contribute to the transport, conductivity and resistivity are zero when the Fermi level lies in the mobility gap where the states are localized.

![Density of States](image)

**Figure 3.8** Landau level energy and separation as a function of the magnetic field. With increasing magnetic field, the Landau levels move up in energy and their separation increases. Each Landau level is emptied as it passes through the Fermi level.

It is not only the energy of the Landau levels that increases linearly with magnetic field but also the energy separation between consecutive Landau levels. A Landau fan diagram is shown in Figure 3.9, illustrating the dependence of the Landau level energies on the strength of the magnetic field. As each Landau level passes through the Fermi level, the Fermi energy starts to oscillate smoothly with the DOS until the Landau levels do not overlap any more, forcing the Fermi energy to jump between successive Landau levels which produces ragged oscillations. As a result of the increasing energy separation of the Landau levels, the frequency of the oscillations of the conductivity and resistivity is continuously lowered with rising magnetic field.

![Landau Fan Diagram](image)

**Figure 3.9** Landau fan diagram showing the energy of successive Landau levels at $k_z=0$ as a function of the magnetic field. The Landau levels separate further in energy with increasing magnetic field. As long as the Landau levels overlap, the oscillations of the Fermi energy are smooth, but at higher magnetic fields, they cease to overlap and the Fermi energy has to jump to the next Landau level.
Shubnikov-de Haas oscillations have been established as a powerful tool to study 2D systems and characterize their properties. Among the most important transport properties to describe a 2D system that can be derived from this technique are the effective mass, the sheet density, the carrier mobility and the quantum and transport relaxation times.

The resolution of the oscillations is strongly dependent on the width of the broadened Landau levels, given by $\Gamma$ which is the full width at half maximum (FWHM). From the Rayleigh criterion follows the condition for $\Gamma$ to be less than the distance between the peaks of two successive Landau levels to resolve any oscillations

$$\Gamma \leq \hbar \omega_c.$$  \hfill (3.52)

In the case of the SiGe heterosystem on the other hand, $\Gamma$ is determined by its energy difference $\Delta E$ as illustrated in Figure 3.10 which in turn can be described according to the principle of uncertainty by

$$\Gamma = \Delta E = \frac{\hbar}{\tau}$$  \hfill (3.53)

and clearly shows that the broadening of the Landau levels is only dependent on the existing scattering in the system. Substituting eqn. (3.53) into eqn. (3.52) leads to the condition for the onset of any oscillations in the longitudinal conductivity and resistivity with magnetic field

$$\omega_c \tau \geq 1$$  \hfill (3.54)

which is equal to the condition

$$\mu B \geq 1$$  \hfill (3.55)

because of the relation

$$\omega_c \tau = \frac{eB}{m^*} = \mu B.$$  \hfill (3.56)

Although the nature of the level broadening depends strongly on the range of scattering potentials, their influence has been neglected in the definition of $\Gamma$ in eqn. (3.53). Hence, the following revised expression for the characteristic broadening of the Landau levels caused by short-range scattering potentials has to be introduced

$$\Gamma = \left( \frac{2}{\pi} \frac{\hbar \omega_c}{\hbar} \right)^{1/2}$$  \hfill (3.57)

which includes a dependence on the magnetic field through the cyclotron resonance frequency. The level broadening is thus not only dependent on the scattering potential but also on the magnetic field. Consequently, both the energy separation between successive Landau levels as well as the level broadening increase with
magnetic field with both effects being illustrated in Figure 3.8. Despite a quite
different definition for the level broadening, substituting eqn. (3.57) into eqn. (3.52)
leads to an only slightly modified condition for the onset of the oscillations

\[
\frac{\pi}{2} \omega_c \tau \geq 1
\]  

(3.58)

or alternatively

\[
\frac{\pi}{2} \mu B \geq 1.
\]  

(3.59)

Eqns (3.58) and (3.59) provide the means for a crude estimate of the mobility and the
transport relaxation time since at the onset of the oscillations, both terms equal one
and the magnetic field \( B \) and thus the cyclotron frequency \( \omega_c \) are known at that point.

As can be seen in Figure 3.10 (a), two successive Landau levels cannot be resolved if
their separation is smaller than \( \Gamma \) and hence no oscillation of the resistivity with
magnetic field can be seen. There are two possibilities to fulfil the condition
necessary to resolve the Landau levels. In Figure 3.10 (b), the magnetic field is
increased which enhances the separation of the Landau levels through an increase of
the cyclotron energy. If the magnetic field exceeds a certain minimum value, the
resistivity starts to oscillate. To achieve oscillations at lower magnetic fields, the
temperature can be decreased, as shown in Figure 3.10 (c), which reduces the
broadening of the Landau levels through an increased relaxation time.

Figure 3.10 Influence of cyclotron energy and relaxation time on the resolution of successive Landau
levels. Two adjacent Landau levels cannot be resolved because their cyclotron energy is
too small compared to their width \( \Gamma \) (a). Increasing the magnetic field and thus the
cyclotron energy separates the Landau levels further until they can be resolved and
oscillations in \( \rho_{xx} \) can be seen (b). Alternatively, decreasing the temperature and thus
increasing the relaxation time reduces \( \Gamma \) until the Landau levels can be resolved (c).

Another necessary condition that has to be fulfilled for the oscillations to be
resolvable is the requirement for the distance between the peaks of two adjacent
Landau levels to be greater than the thermal energy

\[
\hbar \omega_c \geq k_B T.
\]  

(3.60)
This expression also represents the limit for the semi-classical approach from which the magnetotransport is governed by quantum mechanics. The requirement of eqn. (3.60) can be explained by looking at the relaxation time which determines the FWHM of the Landau levels as well as the conductivity. An equation can be introduced which relates the relaxation time to the DOS and the Fermi energy according to

\[
\langle \tau \rangle = \frac{\int \tau N(E) \frac{\partial F(E)}{\partial E} dE}{\int N(E) \frac{\partial F(E)}{\partial E} dE}
\]

(3.61)

where \( F(E) \) denotes the Fermi-Dirac distribution. This relation can be interpreted as a weighting function in which the differentiated Fermi-Dirac distribution multiplied by the DOS is the weighting factor for the single relaxation times to determine the overall relaxation time and thus the conductivity. Figure 3.11 is an illustration of the temperature dependence of the Fermi-Dirac distribution and its derivative.

**Figure 3.11** Diagram of the Fermi-Dirac distribution and its derivative. For the ideal case at 0 K, the Fermi-Dirac distribution jumps from one to zero at the Fermi energy, and its derivative is a \( \delta \)-function (a). At low temperatures, the thermal smearing of the Fermi-Dirac distribution is moderate, and the FWHM of its derivative is small (b). At higher temperatures, the smearing becomes more pronounced, and the derivative is too broad to resolve the Landau levels (c).

Showing both functions for three different thermal energies, it can be seen that the width of the derivative through thermal smearing is strongly dependent on the temperature. As the weighting factor averages over all states within the derivative of the Fermi-Dirac distribution, the resolution of the Landau levels is highest at 0 K when the derivative is a \( \delta \)-function and hence only one state is selected. Raising the temperature lowers the resolution as the weighting factor broadens and thus averages over more and more states.

In order to explain the importance of the weighting function and its effect on the ability to resolve Shubnikov-de Haas oscillations, it has to be reiterated that only the
DOS in the vicinity of the Fermi energy is of importance for the conductivity, and it is the weighting function that determines the exact perimeter of this vicinity. In Figure 3.12, the principle of the weighting function is shown where the DOS of the Landau levels are weighted by the differentiated Fermi-Dirac distribution.

If the thermal energy is much smaller than the cyclotron energy, the broadening of the weighting factor is moderate compared to the spacing of the Landau levels, and since only a few states are weighted, the Landau levels can be resolved as they move through the Fermi energy with increasing magnetic field. Increasing the temperature or decreasing the magnetic field forces the weighting factor to broaden accordingly, and hence, any variations in the DOS change the overall contribution of the weighted DOS to a lesser and lesser extend, thus decreasing the resolution of the SdH oscillations. Eventually, when the thermal energy is larger than the cyclotron energy, the broadening of the weighting factor is comparable to the broadening of the Landau levels, which makes it impossible for them to be resolved. The fact that the weighting of the DOS by the derivative of the Fermi-Dirac distribution determines the total relaxation time and therefore the conductivity explains why only the DOS in the vicinity of the Fermi energy is of importance for the carrier transport. In the ideal case, in which the weighting factor is a δ-function, only one state contributes to the carrier transport at a given magnetic field.

At rather low magnetic fields, just above the onset of the SdH oscillations, the Landau levels are most probably closely spaced so that they overlap. Hence, the conductivity will not fall to zero when the Landau level move up in energy since there are always carriers in extended states to contribute to the transport. On the
other hand, the Fermi energy itself is subject to smooth variations with magnetic field due to the successive emptying of the Landau levels as they move through the Fermi energy. With increasing magnetic field, the space between consecutive Landau levels will become bigger, and from a certain point onwards, the Landau levels will cease to overlap. As a result, gaps of forbidden energy states (mobility gap) emerge which increase with the magnetic field. When a Landau level has completely passed the Fermi energy, the weighting factor would theoretically fall within such a mobility gap. However, since this is energetically not possible, the Fermi energy has to jump to the next Landau level, as shown in Figure 3.13, and consequently, the smooth oscillations of the Fermi energy become interrupted.

Initial considerations deduced a theoretical expression for the conductivity that consists of an oscillatory and a non-oscillatory part but assumed a uniform relaxation time for the whole expression. This was found to describe the oscillations insufficiently, and the expression was modified by considering two different relaxation times, the transport and the quantum lifetime. The full expression for the conductivity of the Shubnikov-de Haas oscillations is given by [82]

\[ \sigma_{xx} = \frac{n_e^2 \tau_t}{m^*} \left[ 1 - \frac{2(\omega \tau_q)^2}{1 + (\omega \tau_q)^2} \frac{\chi}{\sinh \chi} \exp \left( -\frac{\pi}{\omega \tau_q} \right) \cos \left( \frac{4\pi^2 \hbar \tau_q}{g_s \xi_s eB} \right) \right] \]

(3.62)

with

\[ \chi = \frac{2\pi^2 m^* k_B T}{\hbar e B} = \frac{2\pi^2 k_B T}{\hbar \omega_c} \]

(3.63)
Whereas this equation provides an expression for the conductivity, only the resistivity can be measured experimentally, and hence, the conductivity has to be inverted via eqn. (3.10) to give an expression for the resistivity

$$\rho_{xx} = \sigma_{xx}(\rho_{xx}^2 + \rho_{xy}^2)$$ \hspace{1cm} (3.64)

To get a direct relation between the two tensors, the term in the bracket has to be approximated. After a few assumptions and estimations, this term is given by [12]

$$\rho_{xx}^2 + \rho_{xy}^2 = \rho_{xx}^2(B = 0)[1 + (\omega_c \tau)^2].$$ \hspace{1cm} (3.65)

By combining eqns (3.62), (3.64) and (3.65), the resistivity can finally be written as

$$\rho_{xx} = \frac{\rho_{xx}^2(B = 0)n_s e^2 \tau_c}{m^*} \left[ 1 - \frac{2(\omega_c \tau)^2}{1 + (\omega_c \tau)^2} \frac{\chi}{\sinh \chi} \exp \left( -\frac{\pi}{\omega_c \tau} \right) \cos \left( \frac{4\pi^2 n_s}{\hbar s \hbar eB} \right) \right].$$ \hspace{1cm} (3.66)

The most important and astonishing aspect of this transformation is that the resistivity is directly proportional to the conductivity due to the approximation of eqn. (3.65). Hence, only the factor in front of the square bracket changes its appearance and in addition becomes a constant. The whole equation can be viewed as a cosine oscillation whose frequency is determined by the sheet density and the magnetic field and whose amplitude is governed by an exponential envelope function which is dependent on the magnetic field and shrinks with increasing temperature and effective mass. From this expression, all important transport parameters can be deduced.

According to the argument of the cosine, the oscillations should be periodic in the inverse magnetic field, which makes it possible to calculate the carrier density. To establish a relation between the magnetic field and the sheet density, eqn. (2.23) is used which states that the DOS in a 2D system is independent of the energy and thus the Fermi energy is defined as

$$E_F = \frac{2\pi \hbar^2}{g_s g_v m^*} n_s$$ \hspace{1cm} (3.67)

which has to coincide with the Fermi energy in the presence of a quantizing magnetic field

$$\left(n + \frac{1}{2}\right) \hbar \omega_c = E_F = \frac{2\pi \hbar^2}{g_s g_v m^*} n_s$$ \hspace{1cm} (3.68)

neglecting the smooth variations of the Fermi energy with increasing magnetic field. Every time the oscillations of the resistivity are at a minimum, the particular value $B_s$ of the magnetic field has to fulfill this relation for a given $n$. Solving eqn. (3.68) with respect to $nB_s$ leads to
\[ nB_n = \frac{h}{g_s g_v e} n_s \]  \hspace{1cm} (3.69)

which sets the order of the minima of the oscillations and the corresponding value for the magnetic field into a specific relation

\[ n = \frac{hn_s}{g_s g_v e B_n} = B_f \frac{1}{B_n} \]  \hspace{1cm} (3.70)

with \( B_f \) being the fundamental field. Hence, the density can be calculated via the equation

\[ n_s = \frac{g_s g_v e B_f}{h}. \]  \hspace{1cm} (3.71)

This method of calculating the sheet density can only be used when only one subband is occupied. If there were more than one occupied subbands, the oscillations would be the sum of as many sine waves as there are occupied subbands, all with different frequencies. Having determined the sheet density, the carrier mobility is found by simply applying eqn. (3.26) for the longitudinal resistivity at zero magnetic field

\[ \mu_H = \frac{1}{n_s e \rho_{xx}(B = 0)}. \]  \hspace{1cm} (3.72)

Subsequently, the calculation of the transport relaxation time can be calculated according to eqn. (3.56) as

\[ \tau = \frac{m^*}{e} \mu_H. \]  \hspace{1cm} (3.73)

The damping factor for the amplitudes of the oscillations in eqn. (3.66) is given by the envelope function \( \chi / \sinh \chi \cdot \exp(-\pi / \omega \tau_q) \) with both terms being dependent on the magnetic field, but whereas the first term is governed by the temperature, the second term involves the quantum relaxation time which in turn determines the collision broadening of the Landau levels according to eqn. (3.53). The quantum relaxation time can be found using a similar method to the one which determined the sheet density. First, the difference of the values of the resistivity between successive minima have to be calculated for a fixed temperature and subsequently divided by the \( (\chi / \sinh \chi) \) term to eliminate the temperature dependency. Taking the logarithm of this expression and plotting it as a function of \( 1/B_n \) in a Dingle plot should result in a straight line with its gradient yielding the Dingle temperature and which is proportional to the quantum relaxation time.

Finally, the effective mass can be determined by examining the variation of the amplitudes of the oscillations with temperature. By taking the values for the amplitudes \( A_n \) at the maxima and minima of the oscillations at a temperature \( T \) (with varying sign for the cosine term) and putting them into eqn. (3.66), the envelope
function can be characterized. By normalizing $A_n$ for a specific temperature $T_0$, the temperature independent exponential term will be eliminated leaving the relation

$$A_n(B,T) = \frac{T \sinh \left( \frac{2\pi^2 m^* k_B T}{\hbar e B} \right)}{T_0 \sinh \left( \frac{2\pi^2 m^* k_B T}{\hbar e B} \right)}.$$

(3.74)

All parameters in this theoretical expression are known except for the effective mass which can be calculated by experimentally measuring $A_n(B,T)$ and subsequently fitting the theoretical expression to the experimental values. Another more direct way to obtain the effective mass is the measurement of the cyclotron resonance frequency which is related to the effective mass according to eqn. (3.5).

### 3.2.3 Integer Quantum Hall Effect

One of the most important magnetotransport phenomenon affects the transverse conductivity of a high-quality 2D system and, like the Shubnikov-de Haas effect, can be seen in a system of completely confined quantum limit. At low temperatures and quantizing magnetic fields, the Hall resistivity deviates from its classical linear relation to the magnetic field as defined by eqn. (3.23) to form a series of quantized plateaux with Hall resistivities

$$\rho_{xy} = \frac{h}{ve^2}$$

(3.75)

where $v$ is the Landau level filling factor given by eqn. (3.51) and the value of $h/e^2$ equals 25.813 kΩ with an accuracy of better than one part in $10^4$. This integral quantization of the Hall resistivity was first observed and explained in a Si MOSFET [83], but better resolved plateaux were seen later in various heterostructure materials with lighter electron (and hole) effective masses such as GaAs/AlGaAs [84], InP/InGaAs and Si/SiGe. In those materials, the effect can be seen at higher temperatures and lower magnetic fields, thus relaxing the experimental requirements. Measurements of the quantum Hall effect (QHE) in different materials yield the same value for the quantized plateaux independent of the material or shape of the sample. The plateaux of the quantum Hall effect occur at integer filling factors and coincide with the minima of the Shubnikov-de Haas oscillations. At even filling factors, the Fermi energy lies in the mobility gap between successive Landau levels, whereas at odd filling factors, it resides between the spin or valley levels of the corresponding Landau level.

An explanation of this phenomenon can be given by looking at the effect of the existing imperfections in a 2D system which divides the electronic states into extended and localized states within each Landau level. Since electrons in localized states do not contribute to the electrical current which is hence purely composed of
electrons in extended states, the electric response of the system will differ depending on which of the two electronic states the Fermi level lies in. Whenever the Fermi energy resides within an area of extended states, a variation in density (MOSFET) or in magnetic field (2D system) adds carriers to or subtracts carriers from consecutive extended states, and as a consequence, the conductivity of the system changes linearly with the density or the magnetic field. However, as long as the Fermi level coincides with the energy of the localized states, any variation in density or magnetic field only changes the density of electrons in localized states, whereas the density of electrons in extended states and thus the conductivity of the system remains constant, leading to the occurrence of the plateaux in the Hall resistivity. Hence, it is the imperfection of a real 2D system which causes the plateaux as the signature of the quantized Hall effect. In Figure 3.14, the relation between the DOS of the Landau levels with their localized and extended states and the longitudinal and Hall resistivity is illustrated. From this relation it is obvious, that the width of the Hall plateau, which can be correlated with the sample mobility, is dependent on the fraction of localized states within a Landau level.

Maybe the most astonishing fact is the high accuracy to which the values of the plateaux are quantized, which means that the value of the Hall resistivity at a plateau is always exactly equal to $h/ve^2$, independent of the investigated material or structure [86]. Since a given fraction of the carriers are localized, the density of the extended states is diminished and the number of carriers to support the current flow in each Landau level deviates considerably from its value given by eqn. (3.50). Consequently, the Hall resistivity is expected to deviate accordingly from its quantized value $h/ve^2$ and like the width of the plateaux should differ for different samples because their
Landau levels have different fractions of localized states. An answer to this paradox was given by several authors who performed calculations of the current flow under the given circumstances and obtained the startling result that the remaining carriers in the extended states make up in current for the localized carriers by increasing their own velocity [85-89]. The total current is kept unchanged by the potential which localizes some of the carriers but also accelerates the remaining carriers in extended states such that the reduction in the number of mobile carriers is exactly compensated by their increased velocity. Hence, the overall gradient of the Hall resistance can still be used to determine the carrier density given by eqn. (3.28), and subsequently, the Hall mobility can be calculated according to eqn. (3.72).

The high precision of the quantized values is of fundamental physical importance and can be used to calculate the fine-structure constant $\alpha$ which is a measure of the coupling between elementary particles and the electromagnetic field with an accuracy of one part in $10^9$.

### 3.3 Parallel Conduction

Complications in the interpretation of the magnetotransport properties can arise when more than one type of carrier contribute to the transport. The presence of spatially separated carriers conducting in parallel channels with different densities and mobilities requires a model that relates the total conductivity to the individual transport properties of each conducting channel [90, 91]. Basically, two different types of parallel conduction have to be considered, one being the possibility that more than one subband is occupied (high mobility parallel layer), the other being that parallel conduction can occur in the supply layer or underneath the gate in gated structures (low mobility parallel layer). For both types of parallel conduction, the model assumes that each type of carrier experiences locally the same electric field.

The current of each conducting channel is given by eqn. (3.9), and each corresponding conductivity is fully described by the conductivity tensor of eqn. (3.14), which can be written in the simpler form

$$\sigma_i = \begin{pmatrix} D_i & -A_i \\ A_i & D_i \end{pmatrix}$$

(3.76)

with

$$D_i = \frac{n_i e^2}{m_i} \frac{\tau_i}{1 + \omega_c^2 \tau_i^2}$$

(3.77)

$$A_i = \frac{n_i e^2}{m_i} \frac{\omega_c^2 \tau_i^2}{1 + \omega_c^2 \tau_i^2}$$

(3.78)
where \( i \) is the subscript labelling each channel, and \( m_i \) and \( n_i \) are the effective mass and the sheet density of the carriers, respectively. The total conductivity of the system is simply given by the sum of the parallel conductivities, and considering a two layer system, the total conductivity tensor can then be written as

\[
\sigma_{\text{tot}} = \begin{pmatrix}
(D_1 + D_2) & -(A_1 + A_2) \\
(A_1 + A_2) & (D_1 + D_2)
\end{pmatrix}.
\] (3.79)

Usually, the measured quantities are the components of the total resistivity tensor which can be obtained by inversion of the total conductivity tensor with the longitudinal and transverse components given by

\[
\rho_{xx} = \frac{E_x}{I_x} = \frac{D_1 + D_2}{(D_1 + D_2)^2 + (A_1 + A_2)^2}
\] (3.80)

\[
\rho_{xy} = \frac{E_y}{I_x} = R_H B = \frac{-(A_1 + A_2)}{(D_1 + D_2)^2 + (A_1 + A_2)^2}
\] (3.81)

Substituting the appropriate expressions for the variables into eqns (3.80) and (3.81) shows that both the longitudinal resistivity and the Hall coefficient change smoothly between the limits of low and high magnetic field.

(i) **THE LOW MAGNETIC FIELD LIMIT** \((\omega_c \tau_1 << 1, \omega_c \tau_2 << 1)\)

\[
\rho_{xx} = \frac{1}{n_1 e \mu_1 + n_2 e \mu_2}
\] (3.82)

\[
R_H = \frac{n_1 \mu_1^2 + n_2 \mu_2^2}{2(n_1 \mu_1 + n_2 \mu_2)^2}.
\] (3.83)

These well known expressions for mixed conduction in low magnetic fields show that the total longitudinal conductivity of eqn. (3.82) is simply the sum of the conductivity of both layers and thus is dominated by the high mobility carriers, and eqn. (3.83) implies that the measured Hall carrier density is also dominated by the high mobility carriers.

Hence, the higher the mobility of the carriers in the parallel conduction layer (or more precisely, the smaller the difference in mobility between the high and low mobility carriers), the greater their influence on the measured overall mobility. At low temperatures, parallel conduction in the supply layer or the surface channel can therefore be neglected as their carrier mobility is much lower than that in the QW (by several orders of magnitude). Carriers that occupy higher subbands, however, have much higher mobilities which can be up to 25% of those in the lowest subband and thus have a much bigger influence on the measured mobility. With increasing temperature, the mobility of the carriers in the low mobility parallel layers rises, and
at room temperature, it becomes much more comparable to the mobility of the carriers in the QW which decreases dramatically with increasing temperature.

(ii) **THE HIGH MAGNETIC FIELD LIMIT** \((\omega_c \tau_1 >> 1, \omega_c \tau_2 >> 1)\)

\[
\rho_{xx} = \frac{(n_1/\mu_1) + (n_2/\mu_2)}{e(n_1 + n_2)^2} \quad (3.84)
\]

\[
R_H = \frac{1}{e(n_1 + n_2)}. \quad (3.85)
\]

In contrast to the low magnetic field limit, the expression for the resistivity in eqn. (3.84) is now dominated by the low mobility carriers. Consequently, a small number of carriers in a low mobility parallel layer will cause a large magnetic field dependence in the resistivity of the structure, and the higher the mobility of the 2D gas, the fewer low mobility carriers are required to introduce a substantial magnetic resistance. According to eqn. (3.85), the measured Hall carrier density is simply the sum of the sheet carrier densities of both channels and is therefore independent of any mobility. In general, it is possible to obtain carrier densities in the lowest subband much higher than in high or low mobility parallel channels which can then be neglected.

(iii) **MAGNETIC FIELDS FOR WHICH** \((\omega_c \tau_1 >> 1, \omega_c \tau_2 << 1)\)

\[
\rho_{xx} = \frac{1}{e} \left( n_2/\mu_2 + n_1^2/\mu_2 B^2 \right)^{-1} \quad (3.86)
\]

\[
R_H = \frac{1}{e} \left( n_1 + n_2/\mu_2 + n_1^2 B^2 \right)^{-1} \quad (3.87)
\]

A third case has to be considered in which both channels are in a different magnetic field limit, i.e. one channel is in the high magnetic field limit whereas the other channel is in the low magnetic field limit. In this case, the resistivity as well as the measured Hall density have a quadratic dependence on the magnetic field.

At low temperatures, the population of higher subbands can be avoided by keeping the carrier sheet density in the QW below the corresponding critical value for the given structure. In addition, all carriers in the supply layer and underneath the gate should be frozen out, and therefore, neither SdH nor Hall measurements should be influenced by any kind of parasitic parallel conduction. However, if mobile carriers are left in any of the parallel conduction layers, a difference in the density determined by the two experimental techniques will be noticeable because only an average of all densities can be extracted from the Hall measurements, whereas in the SdH effect, the density of each contributing layer is measured individually. The influence of populated higher subbands on the SdH effect is clearly visible as each
subband contributes oscillations of different frequencies to the SdH oscillations where the higher the subband, the lower the frequency. How many subbands are occupied cannot be detected directly from the oscillations themselves but by analysing the power spectrum of the SdH oscillations which will have a peak for each subband at its corresponding frequency. In contrast, low mobility parallel conduction is not observable in SdH oscillations because collision broadening causes an extreme smearing of the Landau levels which can thus not be resolved.

At room temperature on the other hand, parallel conduction in the supply layer or a surface channel as well as in higher subbands is much more likely, and due to a much reduced mobility imbalance compared to the lowest subband and a much higher density, their effect on the measured Hall mobility is quite significant.
According to Bloch's theorem, an electron or hole suffers no collisions during its motion through a periodic structure even though it is exposed to a complex background potential. However, this statement is dependent upon the perfect periodicity of the crystalline potential, a presumption which is never fulfilled in a real crystal where various types of imperfections are responsible for scattering of the electrons or holes between the Bloch states. These imperfections can be intentional such as alloy growth or the introduction of dopants, accidental impurities incorporated during growth or vibrations of the lattice. The effect of these imperfections on the carrier transport can be calculated by applying the first order perturbation theory [11, 12, 36, 92].

A scattering process is fully described by its scattering rate $W$ which is the probability of an electron or a hole being scattered from one Bloch state in $k$-space to another or alternatively by its corresponding relaxation time $\tau$. How electronic collisions are to be described depends on the specific collision process of importance for which the scattering potential has to be identified and its matrix element $M_{kk'}$ has to be evaluated between an initial state $|k\rangle$ and final state $|k'\rangle$ with

$$|k\rangle = u_k e^{ikr} \text{ and } |k'\rangle = u_{k'} e^{ik'r}. \tag{4.1}$$

The scattering rate of a particular collision process in the first order perturbation theory is then given by the Fermi Golden rule

$$W(k,k') = \frac{2\pi}{\hbar} |M_{kk'}|^2 \delta(E(k') - E(k) - \Delta E) \tag{4.3}$$

where $\Delta E$ is the energy change during scattering and the $\delta$-function accounts for energy conservation. In general, the motion of an electron or hole through the crystal is not simply perturbed by a singular scattering process, but there is a mixture of different scattering mechanisms present during carrier transport. If the various scattering processes are independent of each other, the total scattering rate (or scattering probability) $W_{\text{tot}}$ is just the sum of the individual scattering rates where a specific relaxation time is associated with each individual scattering process

$$W_{\text{tot}} = \sum_n W_n = \sum_n \frac{1}{\tau_n}. \tag{4.4}$$

Since it is almost impossible to observe the scattering rate for a single electron or hole directly, experimental techniques typically study phenomenological quantities such as the mobility which are properties of a population of electrons or holes. To deduce a scattering rate from the mobility requires the knowledge of the energy distribution.
of the electrons or holes. At low fields, this distribution can be assumed to be that at
equilibrium but slightly perturbed, and the first order solution of the Boltzmann
equation remains valid. Hence, the basic statistical relation between the mobility and
the relaxation time is given by the elementary transport theory as

$$\mu = \frac{e_0 \langle \tau \rangle}{m^*}$$  \hspace{1cm} (4.5)

where for thermal equilibrium, non-degenerate statistics and parabolic bands, the
averaged relaxation time is given by

$$\langle \tau \rangle = \frac{\langle E(k) \tau \rangle}{\langle E(k) \rangle}.$$  \hspace{1cm} (4.6)

Similar to the relation between individual and total scattering rates in eqn. (4.4), the
final mobility can be written as the sum of all individual mobilities according to
Matthiessen’s rule

$$\frac{1}{\mu_{\text{tot}}} = \sum_n \frac{1}{\mu_n}.$$  \hspace{1cm} (4.7)

However, the individual mobilities can only be added up this way if all existing
scattering mechanisms are totally independent of each other and do not exert any
influence on any of the contributing scattering processes. Furthermore, Matthiessen’s
rule is strictly valid only when all the scattering mechanisms have the same energy
dependence which is usually not the case. However, eqn. (4.7) is widely used
because of its ease of application and reasonable accuracy.

Because of their distinct physical properties, bulk semiconductors and their
_corresponding 2D heterosystems possess different scattering mechanisms which
have to be discussed separately. In principle, carriers in a 2DCG experience the same
scattering mechanisms that are also present in the bulk material, and thus, each
scattering process in a 2D system can be treated in similar fashion as in the bulk with
corresponding adjustments for strain or dimensionality. Apart from those
modifications, further scattering mechanisms have to be added arising from
additional scattering centres due to the pseudomorphic growth and the sequential
layer structure of 2D heterosystems. Taking into account the contrasting structures of
2DEG and 2DHG systems, the dominant scattering mechanisms that influence the
mobility of the electrons in a QW are expected to be quite different from those
governing the mobility of two-dimensional holes.

Two principal scattering mechanisms dominate the carrier movement in bulk
semiconductor materials, one is time dependent and the other one is independent of
time. The time-dependent scattering mechanism is caused by thermal vibrations of
the lattice, and its interaction with electrons or holes is described by acoustic and
optical phonons that can be absorbed or emitted by the carriers. Impurity scattering,
on the other hand, is a time-independent scattering mechanism that can have many
different sources. Charged and neutral impurities are among the main contributors
to impurity scattering as is alloy scattering and electron-electron scattering. All possible scattering events in a bulk semiconductor can therefore be classified under either lattice scattering or impurity scattering. In quantum wells, additional scattering processes have to be taken into consideration such as interface roughness scattering, interface charge scattering, alloy scattering and scattering from threading dislocations. Furthermore, impurity scattering has to be differentiated into remote and background impurity scattering, and strain in the QW will alter the effect of lattice scattering.

4.1 RELAXATION TIMES

Before the various scattering events are explained in detail, it is important to comprehend the concept of the different relaxation times and their significance in relation to the different types of scattering mechanisms. There are two different relaxation times which have to be distinguished, the quantum or single-particle relaxation time \( \tau_q \) and the transport or momentum relaxation time \( \tau_r \). The quantum relaxation time becomes finite in the presence of perturbations produced by scattering potentials and is related to the half-width \( \Gamma \) of the broadened Landau level through

\[
\Gamma = \frac{\hbar}{2\tau_q}.
\] (4.8)

This approach assumes that the thermal broadening of the Landau levels can be represented by a Lorentzian distribution for which \( \Gamma \) is independent of energy or magnetic field. The quantum relaxation time is a measure of the time for which an electronic momentum eigenstate can be defined even in the presence of scattering. In contrast, the transport relaxation time is the rate at which the net momentum of all electrons is randomized and is related to the conductivity of the two-dimensional carrier gas by

\[
\sigma = n_s e \mu = \frac{n_s e^2 \tau_l}{m^*},
\] (4.9)

where \( n_s \) is the carrier density, \( e \) the elementary charge, \( \mu \) the Hall mobility and \( m^* \) the effective mass. One important feature of this relation is the fact that \( \tau_l \) is directly proportional to the mobility.

The quantum relaxation time can be described by integrating the quantum-mechanical transition rates \( W(k,k') \), where \( k \) denotes the initial and \( k' \) the final state, over all scattering angles \( \theta \) and is defined as

\[
\frac{1}{\tau_q} = \int W(k,k')d\theta.
\] (4.10)
$W(k,k')$ is proportional to the probability of scattering from state $k$ to state $k'$ through an angle $\theta$, thus taking into account both small-angle and large-angle scattering, i.e. all scattering events contribute equally to $\tau_r$. The transport relaxation time, on the other hand, is derived from the Boltzmann equation in the relaxation time approximation as
\[
\frac{1}{\tau_i} = \int W(k,k')(1 - \cos \theta) d\theta. \quad (4.11)
\]

Due to the weighting of the quantum-mechanical transition rates with the momentum loss factor $(1 - \cos \theta)$, $\tau_i$ is dominated by large-angle scattering. By definition, the only difference between the two relaxation times is that $\tau_i$ is insensitive to small-angle scattering while $\tau_q$ is sensitive to all scattering events [82, 93-98].

The significance of this relation is that the quantum and transport relaxation times are identical for short-range scattering such as interface charges or ionized background impurities in the QW because the value of $\cos \theta$ vanishes. For long-range scattering like remote impurities, on the other hand, the transport relaxation time will become much larger than the quantum relaxation time. Hence, it is possible to obtain information on the scattering mechanisms by measuring both relaxation times and subsequently evaluating the ratio $\tau_i/\tau_q$ which allows a classification of the dominant scattering process in a 2D system. The ratio $\tau_i/\tau_q$ will be close to unity if isotropic short-range scattering is predominant and much bigger than unity if long-range scattering prevails [93-98].

### 4.2 Lattice Scattering

In all semiconductors, the properties of an electron or hole that travels through the crystal is determined primarily by the periodic potential associated with the array of atoms at the lattice points. In a perfect periodic lattice, electrons and holes can move without any electrical resistance, which is known as ballistic transport. A perfect periodic lattice does not cause electrons or holes to scatter, and hence, all resistivity is due solely to deviations of the potential from the ideal state. Thermal vibrations of the atoms about their equilibrium positions, for example, produce practically instantaneous changes in the energy of electrons and holes and thus introduce a time-dependent component $H_{\text{ep}}$ into the time-independent adiabatic one-electron Schrödinger equation. These perturbations can be expressed in terms of a scattering rate $W$ by the equation of the first-order perturbation theory
\[
W = \frac{2\pi}{\hbar} \int |(k'|H_{\text{ep}}|k)|^2 \delta(E_{k'} - E_k - \hbar \omega_{\text{ph}}) dk'. \quad (4.12)
\]

Here, $H_{\text{ep}}$ is the Hamiltonian of the electron-phonon interaction, $E_{k'}$ and $E_k$ are the energies of the final and initial electron states $k'$ and $k$, respectively, the time
dependence is incorporated in the energy conserving $\delta$-function, and the integral is over all final states $k'$. The electron-phonon coupling matrix element of $H_{ep}$ includes an expression associated with the sum of all individual lattice waves

$$M_{kk'} = \frac{1}{V} \int u_k^*(r) e^{-i k' r} H_q(r) e^{i q R} u_k(r) e^{i k r} \, dr,$$  \hspace{1cm} (4.13)$$

where the interaction constant $H_q(r)$ represents the part of $H_{ep}$ that is dependent on the electron or hole co-ordinate for a given mode and is independent of $R$ for lattice scattering. Suppressing $H_q(r)$ and factorizing the integral over the crystal into an integral over a unit cell and an integral over all unit cells over the volume $V$ leads to

$$M_{kk'} = \frac{1}{V} \int e^{i (k+q-k') R} dR \cdot \int_{cell} u_k^*(r) u_k(r) e^{i (k+q-k') r} \, dr.$$  \hspace{1cm} (4.14)$$

Since the product $u_k^*(r) u_k(r)$ is periodic in the crystal lattice, the summation vanishes unless

$$k + q - k' = K,$$  \hspace{1cm} (4.15)$$

where $K$ is a vector in the reciprocal lattice. A scattering event with $K=0$ is called a normal process in which the crystal momentum is conserved. However, since the periodicity of the crystal involves a finite minimum distance, the conversation law is not exactly obeyed and processes with $K \neq 0$ are allowed. This type of scattering event is said to be an umklapp process which generally does not play a significant role in electron scattering in semiconductors since the integrand in eqn. (4.14) varies rapidly for $K \neq 0$, hence significantly reducing the value of the integral. It is important to note that since $\hbar k$ is the crystal momentum rather than the true momentum of the electron, the apparent violation of the momentum conversation is not a real violation because total momentum is conserved even in an umklapp process.

The interaction between electrons or holes and the thermally vibrating lattice is represented by a scattering process in which phonons are either absorbed or emitted by the carriers. Phonons can be seen as quantized lattice vibrations with energy

$$E(\omega_q) = \left(n(\omega_q) + \frac{1}{2}\right) \hbar \omega_q$$  \hspace{1cm} (4.16)$$

where $n(\omega_q)$ is the statistical average number of phonons in the mode $\omega_q$ which at thermodynamic equilibrium is given by the Bose-Einstein function

$$n(\omega_q) = \frac{1}{\exp(\hbar \omega_q / k_B T) - 1}. \hspace{1cm} (4.17)$$

In agreement with the momentum conversation law, an electron is scattered from an initial state $k$ into a final state $k'$ upon emission or absorption of a phonon of wave-vector $q$ where
In this process, the electron loses or gains momentum $\hbar q$ and energy $\hbar \omega_q$ in accordance with the energy conservation law

$$E_k = E_{k'} \pm \hbar \omega_q.$$  \hspace{1cm} (4.19)

In both cases, the upper sign represents phonon emission and the lower phonon absorption. Although phonons do not carry any momentum, they behave as if they had a momentum $\hbar k$, also called crystal momentum, which will be changed upon interaction with particles such as electrons or holes. The actual physical momentum of the lattice vibrations is zero since the atoms are moving against each other and the crystal as a whole is not moving.

In the case of a lattice with two atoms per primitive unit cell, such as the diamond lattice of Si and Ge or the zinc-blende lattice of GaAs, two different phonon modes have to be distinguished, the acoustic and the optical mode. For the acoustic mode, the two sublattices move in the same direction, whereas for the optical mode, they move in opposite directions. According to their orientation within the lattice, a further classification of the two modes has to be accounted for. All modes that have the displacement vectors of each atom along the direction of the wavevector are called longitudinally polarized modes, and those with atoms moving in the plane normal to the wavevector are called transversely polarized modes. Consequently, the phonon spectrum consists of one longitudinal acoustic mode (LA), one longitudinal optical mode (LO), two transverse acoustic modes (TA) and two transverse optical modes (TO). Hence, three acoustic and three optical modes exist in the bulk material of those semiconductors but only one of these modes can be involved in a single lattice scattering event. The phonon dispersion curves of the three acoustic and three optical modes are illustrated in Figure 4.1 for a lattice with two atoms per unit cell.

![Figure 4.1](image)

**Figure 4.1** Acoustic and optical modes of the phonon dispersion relation for a diatomic lattice. The curves are not continuous but rather very closely spaced points [11].

87
Optical phonons have little dispersion across the whole Brillouin zone, and at small \( k \), the LO phonon frequency is higher than the TO phonon frequency. For long-wavelength acoustic modes, the phonon dispersion relation is given by

\[
\omega_q = v_s q
\]  

where \( v_s \) is the sound velocity. The fact that the frequency is directly proportional to the wavevector is a statement that the sound velocity is independent of frequency in the long-wavelength limit of the acoustic modes. For all other cases, it is often useful to approximate their dispersion by \( \omega = \text{const} \). In Figure 4.2, the measured phonon dispersion relation is shown for Si, Ge and GaAs, illustrating the almost flat and thus dispersionless curves of the optical phonon modes. Whereas there is no splitting of the LO and TO frequencies for group IV semiconductors at \( k=0 \), the ionic crystal of the III-V compounds causes a difference in frequency between the optical modes. Except in the long-wavelength limit, the dispersion of the TA phonon modes can also be assumed to be constant.

![Figure 4.2 Measured phonon spectra in Si, Ge and GaAs along their respective direction of high symmetry. Hence, both TO modes and both TA modes coincide in all three materials [12, 35].](image)

In its simplest form, the perturbation energy depends linearly on strain, acoustic strain in the case of acoustic modes and optical strain in the case of optical modes. One effect of these strains on the electrons or holes are the short-range disturbances of the periodic potential which cause instantaneous changes in energy quantified by deformation potentials. Disturbances of the electron's or hole's motion by this effect is referred to as deformation-potential scattering which is common to all semiconductors. A second type of strain-related disturbance only occurs in polar materials such as GaAs. It is long-range in character and stems from a macroscopic electric field associated with the destruction of the local electric neutrality through the lattice distortions which results in electric polarization. This effect is referred to as piezoelectric scattering where acoustic modes are involved, and polar optical scattering associated with optical modes.
Scattering of an electronic state can take place either by absorption or emission of a phonon. Furthermore, an electronic state can be scattered either within the same energy valley or between two different valleys which in turn can be either equivalent or non-equivalent in energy. In the next two sections, the energy and momentum conversation are discussed separately for intra-valley and inter-valley scattering, and in the subsequent two sections, the most important aspects of acoustic and optical phonons are reviewed.

4.2.1 INTRA-VALLEY SCATTERING

For intra-valley scattering, the different scattering events by either absorption or emission of acoustic or optical phonons have to be calculated separately as shown in Figure 4.3. Each of them is considered to be a normal process in which the crystal momentum is conserved. Hence, the final state \( k' \) to which the initial state \( k \) is scattered is determined by the phonon wavevector \( q \) and the scattering angle \( \theta \) as

\[
k'^2 = k^2 + q^2 \pm 2kq \cos \theta
\]

where the plus sign stands for the absorption of a phonon and the minus sign for the emission of a phonon. Accordingly, the energy of the final state in a spherical parabolic band is given by the energy conversation equation

\[
\frac{\hbar^2 k'^2}{2m^*} = \frac{\hbar^2 k^2}{2m^*} \pm \hbar \omega
\]

The shapes of the corresponding band extrema, which can be either spherical or ellipsoidal, as well as a possible nonparabolicity of the bands have to be accounted for and thus included in the energy calculations. In ellipsoidal parabolic bands, the energy of an electronic state varies with direction and thus is dependent on the magnitude of all three components of \( k \). By transforming all vector components into a suitable space, i.e. the transformation of an ellipsoid into a sphere, the necessary conditions can be established in which the energy is independent of direction. For nonparabolic bands, the energies have to be approximated by truncated Taylor expansions [36].

Scattering by absorption or emission of acoustic phonons is approximately elastic since the energy \( 2 \hbar \nu k \) of the acoustic phonon wavevector \( q \) is small compared to the corresponding thermal energy \( k_B T \), which means that the direction of the initial state \( k \) can change considerably, but its magnitude stays more or less constant. In acoustic phonon scattering events, the average group velocity is an important quantity that determines the limits of the phonon wavevector. The group velocity represents the velocity of the energy transmission in the lattice which is at maximum at the zone centre and zero at the edge of the Brillouin zone. At room temperature, the average group velocity is typically much greater than the sound velocity, and therefore, most electrons can absorb phonons with wavevectors in the range \( 0 \leq q \leq 2k \). Whereas
phonon absorption is allowed for all group velocities, an electron must travel at a group velocity in excess of the sound velocity to emit a phonon, and in addition, only forward emission is allowed.

Due to the significant magnitude of the energy $\hbar \omega_0$ associated with the optical phonon wavevector, scattering by absorption or emission of optical phonons is considered inelastic. However, the modes that can interact with the electrons are those with wavevectors near the zone centre which are long-wavelength modes, just as for acoustic phonon scattering. For very energetic electrons or holes, their energy can be much bigger than that of the optical phonons ($E_k \gg \hbar \omega_0$), and the scattering is approximately elastic. As in the case of acoustic phonon emission, only forward emission of optical phonons is allowed, and the electron must have sufficient energy ($E_k \geq \hbar \omega_0$) in order to emit an optical phonon.

### 4.2.2 Inter-Valley Scattering

Scattering of an electronic state from one valley to another is called inter-valley scattering and involves two valleys that are either equivalent or non-equivalent in energy (Figure 4.4). If both valleys are equivalent in energy, the scattering process requires a short-wavelength phonon to satisfy momentum conversation and an umklapp process may be necessary

$$k_2 + k' = k_1 + k \pm (q_{12} + q) + K$$

(4.23)
where $k'$ and $k$ are vectors relative to the valley extrema situated at $k_2$ and $k_1$, $q_{12}$ is a phonon wavevector such that

$$k_2 = k_1 \pm q_{12} + K,$$  \hspace{1cm} (4.24)

and $K$ is a reciprocal lattice vector. The magnitudes of the wavevectors of both acoustic and optical phonons are an appreciable fraction of the Brillouin-zone-boundary wavevector, and because of the dispersion relation, their frequency dependence can be neglected. Once eqn. (4.24) is satisfied, equivalent-inter-valley scattering proceeds exactly like intra-valley optical phonon scattering.

Scattering between non-equivalent valleys, i.e. from $\Gamma$ to $X$, $\Gamma$ to $L$ or $X$ to $L$, involves zone-edge phonons and can thus be treated like inter-valley scattering between equivalent valleys. Non-equivalent valleys at the same point in the zone demand a treatment similar to intra-valley scattering since only long-wavelength modes are involved and hence require a separate approach for acoustic and optical phonons. In the treatment of both types of non-equivalent-valley scattering, the difference of energy between the extrema as well as the difference of the effective mass have to be accounted for.

Inter-valley scattering, as it has been described so far, exists only for electrons in the conduction band. In Si, inter-valley scattering occurs exclusively between the six equivalent $X$-valleys and in Ge between the eight equivalent $L$-valleys. Inter-valley scattering between non-equivalent valleys cannot take place because the energy differences between the $\Gamma$, $X$- and $L$-valleys are too large. In GaAs, the absorption of an high energy zone-edge phonon can provide an electron with the momentum difference necessary to scatter from a low mass $\Gamma$-valley to a high mass $L$-valley.

Due to the different structure of the valence band, the idea of intra- and inter-valley scattering has to be modified for holes to the more appropriate model of intra- and inter-subband scattering to describe the scattering between the various hole bands. Processes involving inter-subband scattering are very strong on account of the
degeneracy of the HH and LH bands at the zone centre and hence are important at both low and high electric fields. For most semiconductors, the SO band is separated from the HH and LH bands by an energy of several hundred meV's so that scattering into the SO band does not occur until high electric fields. In Si, however, the SO band is only 44 meV below the top of the valence band, an energy difference small enough to allow for inter-subband scattering between the SO band and the HH or LH bands. The ability of the holes in Si to scatter to and from the SO band is one of the factors contributing to the comparatively poor hole mobilities in Si.

4.2.3 ACOUSTIC PHONON SCATTERING

Energy and momentum conservation restricts intra-valley scattering by acoustic phonons to long-wavelength modes which cannot change the energy of an electron other than through the elastic strain associated with them. As neighbouring unit cells all move by almost the same amount, the electronic energy perturbation is not related directly to the long-wavelength acoustic displacement but rather to the differential displacement or strain according to

\[ H_\varphi = \Xi \frac{\partial u}{\partial r} \]  

(4.25)

where \( u \) is the displacement and \( \Xi \) the deformation potential for the particular valley of interest. This type of interaction is therefore fully described in terms of deformation potentials which represent the energy shift per unit elastic strain. If \( R(x_1, x_2, x_3) \) is the position vector of a unit cell, the strain tensor is given by

\[ S_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) = S_{ji}, \]  

(4.26)

and hence, the general interaction between electrons and acoustic phonons can be defined by the Hamiltonian

\[ H_\varphi = \sum_{ij} \Xi_{ij} S_{ij} \]  

(4.27)

where \( \Xi_{ij} \) is the deformation potential tensor. Due to the symmetry of the strain tensor, the deformation potential consists of six components that are related to a system of axes which coincide with the principal axes of the \( \Gamma \)-, \( X \)- and \( L \)-valleys. For cubic crystals, the number of independent components of the deformation potential is halved to three, and the symmetry of the given spheroidal valleys reduces this number further to two. In particular, the two independent deformation potential components are \( \Xi_d \) which is related to a pure dilation and \( \Xi_s \) which is associated with a pure shear involving a uniaxial stretch along the major axis together with a symmetrical compression along the minor axis. Hence, shear strains produce no energy change in the case of intra-valley scattering in a \( \Gamma \)-valley. All six components
of the deformation potential can be expressed in terms of the two independent components $\Xi_d$ and $\Xi_u$, as shown in Table 4.1, where the first three components represent the strain along the principal axes of the $\Gamma$- and the $X$-valleys and the last three components characterize the strain along the principal axes of the $L$-valley.

<table>
<thead>
<tr>
<th>valley</th>
<th>$\Gamma$</th>
<th>$L$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi_{11}$</td>
<td>$\Xi_{dl} + \frac{1}{3} \Xi_{ul}$</td>
<td>$\Xi_{dx} + \Xi_{ux}$</td>
<td></td>
</tr>
<tr>
<td>$\Xi_{22}$</td>
<td>$\Xi_{dl} + \frac{1}{3} \Xi_{ul}$</td>
<td>$\Xi_{dx}$</td>
<td></td>
</tr>
<tr>
<td>$\Xi_{33}$</td>
<td>$\Xi_{dl} + \frac{1}{3} \Xi_{ul}$</td>
<td>$\Xi_{dx}$</td>
<td></td>
</tr>
<tr>
<td>$\Xi_{23}$</td>
<td>0</td>
<td>$\frac{1}{3} \Xi_{ul}$</td>
<td>0</td>
</tr>
<tr>
<td>$\Xi_{31}$</td>
<td>0</td>
<td>$\frac{1}{3} \Xi_{ul}$</td>
<td>0</td>
</tr>
<tr>
<td>$\Xi_{12}$</td>
<td>0</td>
<td>$\frac{1}{3} \Xi_{ul}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.1 Deformation potential components for $\Gamma$-, $X$- and $L$-valleys. Whereas the first three components coincide with the $x$-, $y$- and $z$-axis (according to the $\Gamma$- and $X$-valleys), the last three components are along and in the plane normal to the (111) axis (according to the $L$-valleys) [36].

In order to derive an expression for the acoustic phonon scattering rate, the $q$-dependence of the phonon frequency and occupation number has to be determined. For long-wavelength acoustic phonons, the phonon dispersion relation is simply $\omega_q = v_q q$, and since the low energy phonons dominate the scattering process, their occupation number in a certain mode $\omega_q$ is given by the Bose-Einstein distribution of eqn. (4.17). If the for the scattering processes relevant typical acoustic phonon energy is much smaller than the thermal energy, the acoustic phonon occupancy can be approximated to

$$n(\omega_q) = \frac{k_B T}{\hbar \omega_q} \quad \text{for} \quad \frac{\hbar \omega_q}{k_B T} \ll 1. \quad (4.28)$$

As energy and momentum conversation limit $q$ to $2k$, the typical energy of an acoustic phonon is given by $\hbar v_k$. For electrons obeying non-degenerate statistics at thermal equilibrium and for a parabolic band, the energy can be written as

$$\frac{\hbar^2 k^2}{2m^*} = \frac{3}{2} k_B T \quad (4.29)$$

and thus

$$\frac{\hbar \omega_q}{k_B T} = \left( \frac{3 m^* v^2}{k_B T} \right)^{1/2}. \quad (4.30)$$
With respect to the scattering of thermal electrons in Si, \( n(\omega_q) >> 1 \), and equipartition can be assumed to be maintained for all temperatures above 10 K because \( 3m^* v_s^2/k_B = 3K \) is valid in most cases. Regarding the equipartition regime for thermal holes, the critical temperature is slightly higher depending on the Ge concentration and the effective mass and can be as high as 15 K. For hot electrons and holes at low temperatures, this assumption may break down in which case \( n(\omega_q) \ll 1 \) is usually assumed. The acoustic phonon scattering rate for a spherical band in equipartition is

\[
W(k) = \frac{\Xi^2 k_B T}{8\pi^2 \hbar v_{sl}^2} \delta_{k+q-k',0} \delta(E_{k'} - E_k + \hbar \omega_q) d\mathbf{k}'
\]  

(4.31)

where \( \rho \) is the mass density, \( v_{sl} \) is the sound velocity of longitudinal modes and the upper sign is to be taken for absorption and the lower for emission. The total scattering rate is the sum of absorption and emission rates which are nearly identical for acoustic phonon scattering in equipartition. Assuming elastic scattering and hence neglecting the phonon energy \( \hbar \omega_q \) in the energy delta function, the total scattering rate is given by

\[
W_{tot}(k) = \frac{2\pi \Xi^2 k_B T N(E_k)}{\hbar v_{sl}^2}
\]  

(4.32)

where \( N(E_k) \) is the density of a given spin per unit energy in the band which can be written as

\[
N(E_k) = \frac{(2m^*)^{3/2} E_k^{1/2}}{4\pi^2 \hbar^2}.
\]  

(4.33)

As expected, the acoustic phonon scattering rate has a linear dependency on the temperature which can be ascribed to the assumption made about the phonon energy in eqn. (4.28) and which is only valid when the thermal energy is greater than the acoustic phonon energies.

A much more complicated case regarding the acoustic phonon scattering rate arises for the spheroidal parabolic bands of the L- or X-valleys due to their division into longitudinal and transverse components. Accordingly, the scattering rate in equipartition can be described for the longitudinal and transverse modes by

\[
W_{L,T}(k) = \frac{k_B T}{8\pi^2 \hbar v_{sl}^2} \int \Xi_{L,T}^2(\theta_q) \delta_{k+q-k',0} \delta(E_{k'} - E_k + \hbar \omega_q) d\mathbf{k}'.
\]  

(4.34)

Here, the definition of the effective deformation potential for longitudinal modes is given by

\[
\Xi_L(\theta_q) = \Xi_a + \Xi_u \cos^2 \theta_q
\]  

(4.35)

and for the sum of both branches of the transverse modes by
where $\theta_q$ is the angle between $q$ and the principle axis of the spheroidal valley. Since $E_{k'}$ varies with direction in the spheroidal band for a given magnitude of $k'$, the delta function is strongly dependent on the angle. The total scattering rate can be expressed in terms of a longitudinal and a transverse rate as

$$W_{L, T}(k) = \frac{2\pi \Xi_i^2 k_T N(E_k)}{h \rho \nu_{st}}$$

where $\Xi_i^2$ and $\Xi_d^2$ both are an intricate combination of $\Xi_i$ and $\Xi_d$, and $N(E_k)$ is given by eqn. (4.33) with a total effective mass of $m^* = (m_i^* m_d^2)^{1/3}$ [36].

Due to the warped nature of the degenerate valence bands which results in the density of states being highly anisotropic, the scattering of holes is far more complex and strongly dependent not only on the polar angle $\theta$ but also on the azimuthal angle $\phi$. Hence, it is extremely difficult to calculate the scattering rate for holes analytically, and numerical techniques are needed to obtain accurate results for lattice as well as impurity scattering. However, reasonable results can be achieved if isotropic bands are used that have the same DOS as the warped bands. Three independent deformation potentials $a$, $b$ and $d$ are associated with the acoustic strains in the valence bands where $a$ corresponds to pure dilation and $b$ and $d$ to uniaxial strain.

### 4.2.4 OPTICAL PHONON SCATTERING

Because of the relatively low frequencies of the phonons involved, acoustic phonon scattering is assumed to be elastic within a good approximation. For optical phonon scattering, this assumption cannot be made since the zone centre optical phonons have energies which are several tens of meV’s. The electron-phonon interaction Hamiltonian takes the form

$$H_{ep} = D_0 u$$

where $D_0$ is the optical deformation potential constant and $u$ is the relative displacement of the two atoms in the unit cell. As in the case of acoustic phonon scattering, energy and momentum conversation restricts intra-valley scattering by optical phonons to long-wavelength modes where the optical displacement may effect the electronic energy directly. However, the interaction in eqn. (4.38) is strictly valid only for the diamond lattice in which the two atoms in the unit cell have the same mass. On the other hand, optical phonon scattering is an important process only in diamond lattice structures, whereas in zinc-blende structures, polar optical phonon scattering is the dominant scattering process. As the optical phonon frequencies near the zone centre have very little dispersion, they can be assumed constant with no $q$ dependence. In a diamond structure, the transverse and longitudinal phonons have the same frequency at the zone centre, which is not the
case for the zinc-blende structures, but the differences are quite small and can be neglected. With this dispersionless \( \omega \) approximation, the energy conserving delta function for the scattering rate becomes independent of the scattering angle, which leaves a simple integration over the final energy states. Hence, the scattering rate can be written as

\[
W(k) = \frac{\pi D_0^2}{\rho \omega_0} \left[ n(\omega_0)N(E_k + \hbar \omega_0) + (n(\omega_0) + 1)N(E_k - \hbar \omega_0) \right]_{\text{absorption}}^{\text{emission}}
\]

where \( N(E_k) \) is identical to the density of states for electrons in a spheroidal parabolic band. As in the case for acoustic phonon scattering, the density of states for holes in the valence band has to be determined by performing a numerical integration of the warped valence bands. There is only one deformation potential \( d_o \) associated with the optical strains in the valence bands which is usually defined as an energy rather than an energy per unit displacement as is the case for electrons.

### 4.3 Impurity Scattering

An altogether different scattering process, impurity scattering is caused by a change in potential introduced by an impurity in the lattice. Impurity sites are created either by background doping or by point defects such as missing ions or an occasional ion in the wrong place. Unlike lattice scattering which is caused by vibrational states spread uniformly over the crystal volume, the interaction between an electronic state and an impurity has a local character. Whereas scattering events involving phonons take place in all unit cells, impurity scattering will occur with reasonable probability only in the vicinity of the impurity site. Hence, the carriers can be seen as wave packets, partially localized in the proximity of the scattering centre, and the scattering event can be described in terms of either a scattering cross section or a scattering rate. However, since the scattering potential of an individual impurity has infinite range and there are always many impurity sites present in the crystal, an electron or hole has to be considered as being scattered continually, changing the properties of the carriers from those of the Bloch wave particle. Due to the involvement of a large change in crystal momentum, inter-valley scattering by an impurity site is an extremely weak process. Therefore, the impurity scattering process is assumed to always be an intra-valley scattering event.

**Charged Impurity Scattering**

Scattering by charged or ionized impurities is governed by the same physical principle as Rutherford scattering where the trajectory of an electron or hole is diverted by the Coulomb potential of an ionized impurity site. The coulombic interaction of two charged particles is inversely proportional to the distance between
the interacting particles and it is strongest for a long interaction time. A long interaction time is given for slowly moving electrons and holes as is the case for a non-degenerate carrier gas at low temperatures. The low-temperature mobility is therefore a measure of the impurity and defect content in non-degenerate doped semiconductors. Due to the infinite range of the scattering potential of each impurity, the scattering cross section is effectively infinite, and there are a number of competing techniques to overcome this problem. Calculating the cross section of this kind of elastic scattering event is further complicated by the fact that the interacting particle is surrounded by other carriers, making the scattering event a many-body process. As a result, the impurity does not only interact with the particle under observation but also with the surrounding particles, thus giving rise to screening. In general, screening can be seen as a dielectric response of the system to the introduction of the charged impurity which screens the impurity from passing carriers and thereby reduces the scattering rate substantially.

A special signature of the ionized impurity scattering process is its characteristic temperature dependence given by $\mu \sim T^{3/2}$. This increase of the mobility with rising temperature can be explained by the fact that the electrons travel at a higher speed at higher temperatures and thus are less affected by the ionized impurities. In contrast, the mobility limited by phonon scattering decreases with increasing temperature, and since the impurity concentration does not change significantly with temperature, the ionized impurity scattering process becomes dominant at lower temperatures where lattice and electron-electron scattering become increasingly weak.

**Neutral Impurity Scattering**

At low temperatures, non-ionized shallow impurities can act as scattering centres by quasi binding a corresponding carrier to form a charged ion. Because there is no Coulomb potential involved in this type of interaction, scattering by neutral impurities is much weaker than scattering by charged impurities. As for charged impurity scattering, a term for screening has to be included in the calculation of the cross section, again reducing the scattering rate significantly.

**Alloy Scattering**

The average potential and band structure of an alloy of the form $A_xB_{1-x}$ is best described to the lowest order by the virtual crystal approximation in which an average periodic pseudo-potential is assumed corresponding to a uniform distribution of the components of the alloy over the lattice sites. In reality, the random dispersion of the ions introduces a disorder component, giving rise to potential fluctuations experienced on average by the free carriers and thus leading to elastic scattering. If $V_a$ and $V_b$ are the potentials associated with species A and B of the alloy, the average potential is

$$V_0 = V_ax + V_b(1-x). \tag{4.40}$$
In the region where the fractional occupancy of A is \( x' \), the deviation of the real potential from the virtual potential is represented within each unit cell by a highly localized potential

\[
V' - V_0 = (V_a - V_b)(x' - x).
\]

(4.41)

The interaction potential \( V_a - V_b \) has been variously interpreted as the difference in band gaps of the two species, the difference in electron affinity or the difference in band offsets for heterostructures made up of the two systems. However, elastic strain will also add a deformation-potential component, so the interaction or alloy potential is usually regarded as a quantity to be determined empirically. Like for charged and neutral impurity scattering, screening of the potential fluctuations results in a decreased cross section of the alloy scattering process. Due to the fact that most research into SiGe alloys is concentrated on strained layers in 2D systems, no systematic mobility investigations on bulk SiGe alloys have been done so far.

**ELECTRON-ELECTRON SCATTERING**

The fact that an electron is surrounded by many other electrons in the conduction band gives not only rise to screening, which limits the scattering rates involving all types of impurities, but also has an effect on the motion of the electron itself. Interactions of an electron with its surrounding electrons lead to collisions between identical particles for which the cross section can be expressed in terms of the relative momentum relaxation rate. For fast incident electrons, this relaxation rate can be calculated quite easily and is found to be inversely proportional to the carrier density. In order to compete with typical scattering times, the relative momentum relaxation rate has to be in the picosecond range, and therefore, the carrier density has to exceed \( 10^{17} \) cm\(^{-3} \). In non-degenerate semiconductors, scattering rates due to electron-electron collisions are usually negligible. Evidently, the principle of electron-electron scattering has to be applied correspondingly for the movement of holes in the valence band.

**4.4 SCATTERING IN QUANTUM WELLS**

Clearly, lattice and impurity scattering are not only present in bulk semiconductors but also influence the transport behaviour of the carriers in a 2D system to a greater of lesser extent. However, pseudomorphic growth and strain alter the crystal structure and thus the physical properties of such a system. Consequently, existing scattering mechanisms have to be modified and additional scattering events have to be considered for the calculation of the carrier mobilities in a 2D SiGe heterosystem.

Looking at the growth characteristics of heterostructures, certain scattering centres have to be dealt with that have their origin in the specific layer structure of quantum
wells and are not present in bulk material. Among those additional scattering mechanisms are scattering by interface roughness, by interface charge and by threading dislocations. Due to the modulation doping technique employed in the design of quantum wells, impurity scattering has to be divided into remote impurity and background impurity scattering. Alloy scattering only occurs in hole gas structures where the channel consists of SiGe. Lattice scattering has to be modified because of the strain in the QW and will be discussed in detail in chapter 11. When the carrier density is high enough for more than one subband to be occupied (generally at higher temperatures), inter-subband scattering is possible, but since the relaxation times are determined at low temperatures where usually only one subband is occupied, inter-subband scattering will not be considered further. It is useful to look at all of the different scattering mechanisms separately and determine their effect on the relaxation times and thus the mobilities individually.

Assuming the disorder in a 2D carrier gas is characterized by the scattering potential \( U_s \), the lowest order results for the transport and quantum relaxation times are given by the following expressions (with \( h = 1 \) for simplification) [95, 99, 100]

\[
\frac{1}{\tau_i} = \frac{1}{2\pi E_F} \int_0^{2k_F} \frac{q^2}{(4k_F^2 - q^2)^{1/2}} \frac{U_s^2}{e_q^2} dq, \tag{4.42}
\]

\[
\frac{1}{\tau_q} = \frac{1}{2\pi E_F} \int_0^{2k_F} \frac{2k_F^2}{(4k_F^2 - q^2)^{1/2}} \frac{U_s^2}{e_q^2} dq. \tag{4.43}
\]

Here, \( e_q \) is the static dielectric function, \( E_F \) is the Fermi energy and \( k_F \) is the Fermi wavevector. The scattering potential depends on the scattering type and the shape of the confining potential. Because of their simplicity, these two equations are widely used in the interpretation of experimental data. If multiple scattering effects are taken into account, the relaxation times are reduced in comparison to the lowest-order calculation. Defining the appropriate scattering potentials of the different scattering types allows a separate analysis of each scattering process.

**Acoustic Phonon Scattering**

At temperatures low enough for the carrier distribution to be degenerate, Pauli exclusion becomes important in limiting the scattering that is allowed. In such a degenerate system, the occupation probability of the final state in a scattering process is not negligible anymore but has to be taken into account by assuming that a Fermi-Dirac distribution is maintained at any time. However, high carrier concentrations also permit the assumption that electron-electron interactions are strong enough for the distribution function to maintain the Fermi-Dirac form which is characterized by an electron temperature \( T_e \). This concept remains valid even for hot electrons which have a temperature above that of the lattice. As the Pauli exclusion leads to rapid variations with temperature, the scattering rates can be formulated in terms of electron temperature rather than electron energy.
REMOTE IMPURITY SCATTERING

Although the principal advantage of modulation-doped quantum wells is the enhancement of the mobilities due to spatial separation of the mobile carriers from their ionized parent impurities, the remote ionized impurities maintain their character of coulombic scattering centres. The Coulomb force exerted on the carriers in the QW by the remote impurities decreases with the square of the distance and linearly with the dopant concentration. Hence, the doping concentration in the dopant layer and the spacer layer thickness between the dopant layer and the QW are the most important parameters to control the strength of remote impurity scattering. A further important factor is the effect of screening which is strongly dependent on the carrier concentration itself. The higher the carrier concentration, the stronger the screening effect of the additional carriers, which explains the specific dependence of the mobility limited by remote impurities on the carrier concentration given by $\mu_{RI} \propto N_{\text{doping}}^{-1} n_s^{3/2}$.

BACKGROUND IMPURITY SCATTERING

The purpose of the modulation technique is to leave the quantum well free of intentional doping. Practically this is not feasible because of growth effects such as segregation of the dopant or Ge and incorporation of impurities from the growth chamber, all leading to unintentional doping in the channel. Although the density of impurities in the QW is generally low, their proximity to the electron or hole gas can result in a significant contribution of the background impurities to the scattering process. With increasing spacer layer thickness the remote impurities become less effective scattering centres, and eventually, background impurities become the dominant source for scattering. It is obvious that the higher the background impurity density, the smaller the spacer thickness at which background impurities become the dominant scattering centres. Due to screening, the momentum relaxation time for background impurity scattering has a similar dependency on the carrier density as it has for remote impurity scattering. Hence, the reduction of the carrier density with increasing spacer thickness results in a corresponding reduction in the mobility limited by background impurities which is given by $\mu_{BI} \propto N_{\text{background}}^{-1} n_s^{3/2}$.

INTERFACE ROUGHNESS SCATTERING

At least two distinct features can be isolated to cause interface roughness. The effectively random distribution of Ge atoms in the alloy means that one or two monolayers at the heterojunction should be considered as a transition region between the SiGe layer and the Si layer. On the other hand, inappropriate selection of the growth temperature and deposition rate can lead to growth by 3D islanding rather than by advancing terrace edges in 2D which leads to a non-planar interface. In both cases, the associated variation of the potential well thickness causes a perturbation of the eigenstate energies, thus giving rise to an effective scattering potential. The conventional method to characterize interface roughness scattering assumes the fluctuations along the interface to be random and correlated in a
Gaussian manner with average height $\Delta$ and correlation length $\Lambda$. However, such an approach is justified only if $\Delta$ is much smaller than the width of the well while $\Lambda$ is much larger than the width. The dependency of the mobility limited by interface roughness on the average height, the correlation length and the carrier density is given by $\mu_{ir} \propto (\Lambda\Delta)^{-2}n_s^{-1}$.

**INTERFACE CHARGE SCATTERING**

Some degree of impurity and contamination of layers during and after growth is inevitable, but it is thought that the nature of the growth environment is the most crucial aspect in determining impurity levels in the material. Many metals are known to be fast diffusers in Si while interfaces can act as efficient gettering sites, so impurities from the whole sample may potentially be trapped at the heterojunction. Furthermore, a dislocation network generated from the heterointerface upon relaxation of the strained layer would further enhance the trapping of impurities. When such impurity or intrinsic effects become charged, the carrier mobility will be limited by Coulomb scattering off these centres. An expression to describe interface charge scattering can be obtained using the same theory as for interface roughness scattering with an additional impurity form-factor.

**ALLOY SCATTERING**

SiGe alloys are almost exclusively used in the design of quantum well structures, and hence, scattering due to the fluctuations of the Coulomb potential in the SiGe alloy is restricted to heterostructures. Obviously, the two different types of quantum wells for electron and hole transport have to be distinguished and treated separately because of their opposite layer structure.

In the case of a 2DEG, the channel consists ideally of pure Si, and the electrons can therefore not be scattered by potential fluctuations in the alloy. Nevertheless, the Si channel is sandwiched between two layers of SiGe, and because the electron wavefunction penetrates at least one of those layers to a certain extent, a small part of the wavefunction will experience some alloy fluctuations. A second source of alloy scattering for electrons in a Si channel can be unintentional incorporation of a few percent Ge into the quantum well that can for instance originate from segregation during growth. Under these circumstances, the alloy scattering rate can increase dramatically.

The main problem of alloy scattering, however, occurs in the SiGe quantum well of 2DHG structures where the percentage of Ge has to be much larger and thus limits the hole mobility to a far greater extent. Following a relation according to $x(1-x)$, the scattering rate will increase with growing Ge concentration $x$ up to a Ge content of 50% in the alloy where it will reach a maximum. Yet, the corresponding decrease of the hole mobility will be at least partly offset by the reduction of the hole effective mass due to strain-induced band bending. Again, it is important to include an appropriate term for screening in the description of the scattering rate to characterize
the effect of alloy scattering on the hole mobility as accurately as possible, and the
mobility dependency is therefore given by $\mu_{\text{All}} \propto m^* n^2 (x(1 - x))^{-1} n_s^{-1}$.

**SCATTERING FROM THREADING DISLOCATIONS**

Each dislocation represents a row of atoms which have a missing bond with their
neighbours, resulting in deep levels in the bandgap of the semiconductor. The
dangling bond of each atom of a dislocation can accept an electron to lower its
energy and get negatively charged, thus forming a positively charged space charge
region. If the density of available trap sites of a dislocation is high enough to pin the
Fermi energy at mid-gap, an electron barrier is created around the dislocation
through band bending, and all free electrons in the channel will then be entirely
repelled. Since the spacing of the charged impurities on the dislocation line is on the
order of the lattice constant, the space charge usually extends over a micron or so,
wheras the radius of the cylindrical depletion region is of the order of the 2D Debye
length which is typically about a third of a micron. The mobility limited by threading
dislocations depends on the density of threading dislocations and the carrier density
according to $\mu_{\text{TID}} \propto N_{\text{thread}}^{-1} n_s^{-1/2}$.

Because nucleation of misfit dislocations and the subsequent propagation of
threading dislocation occur only in pseudomorphically grown strained layers thicker
than the metastable critical thickness, 2DHG structures are in general free of
dislocations by virtue of their layer design. In 2DHG structures, the sole layer grown
under stress is the SiGe channel, which is fully strained and thus does not exhibit any
dislocations. However, since the channel width is usually between the equilibrium
and the metastable critical thickness, partial relaxation of the metastable strain can
eventually take place by creation of misfit and threading dislocations. As the
threading dislocations will propagate across the active region of the structure, the
quality of the 2DHG will then become very poor. However, for appropriate growth
conditions and moderate thermal processing, strain relaxation can be avoided
completely in these structures.

On the other hand, 2DEG structures employ a thick and fully relaxed SiGe buffer
layer for the subsequent fully strained Si channel. As a consequence, misfit
dislocations nucleate at the interface between buffer layer and substrate, terminated
by threading dislocations propagating towards the top of the buffer layer. When the
distance between dislocations is significantly longer than the electron mean free path,
the electrons can move freely in the region between dislocations. Once the density
increases to the extent that there is an overlap between the depletion regions of
adjacent dislocations, the electrons are forced to scatter off the dislocations, and the
mobility decreases dramatically. Assuming the Debye length to be in the order of a
third of a micron, the overlap is expected to take place when the density of threading
dislocations is about $10^6$ cm$^{-2}$ which has to be compared to a density of over $10^{11}$ cm$^{-2}$
for single-step buffers. The use of graded buffers can reduce the density of threading
dislocations by five to six orders of magnitude into the $10^5$ cm$^{-2}$ range where this
scattering process is entirely negligible.
4.5 SCATTERING IN 2DEG STRUCTURES

When investigating 2D structures with regard to the determination of the dominant scattering mechanism at low temperatures, one of the most powerful techniques to distinguish between the dominance of short-range and long-range scatterers is the evaluation of the ratio $\tau_s/\tau_q$ by measuring the transport and quantum relaxation times. In $n$-type structures with graded buffers, the this ratio is usually greater than ten, indicating that the electron mobility in the QW is predominantly limited by long-range scattering. Although the use of grading techniques allows for buffer growth with threading dislocation densities below the critical limit at which they restrict the electron mobility, two further conditions have to be fulfilled in order to avoid misfit dislocations becoming the dominant scattering mechanism. As misfit dislocations are generated during relaxation of the graded buffer, a fully relaxed buffer layer of a certain minimum thickness (~0.5 μm) has to be grown on top to eliminate any influence of the existing misfit dislocations in the graded buffer on the electrons in the QW. Another source for the formation of misfit dislocations is the strain in the active Si channel. If the layer thickness of the strained Si channel exceeds the equilibrium critical thickness, the threading dislocations in the graded buffer will start to glide and form misfit dislocations at the bottom interface of the Si channel. Due to their proximity, these misfit dislocations induce a very strong scattering potential on the electron gas. Figure 4.5 illustrates the effect of the grading rate of the Ge content in the buffer layer on the density of the threading dislocations and the electron mobility. Clearly, dislocation densities well below $10^9$ cm$^{-2}$ can be realized by growing the buffer layer with an increasingly shallow gradient, but electron mobilities do not improve any further for densities below $10^7$ cm$^{-2}$, suggesting other scattering mechanisms to become dominant for these high mobility 2DEG structures.

![Figure 4.5 Density of threading dislocations and electron mobility (measured at 0.4 K), each as a function of the grading rate of the Ge content in the buffer layer which is ramped to a final Ge concentration of 30% [101].](image-url)
Indeed, once the density of threading dislocations is in the order of $10^6 \text{ cm}^2$ and below, electron mobilities are limited by remote ionized impurities rather than by the quality of the buffer layer. Minimizing the scattering potential of remote impurities is a complex process as the influence of their Coulomb force on the electron gas is dependent on the doping concentration in the supply layer, the thickness of the spacer layer and the carrier density in the QW. Hence, tradeoffs are inevitable when adjusting those parameters. Usually, the doping concentration is already chosen to be at a minimum to avoid parallel conduction in the supply layer, and a further reduction in the doping concentration will result in a reduced carrier density which in turn will decrease the effect of screening. Increasing the spacer thickness has a similar effect on the electron mobility as it reduces the scattering potential by increasing the distance between the remote impurities and the electrons in the QW. However, at the same time, it also reduces the carrier density and thus the effective screening. The relation between the carrier density, the spacer thickness and the electron mobility is shown in Figure 4.6. In the graph on the left side, the electron density and mobility are both calculated as a function of the spacer thickness, and as expected, the density decreases with increasing spacer thickness. Within the limits of the spacer thickness for which the mobility is calculated, the mobility increases monotonically with the width of the spacer layer. Additionally, different background doping levels are taken into consideration to investigate their effect on the carrier density and mobility. A higher concentration of background impurities results in a reduced carrier density for the same spacer thickness, and the much smaller gradient of the mobility curve for a higher background doping level suggests that background impurity scattering becomes more important with decreasing screening.

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{figure4.6.png}
\caption{Calculated electron mobilities and densities versus spacer layer thickness at 1.5 K for a 30\% Ge buffer layer and a conduction band offset of 180 meV (left) \cite{102}. The doping concentration in the supply layer is $N_D = 2 \times 10^{18} \text{ cm}^3$ and two background doping levels of $N_A = 10^{14} \text{ cm}^3$ (solid curves) and $N_A = 10^{15} \text{ cm}^3$ (dashed curves) are chosen. Measured electron mobility versus sheet density at 0.4 K (right) \cite{103}. For the upper curves, the wavefunction is increasingly pulled away from the top interface by biasing the back gate, effectively increasing the thickness of the spacer layer ($l_0$ is the distance between the top interface and the wavefunction).}
\end{figure}

To investigate the direct dependence between carrier density and mobility, the electron mobility is plotted as a function of the carrier density in the graph on the right side of Figure 4.6. Here, the carrier density is varied solely by means of front-
gate biasing, showing clearly that an increase in screening due to higher carrier densities (without the corresponding increase of the remote impurity concentration or decrease of the spacer layer thickness) improves the electron mobility significantly. Obviously, this is of great benefit regarding device applications where a high carrier density is required. A shift of the wavefunction towards the centre of the channel can be achieved by back-gate biasing, resulting in an increased effective spacer layer thickness without compromising the sheet carrier density. Pulling the wavefunction away from the top interface towards the centre of the channel also reduces the scattering from interface roughness and interface charge. However, with modern growth techniques, the quality of the interface can be grown to such a high standard that none of these scattering mechanisms is expected to limit the mobility in these high-quality and high-mobility 2DEG structures. On the other hand, alloy scattering from the top interface cannot be reduced by improving the quality of the interface, which therefore maintains its mobility limiting character if parts of the wavefunction penetrate into the SiGe alloy spacer layer. Consequently, by moving the wavefunction away from the top interface, less and less of the wavefunction penetrates the alloy layer, thus decreasing this type of scattering process. At room temperature, the electron mobility is predominantly limited by phonon scattering and the measured Hall mobility might be influenced by a potential low-mobility parallel conduction channel in the supply layer or in populated higher subbands.

4.6 SCATTERING IN 2DHG STRUCTURES

Measurements of the transport and quantum relaxation times at low temperatures in p-type structures generally yield a $\tau_r/\tau_q$ ratio close to unity, suggesting large-angle scattering to be dominant in such structures. Consequently, scattering by remote impurities is not considered to be the dominant scattering mechanism that limits the hole mobility at low temperatures. Theoretical calculations show that remote impurity scattering only starts to limit the hole transport for mobilities in excess of $5 \times 10^5 \, \text{cm}^2/\text{Vs}$. One of the possible short-range scattering mechanisms that could be responsible for the limitation of the hole mobility at low temperatures is alloy scattering which is dependent on the Ge content $x$ in the SiGe channel. To understand the influence of the composition in the channel on the limitation of the hole mobility, the increase of alloy scattering with increasing Ge concentration has to be weighed against the concurrent decrease of the hole effective mass $m^*$. This leads to a complex dependence in which alloy scattering is a function of the effective mass, the carrier density and the Ge content. On the other hand, the effective mass itself depends on $x$ and $n_s$ since strain reduces $m^*$ as $x$ increases, whereas the pronounced nonparabolicity of the HH band leads to an increase of $m^*$ as $n_s$ is raised. These effects are coupled since the higher valence band offsets achievable with greater $x$ are usually exploited for enhancing $n_s$ and hence, the strain induced mass reduction is partly compensated by nonparabolicity effects. Overall, a monotonic decrease of the mobility limited by alloy scattering can be seen as $x$ increases.
A matter of controversy is the determination of the alloy scattering potential which has been measured to lie between 0.2 eV and 1.0 eV [104-111]. Depending on which of the two limits are chosen for the scattering potential, alloy scattering should not be (lower limit) or could be (upper limit) the dominant scattering mechanism at low temperatures. In Figure 4.7, calculated mobility curves for both limits are plotted against the Ge concentration in the channel where in a first approximation, a constant effective mass $m^* = 0.3 m_0$ is assumed, and $\mu_{\text{alloy}}$ is calculated for carrier densities $n_i = 2 \times 10^{11} \text{ cm}^{-2}$ (dashed line, sensible for low $x$) and $n_i = 10^{12} \text{ cm}^{-2}$ (dotted line, reasonable for $x > 0.3$). In order to achieve a more realistic picture, the variations of the effective mass as a function of both $x$ and $n_i$ is included in the calculations as well as the maximum possible increase of $n_i$ with $x$. The increase of the mobility with decreasing $x$ is only of limited benefit with regard to device applications where high carrier densities are advantageous as the strongly reduced valence band offsets at lower Ge concentrations restrict the 2D carrier densities to the low $10^{11} \text{ cm}^{-2}$ range.

Although screening is very effective at low temperatures, the effect of screening has been neglected in these calculations due to the short-range nature of alloy scattering, leading to a possible overestimation of the effective scattering potential. However, due to its temperature independence, alloy scattering is not regarded to be a potential dominant scattering mechanism at room temperature where phonon scattering is the most plausible scattering process to limit the hole mobility. Another problem during Hall measurements at room temperature may arise from the presence of parasitic parallel low-mobility conduction channels in the supply layer or populated higher subbands (as for $n$-type structures), which results in a significant underestimation of the true mobility in the QW.

![Figure 4.7](image-url)

*Figure 4.7* Hole mobilities limited by alloy scattering as a function of the Ge content for an alloy scattering potential of 0.28 eV (top curve) and 0.74 eV (bottom curves) at low temperature [28]. A constant effective mass and a constant carrier density is assumed for the two broken lines, whereas their full dependency on $n_i$ and $x$ is accounted for by the solid lines. Experimentally acquired data points for the mobility are shown for comparison to the theoretical curves [112].
Other possible scattering mechanisms limiting the hole mobility at low temperatures are interface roughness and interface charge scattering. Whereas the strength of interface charge scattering decreases with increasing carrier density, that of interface roughness increases with carrier density. Assuming alloy scattering to be negligible when screening is included in the calculations, the hole mobility is hence limited by interface charge at lower sheet densities (low $10^{11}$ cm$^{-2}$) and by interface roughness at higher hole densities (high $10^{11}$ cm$^{-2}$) [112-114]. Associated with the rough nature of the interface are local variations in strain within the channel which can act as an additional source of scattering through coupling to the deformation potential. The influence of all the major scattering mechanisms on the hole mobility is shown in Figure 4.8 in which the total mobility is calculated according to Matthiessen's rule.

The demand for high sheet densities in device applications would suggest interface roughness scattering to be dominant in those structures which thus requires preferential treatment to minimize its detrimental effect on the mobility. However, the analysis of the scattering rates via their mathematical function remains ambiguous due to their dependence on poorly understood fitting parameters. Therefore, another method can be used to complement the measurements of the density dependence by employing a back-gate bias to move the wavefunction away from the interface. Using this approach, Coulomb scattering by the impurities in the spacer layer and/or at the interface are identified to limit the mobility for sheet densities in the range of $2 - 5 \times 10^{11}$ cm$^{-2}$ [115].

![Figure 4.8](image_url)

**Figure 4.8** Calculated mobility restrictions of the most important scattering mechanisms at low temperatures and the resulting maximum overall hole mobility [114].

A third method to distinguish more clearly between the effect of interface charge and interface roughness is by investigating an inverted 2DHG structure where the doping layer is located between the substrate and the SiGe channel, which enables the wavefunction to be established at the lower, smoother interface. Problems arise in this structure due to surface segregation of boron during growth and the much
increased density of background impurities which in this specific case seem to limit
the mobility rather than the quality of the interface. In normally grown p-type
structures, however, background impurities are in general found to have no limiting
effect on the hole mobility.

Unlike in 2DEG structures, where the mobility of electrons in high-quality materials
at low temperatures is effectively limited by the Coulomb potential scattering of
remote ionized impurities, the equivalent dominant scattering mechanism in p-type
structures cannot be specified in a similarly general way. Due to the more complex
interdependence of the material properties for a changing alloy composition in the
channel and stronger interactions of the existing scattering centres with the holes, it
is not possible to clearly identify a single scattering process to be dominant in a
2DHG structure of arbitrary Ge composition at low temperatures. Depending on the
specific choice for the value of the numerous parameters involved, \textit{ab initio}
calculations tend to suggest different scattering sources to be dominant when fitted
to the experimentally acquired data.
5 Band Structure Modelling

Providing the general basis for enhancing the performance of semiconductor devices, the concept of band structure engineering constitutes the most essential technique to improve the electrical and optical properties in modulation-doped heterostructures. When designing transport devices, it is vital to know and understand the influence of the band structure on each individual electrical parameter since the improvement of the transport behaviour relies ultimately on a carefully chosen compromise regarding the optimisation of these parameters. Depending on the chosen aspect of the device performance that should be enhanced, the band structure can then be tailored accordingly to optimize the corresponding parameters. Within this context, band structure calculations can be used prior to growth in order to optimize and also analyze heterostructures before they are grown. Together with the feedback of the qualitative and quantitative results from the experimental analysis of these structures after they are grown, band structure modelling is a powerful tool that reduces the risk of growing heterostructures of inferior quality and can considerably shorten the time needed to design a 2D system with the required characteristics.

When using band structure calculations to improve the transport behaviour of 2D heterostructures, there are four principal aspects that should be considered optimizing, i.e. the sheet density, the parallel conduction in the supply layer, the effective mass and the position of the wavefunction. Concerning the design of room-temperature devices, for example, a reduction of the doping concentration is a crucial requirement to suppress parallel conduction in the supply layer but it also has to be weighted against the need for a minimum sheet density to sustain electronic transport in the channel. In p-type structures, the value of the hole effective mass can be changed by varying the Ge composition and the wavefunction can be moved by grading the Ge content. The position of the wavefunction in n-type structures is of less significance because scattering is dominated by remote impurities and the electron effective mass remains constant when the structure is altered. For this reason, band structure calculations are only performed for the valence band of 2DHG structures. It should be noted that the process of optimizing individual parameters might occasionally lead to heterostructures of inferior quality inadvertently caused by an increase in scattering due to unfavourable growth conditions.

5.1 Self-Consistent Poisson-Schrödinger Solver

The presence of electronic charge in a semiconductor, as represented by the carriers and the dopants, has an effect on the confinement potential through the associated electrostatic potential which complicates the band structure calculations significantly.
because the Schrödinger and the Poisson equation have to be solved simultaneously. This is due to the fact that the eigenstates of the Hamiltonian are determined by the confining potential which in turn is affected by the eigenvalues (their position with respect to the Fermi energy) via the Poisson equation. This problem can be resolved by employing a self-consistent procedure in which an iterative approach is used to obtain solutions for both equations concurrently. Starting with an initial guess for the confining potential, the procedure subsequently changes the potential until it has converged to the form in which the charge distribution as given by both equations is identical. Such a routine is often referred to as a self-consistent Poisson-Schrödinger (SCPS) solver and can provide information about the band structure, the subband populations and the corresponding wavefunctions.

For all calculations of the valence band parameters, an in-house SCPS programme has been used which describes continuous functions such as potentials, envelope functions and doping profiles at a finite number of discrete mesh points. This process transforms the differential equations into a finite difference form and the Poisson and the Schrödinger equation consequently take the form of a matrix equation. In order to reduce the numerically demanding calculations associated with solving the Schrödinger equation, the much-simplified $k$-$p$ approach, or effective mass theory, is employed after some modifications have been made to include strain effects. Although the principal aim is to model compositionally varying structures with respect to quantities like the confinement potential, envelope functions and doping distributions, it is also possible to calculate the mobility of the holes if an adequate scattering model which describes the scattering processes in the heterostructure fairly accurately is used in combination with the SCSP programme. However, implementing the appropriate calculations for the scattering rates of a $p$-type SiGe structure is rather complicated because the system is dominated by short-range scattering and the dominant scattering mechanism is also dependent on the structural design of the system. The description of a suitable scattering model is thus very complex, especially if phonon scattering is to be included at higher temperatures, which ultimately renders the implementation of such an intricate scattering model uneconomical and the calculation of the hole mobilities is therefore not included in this programme.

The iterative approach of the SCPS programme, in which successively better solutions for both equations are obtained, initially assumes an arbitrary confinement potential which includes the corresponding electrostatic component and also the local exchange and correlation potential which represents the many-body correction to the one-electron Schrödinger equation for high-density systems. The Hamiltonian is then solved for the hole eigenfunctions, distributions and densities through the effective mass equation which in turn supplies information about the hole eigenvalues, wavefunctions and dispersion relations. Subsequently, the true electrostatic contribution corresponding to these hole densities and distributions is obtained by solving the Poisson equation thereby defining a new confinement potential. To facilitate the corresponding calculations, the hole density distribution is split up into a quantum region close to the heterointerface and a bulk region sufficiently far away from the quantum well. The resulting potential function also
has to satisfy two boundary conditions to guarantee it represents a unique solution. One boundary condition is given by the assumption that the substrate remains in a charge neutral state, effectively pinning the Fermi level at mid gap at the semiconductor-substrate interface. The other boundary condition follows from an experimental observation of the Fermi-level pinning at the air-semiconductor surface. A simultaneous solution for both equations is obtained if both confinement potentials are equal within a very small margin of error. Otherwise, the programme proceeds by constructing a confinement potential which is expected to be closer to the exact solution and the whole procedure is repeated until the solution has converged to a point where it satisfies both the Poisson and the Schrödinger equation. The actual application of the SCSP solver is fairly straightforward and only requires the input of a few experimental parameters. Apart from supplying the appropriate material constants such as band gaps, mass parameters and deformation potentials, the programme simply requires an accurate description of the structural composition of the heterostructure.

5.2 **Band Structure of 2DHG with Graded QW**

Since the low-temperature transport properties of \( n \)-type SiGe heterostructures are predominantly improved by reductions in ionized remote impurity scattering and enhancements in the quality of the virtual substrate, the present band structure calculations are concentrated on the valence band of \( p \)-type SiGe systems. As part of the general concept of band structure modelling, valence band calculations have been performed at room temperature to eliminate the parallel conduction path in the supply layer by varying the doping concentration, the doping profile and the width of spacer and supply layer. The same types of calculations have also been carried out in the low-temperature regime with the main intention of modulating the sheet density in the channel. Due to space restrictions, however, the results of these calculations cannot be presented in this chapter which instead focuses exclusively on the attempt to reduce the hole effective mass and vary the position of the wave-function. Principally, both objectives can be realized by varying the profile of the Ge composition in the quantum well. A reduction in \( m^* \) is essentially achieved by increasing the Ge content in the channel because the hole effective mass in pure Ge is lower than in pure Si and due to bigger strain. It should be noted that although the influence of strain on \( m^* \) is fully accounted for in the SCPS programme, its effect on the deformation potential has been disregarded and the value of the deformation potential in the channel is simply obtained by a linear interpolation between the bulk values for Si and Ge. In addition, the wavefunction can be moved away from the upper interface by grading the Ge content in the channel. This aspect of band structure modelling is of great importance because the upper interface is of much lower quality than the lower interface and therefore is the source of the dominant scattering mechanism. Whether the wavefunction is moved to the middle of the quantum well or near the lower interface
is only dependent on the grading profile. A linearly graded channel with a decreasing Ge content in growth direction pulls the wavefunction into the vicinity of the lower interface whereas a triangular profile shifts the envelope function to an area in the middle of the channel where the highest Ge content is located. However, the superiority of a profile over another with respect to its effect on the scattering rate can only be established experimentally by measuring the transport properties.

A few selected models are presented in this section to demonstrate the dependence of the wavefunction regarding its shape and position on the profile and the maximum Ge composition. Due to restrictions in the input file of the SCPS programme related to its discrete nature, the Ge content cannot be altered continuously but has to be graded in steps of constant composition. Figure 5.1 shows the band diagram of four step-graded quantum wells at 3 K, two with a linearly decreasing Ge concentration in growth direction and two with a triangular profile. Since the HH1 band is the only occupied hole band at low temperatures, the most important aspect of the diagrams is the shape and position of the envelope function of the HH1 band. Another point of interest is the separation between the HH1 band and the other subbands (HH2, LH1, LH2 and SO1) which among other things affects the nonparabolicity of the HH1 band and the hole effective mass.

![Band Structure Diagrams](image_url)

**Figure 5.1** Calculated valence band diagrams for step-graded 2DHG structures with corresponding envelope functions of the hole bands. Each QW consists of three steps with a 5% change in Ge content and has a total width of 15 nm. The structural composition of each QW is

- a) Ge %: 25/20/15, step size: 3 x 5 nm
- b) Ge %: 15/10/5, step size: 3 x 5 nm
- c) Ge %: 20/25/20, step size: 3 x 5 nm
- d) Ge %: 10/15/10, step size: 3 x 5 nm

112
As expected, a profile with a linearly decreasing Ge concentration in growth direction pulls the envelope function of the HH1 band close to the lower interface which should reduce scattering from the upper interface considerably. When comparing the wavefunction of the lowest subband in structures $a$ and $b$ of Figure 5.1, it can be seen that the position of the HH1 envelope function is marginally closer to the lower interface in structure $b$ and its shape is significantly less broadened which indicates a wavefunction that is much more localized. Compared to structure $b$, the energy separation in structure $a$ between the HH1 and the HH2 band is notably smaller which indicates that the required hole energy to populate the second subband is relatively lower. The separation of the HH1 and the LH1 band, on the other hand, is considerably larger in structure $a$. As a result, the onset of anti-crossing between both subbands is shifted towards higher energies which causes less nonparabolicity in the HH1 band. In addition, the Fermi energy is larger in structure $a$ due to a stronger confinement potential at the position of the wavefunction. The step-graded structures $c$ and $d$ of Figure 5.1 have a triangular profile which shifts the envelope function of the HH1 band close to the centre of the quantum well. Similar to the relation observed in the first two structures, the lower Ge concentration in structure $d$ results in an envelope function that is slightly more centred and much less broadened compared to the one in structure $c$. Concerning the energy separation of the HH1 and HH2 bands, the difference is only marginal between both structures but the LH1 band in structure $c$ is much further separated in energy from the corresponding HH1 band than in structure $d$. Again, the Fermi energy is larger in structure $c$ due to a higher Ge composition in the middle step.

The observed trend of the envelope functions as described for the structures in Figure 5.1 demonstrates the need for a compromise regarding the general spread of the wavefunction and the nonparabolicity of the HH1 band when choosing the Ge concentration of the three steps in the quantum well. Possible solutions that address this problem are in general associated with modifications to the grading profile and comprise in particular an adjustment to the number of steps, the step width, the amount of change in the Ge content per step and the maximum Ge concentration in the channel. However, certain restrictions have to be observed when varying the step width and the Ge change per step. As an approximate guide to the shape of the HH1 envelope function, the ratio $\eta$ of step width to Ge change per step determines the broadening of the peak. The minimum value of the ratio is limited by the total width of the quantum well. When the ratio is increased, the peak broadens until the step width becomes too large for the change in Ge content to pull the wavefunction away from the upper interface and the HH1 envelope function is that of a uniform channel.

To demonstrate the dependence of the wavefunction on these parameters, the band structure of four step-graded channels consisting of five and eleven steps are shown in Figure 5.2. By increasing the number of steps to five, the linearly ramped profile of structure $a$ generates a HH1 envelope function which is less broadened and closer to the lower interface than in structure $b$ of Figure 5.1 but the energy separation between the HH1 and the LH1 band is similar to that in structure $a$ of Figure 5.1. Exactly the same effect is achieved for the envelope functions of the triangular profile in structure $b$, also by increasing the number of steps. This improvement of the
envelope function can be explained by the fact that the step width is smaller in structures \(a\) and \(b\) of Figure 5.2 but the change in Ge content per step is the same as in Figure 5.1 thus reducing the ratio \(\eta\) from 1 to \(3/5\) (in \(\text{nm}^\%\)).

Increasing the number of steps even further to eleven steps does not improve the HH1 envelope function with respect to its shape and position or alter the higher subbands significantly. In fact, the HH1 envelope function of structures \(c\) and \(d\) of Figure 5.2 are marginally broader than in structures \(a\) and \(b\) due to a slightly larger ratio \(\eta\) of \(3/4\). By increasing the number of steps, however, the discrete band structure model progressively approaches the situation of a continuously graded profile which ultimately should provide the best transport behaviour as there are no interfaces within the channel to cause additional scattering. Hence, it is possible to calculate the envelope functions of continuously graded profiles by approximating the corresponding structure with a large number of steps (eleven steps for example).

Since the present SCPS programme does not include a scattering model, a prediction of the mobility as a function of the position of the wavefunction is not possible. In order to identify the point in the quantum well where the influence of scattering from the upper and the lower interface is at a minimum, interface charge as well as
interface roughness scattering of both interfaces has to be investigated. Depending on the location of this point, the wavefunction can subsequently be positioned exactly at this point by employing an appropriate triangular profile which may also be asymmetric if necessary. Another very important benefit of a graded quantum well is the much lower Ge content at the upper interface thus ensuring less scattering than in a uniform channel with a Ge concentration equivalent to the maximum of the graded profile which is necessary to achieve a similar confinement potential. Theoretically, scattering from the upper and lower interface can be reduced to a minimum by employing a continuously graded triangular profile which starts and finishes with a zero Ge composition at the edges of the quantum well.

5.3 **Band Structure of Grown 2DHG Samples**

The necessity to obtain information about the scattering processes involving the upper and lower interface requires the growth of uniform quantum wells prior to the modelling and eventual realization of graded structures. If gated Hall bar structures are used, modulation of the sheet density, depletion of the supply layer and positioning of the wavefunction can also be achieved in 2DHG structure with a uniform channel by adjusting the bias accordingly. Although the application of a gate requires an additional photolithographic step consisting of the thermal deposition of an insulating oxide and the evaporation of the gate metal, the resulting flexibility justifies the slightly more complicated processing. Following this approach, the influence of scattering from the upper as well as from the lower interface on the hole mobility can be investigated by measuring the transport properties as the wavefunction is shifted along the width of the channel towards the lower interface. Taking into account the variations in the sheet density, the measurement with the highest mobility should also indicate the position of the wavefunction for which the scattering from both interfaces is lowest. Apart from the fact that single-step structures are much simpler to grow, it is also easier to investigate the dependence of the hole effective mass on the Ge content in a uniform channel.

Considering the relevance of these arguments, three single-step structures were initially grown, each with a different Ge composition in the channel, for which the band diagram is shown in Figure 5.3. The width of the uniform quantum well in these structures was chosen to be very large in order to allow for a separate examination of the upper and lower interface. When comparing the position and shape of the HH1 envelope function in the three structures, the peak of the envelope function can be seen to move closer to the upper interface with increasing Ge content and a corresponding drop in the width of the peak is clearly visible. Evidently, both effects are caused by an increasing confinement potential at the upper interface. It should be noted that the difference in shape and position of the peak is greater between structures $a$ (10%) and $b$ (20%) than between structures $b$ and $c$ (30%). The energetically lowest subband (HH1) and the second lowest subband (HH2) progressively separate in energy with increasing Ge concentration and so does the
hole energy required to populate the HH2 subband. A second occupied subband would increase the scattering rate due to additional intersubband scattering. The energy separation between the HH1 and the LH1 band also increases with the Ge content which indicates a reduction in the nonparabolicity of the HH1 subband.

Figure 5.3 Calculated valence band diagrams for the grown 2DHG structures with corresponding envelope functions of the hole bands. The width of the QW in the single-step structures is 40 nm and the Ge concentration is 10% (a), 20% (b) and 30% (c). The step-graded structure in (d) is identical to the structure in Figure 5.1 (a).

As shown in structure d, a linearly graded quantum well consisting of three steps was finally grown to compare its transport properties with those of the single-step structures. The graded profile was chosen to be realized with only three steps as described in Figure 5.1 (a) mainly due to the lack of experience concerning the growth of graded Ge profiles prior to the growth of this sample. In accordance with this approach, a triangular step-graded profile as described in Figure 5.1 (c) was also grown. However, the SdH and Hall channels of the corresponding samples became highly resistive when cooled down thus preventing any meaningful measurements at low temperatures. An exact description of the layer structure and growth conditions of the four samples that are analyzed experimentally is given in chapter 7. Unfortunately, the quality of the gate oxide is very poor thus causing a relatively high leakage current to the gate which prohibits the use of a gated Hall bar structure. Hence, none of the associated benefits referred to in this section can be exploited in the experiments involving the single-step structures.
5.4 Summary

Using a self-consistent Poisson-Schrödinger solver to calculate the corresponding band structures and envelope functions, it has been shown that the wavefunction in the channel of a 2DHG structure can be moved away from the low-quality upper interface by grading the Ge content accordingly. Three important aspects have to be considered when designing the Ge profile of the channel. First, the shape and position of the envelope function determines the location and degree of spread of the wavefunction. Secondly, the energy separation between the HH1 and HH2 band gives an indication for the population of the second subband, and thirdly, the energy separation between the HH1 and LH1 band affects the nonparabolicity of the HH1 band. The structural parameters that control these aspects are the grading profile (linearly graded or triangular profile), the number of steps, the step width, the percentage change of Ge per step and the maximum Ge content in the channel (determines the hole effective mass). Concerning the reduction of interface scattering, the ideal structure is a continuously graded profile which can be approximated in the model with a large number of steps and a corresponding small step width. The band structure and envelope functions of the experimentally examined samples are also presented to demonstrate the importance of the SCPS solver in designing high-quality 2DHG structures. Apart from the three samples with a uniform channel, only one of the step-graded structures could be investigated at low temperatures despite having grown a range of samples with different profiles. This lack of consistency to reproduce step-graded structures of high quality demonstrates the difficulties and current inexperience concerning the growth of such structures.
6 EXPERIMENTAL TECHNIQUES

A selection of different magnetic systems was chosen for their distinct characteristic performance to investigate certain aspects of the samples. As a result, magneto-transport measurements could be performed across the whole temperature range from room temperature down to the milli-Kelvin regime. Altogether six magnetic systems were employed to carry out a variety of experiments on the present $n$- and $p$-type SiGe heterostructures:

1. Probing Station,
2. 8T Variable Temperature Insert,
3. 4T Sorption Pumped $^3$He Insert,
4. 15T Variable Temperature Insert,
5. 15T Dilution Refrigerator,

Directly after the samples had been grown, they were tested on a probing station where basic van der Pauw measurements could be carried out at room temperature and 77 K. From these measurements, vital information such as sheet density and carrier mobility could be extracted without having to process the samples photolithographically. Thus, the quality of the grown structures could be determined immediately. After this initial test, the samples with the best transport properties were processed into Hall bars and subsequently measured in the cryomagnetic systems.

Initial measurements at low temperatures and high magnetic fields were carried out in a cryostat with a Variable Temperature Insert (VTI) at the Interdisciplinary Research Centre for Semiconductor Materials (IRC) at Imperial College in London. With the system's 8T-superconducting solenoid, magnetic sweeps can be performed at temperatures between 4.2 K and 1.7 K from which the carrier concentration and mobility can be determined as well as the quantum and transport relaxation times. For measurements at lower temperatures, the Sorption Pumped $^3$He Insert of the Electrical Engineering facilities at Imperial College in London was made available. This cryomagnetic system has a 4T-superconducting magnet and is capable of reaching temperatures as low as 300 mK, thus allowing additionally the calculation of the effective mass and the energy-loss rate. For temperature measurements above 4.2 K, the closed cycle cryostat at University College London was employed which operates in a temperature range between 10 K and 300 K. As magnetic sweeps are of less interest at those temperatures, temperature sweeps are performed at a constant magnetic field of 0.314 T from which the temperature-dependent carrier concentration and mobility as well as the deformation potential can be extracted.
To obtain even more detailed information about the investigated 2D structures, the best performing samples were taken to the High Magnetic Field Laboratory in Grenoble where two different cryomagnetic systems were available. All samples were first measured in the VTI employing a 15T-superconducting solenoid with which magnetic sweeps at temperatures between 4.2 K and about 1.3 K can be performed. The second cryostat consists of a 15T-superconducting magnet and a dilution fridge enabling magnetic sweeps at any temperature between 4.2 K and the base temperature which can be as low as 20 mK but strongly depends on the cooling process and consequently varies between individual cool-down cycles.

Employing so many cryomagnetic systems resulted in the problem of having to mount the samples on different header systems according to the cryomagnetic equipment used. Altogether three distinct arrangements were chosen: (i) thermo-compressive bonding of aluminium wires onto the contacts of the sample which is mounted on a ceramic chip carrier, (ii) silverpainting of gold coated copper wires to the sample on a DIL socket and (iii) bonding of aluminium wires to the sample on a TO8 header. The obvious problem arising from the use of three different header systems is the necessity to frequently re-connect the samples when used in different systems to accommodate the matching header system. Not only does this process take up time, but the quality of the contacts also suffers from frequent re-connections. A way around this problem is to have three specimen of the same sample, each mounted to one header type. The biggest problem with this approach is that even if the specimens are taken from adjacent places on the wafer, they still can exhibit different properties due to anisotropies or inhomogeneities of the material. Hence, the same specimen of each sample was used whenever possible.

### 6.1 Description of Equipment

In order to point out the differences between the cryomagnetic systems employed in the magnetotransport experiments, a short description of the basic principle of design and also of the mode of operation is presented for each system.

**Probing Station (300K and 77K)**

This is the most simple and straightforward of the employed systems and does not need a lot of preparation before it can be used. Its design is very elementary and consists of four electrical probes and a permanent magnet to perform the Hall effect measurements. The permanent magnet produces a magnetic field of 0.32 T to within ±1% over a radius of 12 mm from the pole centre. Hall effect measurements can be carried out with the sample having either the van der Pauw or the Hall bar configuration. In both cases, the whole magnet assembly has to be rotated by 180° to reverse the magnetic field. The sample current is generated by an integral current source that can provide currents between 1 µA and 20 mA. The sample stage consists
of a small open dewar which can hold the liquid nitrogen for the measurements at 77 K and a copper block inside the dewar to stabilize the temperature.

Since there is a big interest in a quick semi-quantitative characterization of the 2D heterostructures immediately after they are grown, indium dots are annealed close to the corners of the sample to form the contacts for both p- and n-type structures for VdP measurements. For all p-type structures, indium forms an ohmic contact, and in the case of the n-type structures, a tunnel contact with ohmic behaviour is being formed. No adverse effects from the contacts could be observed in the temperature range of the system.

After placing the sample on the sample stage and positioning the contact probes, the contacts are first checked whether they are ohmic in nature. Due to the limitations of the system, the maximum resistance between any two contacts cannot exceed 1 MΩ. As the internal voltmeter is capable of performing phase-sensitive measurements, an AC current at 200 Hz is chosen for the actual VdP measurements to reduce the noise and increase the accuracy of the measured signal. Thus, much lower sample currents can be used to avoid heating effects. Before measurements are performed at 77 K, the contacts have to be checked again for ohmic behaviour.

8T VARIABLE TEMPERATURE INSERT (4.2K AND 1.7K)

In principle, this type of top loading cryostat consists of two concentric dewar vessels separated by a vacuum chamber, a superconducting solenoid and a sample chamber. The inner dewar vessel is filled with liquid helium and covers the Niobium Titanium (NbTi) superconducting solenoid which operates at liquid helium temperature. Around the inner dewar vessel, a vacuum chamber acts as an insulating thermal shield against the higher temperatures outside the cryomagnetic system. An additional thermal shield is being provided by the outer dewar vessel containing liquid nitrogen. The sample chamber itself is situated inside the inner dewar vessel and lies exactly in the middle of the solenoid. As can be seen in Figure 6.1, the vacuum chamber and an additional radiation shield separate the sample chamber from the main helium bath, thus insulating the sample chamber thermally and facilitating the cooling process of the insert to temperatures below 4.2 K or alternatively the heating to higher temperatures.

Through the sample tube, a probe containing the sample holder can be inserted into the system, placing the sample directly in the centre of the solenoid. To guarantee a constant temperature of 4.2 K during the measurements, the sample chamber itself can be filled with liquid helium through a siphon and a needle valve which connect the inner dewar vessel and the sample chamber. A copper block is used as the sample holder to keep the temperature stable over a period of time for the measurements which are performed at temperatures other than that of liquid helium. To heat the sample above 4.2 K, the copper block contains a resistor acting as a heater and a diode to measure the temperature in the copper block. Temperatures below 4.2 K can be achieved by evaporation cooling of liquid helium whereby the helium vapour in the sample chamber is pumped away and with it the latent heat of vaporization of the liquid helium. As the equilibrium vapour pressure drops with
falling temperature, thus reducing the rate at which helium gas and its heat of vaporization can be extracted, the lowest accessible temperature depends on the vacuum and lies around 1.7 K for this cryostat. Depending on the amount of helium gathered in the sample chamber, this temperature can be kept for 30 - 60 min until the chamber is empty. Leaving the needle valve a fraction of a turn open can prolong the pumping procedure marginally. There are no metastable temperature regimes during the cooling process, making it necessary to use the heater to stabilize at a temperature between 4.2 K and 1.7 K. Since this process is very unstable and reduces the pumping time drastically, this option was never used. Neither was the heater used to achieve temperatures above 4.2 K because of the resulting severely increased helium boil-off. Hence, measurements were only carried out at liquid Helium temperature and 1.7 K.

![Schematic diagram of a top loading cryostat.](image)

Figure 6.1  Schematic diagram of a top loading cryostat.

Another cause for an increased helium boil-off is the cooling process of a newly inserted sample which can either be at room temperature or pre-cooled at 77 K. How quickly the sample is cooled down depends on whether the sample chamber contains
liquid helium, in which case the sample will be cooled almost instantaneously to 4.2 K causing an intense helium boil-off. If there is no liquid helium in the sample chamber, the cooling process can take up to 60 min at a much lower boil-off rate. It is not known which of the two methods boils off more helium over all.

As supply for the superconducting solenoid, a programmable current source is used which is capable of generating the 75 A necessary to reach a maximum magnetic field of 8 T in a 40 mm bore. Each time a magnetic sweep is performed, the solenoid is heated, which results in an increased helium boil-off. For a sweep over the full magnetic field range, a minimum time limit of at least 5 min should be kept to avoid high inductive currents within the solenoid which in turn could lead to an overheating and eventually the quenching of the magnet.

To connect the samples to the probe, the copper block of the insert accommodates a 10-pin DIL socket. The standard method to mount the samples onto a DIL header is the usage of silver paint to connect gold coated copper wires to the contacts of the Hall bar structure and subsequently solder the other ends of the wires to the pins of the DIL header. In addition, a second method was employed, in which the samples are bonded to a 20-pin ceramic chip carrier using aluminium wires. Subsequently, the chip carriers can be placed in an ‘easy insert’ chip socket which is directly connected to a 10-pin DIL header. This second method has been introduced as an attempt to achieve a standard header system for all cryomagnetic systems using the ceramic chip carriers.

4T SORPTION PUMPED $^3$He INSERT (4.2K - 300mK)

An altogether different design, this system relies on a $^3$He refrigerator to achieve temperatures below 1 K. Although it relies on the same evaporation cooling process, the use of the expensive helium isotope as the working substance requires a more complicated design. The $^3$He gas is stored within the system at all times, making the insert relatively easy to run. Apart from the $^3$He refrigerator, the insert consists of a NbTi superconducting solenoid and a cold finger which accommodates the sample. The refrigerator itself consists of a sorb heat exchanger, a 1 K pot and a $^3$He pot to which the cold finger is connected. The charge of $^3$He is sealed into a self-contained storage vessel to keep the $^3$He gas in a closed cycle within the insert. A structural description of the important parts of the insert is given in the schematic diagram in Figure 6.2. To operate the insert, the whole system has to be mounted in a liquid helium storage dewar.

While mounting the sample, it is important to establish a good thermal contact between the ceramic chip carrier and the cold finger in order to cool the sample to the base temperature. The copper wires of the cold finger can be soldered directly to the gold contacts of the chip carrier. A metal tube containing the magnet is placed over the arrangement of refrigerator and sample to form the inner vacuum chamber (IVC). Finally, the IVC is sealed, pumped to a rough vacuum and lowered slowly into the liquid helium reservoir. While carefully being lowered into the dewar, the insert is pre-cooled by the cold helium gas resulting in a reduced helium boil-off. Taking advantage of this effect, it takes typically up to three hours to cool the insert.
from room temperature to below 10 K. To assist this part of the cooling process, a small amount of helium gas has to be inserted into the IVC, thus establishing a better thermal conduction between the fridge and the metal tube. Once the insert has reached temperatures below 10 K, the residual helium gas within the IVC is absorbed by a piece of charcoal. Together with the cryopumping effect, a good vacuum should be established to thermally insulate the fridge while it is being cooled down further to temperatures below that of liquid helium.

During the process of cooling down, the temperature of the sorb heat exchanger has to be kept above 40 K so that it will not absorb any $^3$He gas. After the fridge has cooled down to about 5 K, the 1 K pot can be filled with liquid $^4$He through a pick up tube and a needle valve, thus cooling it further down to 4.2 K. While leaving the needle valve slightly opened to maintain a certain helium level, the 1 K pot should be pumped, which will cause the temperature to start falling to around 1.5 K within a time period of 30 - 60 min. As the 1 K pot cools below 3 K, the $^3$He gas will start to condense and run down into the $^3$He pot. The condensation process generally takes between 10 and 20 min. After all $^3$He has been condensed in the $^3$He pot, the heater of
the sorb can be switched off. While cooling down, the sorb starts pumping on the liquid $^3$He, lowering the temperature of the $^3$He pot even further until it reaches a base temperature after 20 - 30 min. Since $^3$He has a higher vapour pressure than $^4$He, the base temperature of the $^3$He pot is much lower than that of the 1 K pot. To operate the sorb at its most effective, its temperature has to be kept as low as possible. Hence, $^4$He can be pumped through the sorb via a second pick up tube and needle valve to keep its temperature below 5 K. The process of condensing the $^3$He and the subsequent evaporation cooling of the insert is illustrated in Figure 6.3. In principle, the base temperature of the fridge can reach 300 mK and can be kept for several hours, but the contribution of different factors, such as the mass of the sample, current induced sample heating and inductive heating caused by the changing magnetic field, can lead to an increased base temperature and a reduced time period in which the base temperature is stable. For most samples, a base temperature of around 400 mK could be achieved for a duration of about two to three hours.

![Figure 6.3 Operating principle of a sorption pumped $^3$He system.](image)

Because the superconducting solenoid has to be relatively small to fit inside the metal tube of the IVC, its bore is with 20 mm only half that of the 8T VTI. At 45 A, it produces a magnetic field of 4 T with a homogeneity of 0.1% over 10 mm axial
length. Notwithstanding a maximum sweep rate of 12 A/min, it is prudent to allow 10 min per sweep to avoid excessive inductive heating of the sample which would effect the base temperature and hold time.

**15T VARIABLE TEMPERATURE INSERT (4.2K-1.3K)**

Although very similar in design to the 8T VTI of the IRC, this top loading cryostat has a very different insert system to accommodate four samples at the same time. To be able to do that, four 8-pin DIL connectors are arranged in-line along the tip of the probe. All samples are mounted to the DIL headers using the standard method described earlier. After having connected four samples this way, a metal tube is being pulled over the probe, sealed off and evacuated. Once put inside the cryostat, each sample can be placed in the centre of the magnet by moving the whole insert up or down. This way, four samples can be measured at low temperatures within a single cooling cycle. In order to cool down the samples, a good thermal link to the metal tube and thus the sample chamber has to be established. Hence, the insert has to be filled with helium gas before being lowered into the cryostat. It can take between 30 and 60 min to cool the insert from room temperature to 4.2 K, depending on how fast it is being lowered into the cryostat. No special care concerning the helium boil-off has to be taken because of the installed helium retrieval system.

The most important difference to the cryostat of the 8T VTI with respect to the temperature control is a pumping system with more powerful pumps and a more accurate control of the needle valve to let the liquid helium flow dynamically through the siphon from the main helium bath into the sample chamber. Thus, a continuous flow of helium can be achieved while pumping on the sample chamber, providing an infinite supply of liquid helium for the evaporation cooling process. Variable pumping speeds and fine-tuning of the needle valve allow for a very precise and stable temperature control below 4.2 K. While pumping on the helium vapour in the sample chamber, the temperature can drop to 1.3 K and even below due to a better vacuum compared to the 8T cryostat. Since there is virtually no time limit to the evaporation cooling process in the sample chamber, it is possible to measure all four samples at a constant stabilized temperature.

To achieve magnetic fields up to 15 T, a different type of superconducting magnet has to be employed in this cryostat. Whereas for magnetic fields up to 9 T NbTi can be used for the design of the solenoid, higher fields (up to 20 T) require the use of the expensive and brittle intermetallic compound Nb₃Sn. Yet, for economic reasons, the Nb₃Sn is only used for the inner section of such a magnet where the magnetic field is high. For the outer sections, the cheaper NbTi is used. In addition, the solenoid has to be big enough to accommodate the diameter of the entire insert including the encapsulating metal tube. At 52 mm, the magnet's bore is therefore rather large for induced magnetic fields that can reach up to 17 T. However, due to safety restrictions, the maximum magnetic field is restricted to 15 T at 100 A which is generated by a current source that is precise to within 1%. To avoid excessive heating of the magnet, a minimum sweep time of 10 min over the whole field range has to be followed.
15T DILUTION REFRIGERATOR (4.2K - 30mK)

Being by far the most complicated cryogenic system, the top loading dilution refrigerator is capable of reaching the lowest temperatures. It uses a closed cycle of $^3$He/$^4$He mixture to achieve temperatures as low as 30 mK. The high throughput of the fridge is specially designed to allow cooling from room temperature to 50 mK in less than three hours. Further cooling to the ultimate base temperature of around 30 mK can take a few hours and is typically done overnight. Due to the magnet's large cold bore of 34 mm, a very large volume of $^3$He/$^4$He mixture is needed.

As shown in the phase separation diagram of a $^3$He/$^4$He mixture in Figure 6.4, the two liquids are miscible in all proportions above the tricritical point at 0.87 K. Since both helium isotopes become superfluid at different temperatures ($^4$He at 2.17 K and $^3$He at just below 1 mK), the transition temperature between superfluid and normal liquid state is dependent on the $^3$He concentration in the mixture. Below the tricritical temperature, liquid $^3$He and $^4$He are immiscible over a wide composition range, thus separating the mixture into two different phases, the concentrated $^3$He phase and the dilute $^3$He phase. Hence, a mixture with composition in the unstable regime will decompose into two separate phases whose compositions are given by the two branches of the $\lambda$-curve.

![Phase separation diagram of a $^3$He/$^4$He mixture.](image)

Figure 6.4  Phase separation diagram of a $^3$He/$^4$He mixture.

For temperatures towards 0 K, the concentrated phase becomes essentially pure $^3$He, whereas the $^3$He concentration of the dilute phase drops to about 6.3%. The concentrated phase of the mixture can be regarded as liquid $^3$He and the diluted phase as a quasi-gas of $^3$He atoms dissolved in liquid $^4$He. For mixtures with a $^3$He concentration exceeding 6.3%, the excess condenses into concentrated liquid $^3$He,
thereby liberating latent heat. By evaporating liquid $^3$He across the phase boundary from the concentrated to the dilute phase, latent heat is consumed. This process constitutes the basic principle of a helium dilution refrigerator [116, 117].

Shown in Figure 6.5, the schematic diagram of such a refrigerator illustrates the working principle of the cooling process. In the still, $^3$He is selectively distilled from the $^3$He/$^4$He mixture due to a higher vapour pressure and pumped off. To obtain a useful $^3$He evaporation rate, the temperature of the still has to be raised to about 0.6 - 0.7 K, still guaranteeing a distinctly smaller $^4$He vapour pressure. The lower concentration of $^3$He in the still causes $^3$He atoms from the $^3$He/$^4$He mixture in the mixing chamber to diffuse through a counterflow heat exchanger into the still, thus decreasing the $^3$He concentration in the dilute phase. In order to restore the equilibrium concentration, quasi-evaporation of $^3$He atoms from the concentrated phase to the dilute phase will take place under consumption of latent heat, which causes the mixture to cool down. The pumped-off $^3$He is returned to the system through a condenser that is cooled to about 1 K by contact with a pumped $^4$He bath. A narrow capillary tube creates a flow impedance to maintain a high enough pressure in the condenser for the gas to condense. Before re-entering the mixing chamber, the liquefied $^3$He is further cooled, first in the still, then in the counterflow heat exchanger.

Figure 6.5  Operating principle of a $^3$He/$^4$He dilution refrigerator.
As the probe contains a 14-pin DIL socket, two 8-pin DIL headers can be placed directly next to each other after having modified the pin configuration of one of the headers. While cooling down from 4.2 K to about 1.5 K, the temperature change is relatively slow, making it possible to perform magnetic sweeps at intermediate temperatures. For each sweep cycle, which consists of an up-sweep and a subsequent down-sweep, the temperature is measured at zero magnetic field at the beginning of the up-sweep and also at the end of the down-sweep due to the sensitivity of the temperature sensor to high magnetic fields. The superconducting magnet of the dilution fridge is of the same type as the magnet for the 15T VTI, capable of reaching 17 T in a 34 mm bore but restricted to a magnetic field of 15 T at 100 A as a precaution and limited to a maximum sweep rate of 1.5 T/min. At base temperature, the sweep rate has to be lowered to around 0.5 T/min to minimize the effects of inductive heating of the sample and the mixing chamber.

**CLOSED CYCLE CRYOSTAT (300K - 10K)**

All cryomagnetic systems mentioned so far operate in the low-temperature regime, making use of the Shubnikov-de Haas and the quantum Hall effect by ramping the magnetic field at constant temperatures. In the higher temperature range, these effects vanish, which leaves only simple Hall effect measurements to investigate the temperature dependence of some physical properties. Hence, superconducting magnets capable of high magnetic fields are not necessary in the region between 10 K and room temperature, rendering the design of this type of cryomagnetic system much more simple. By pumping helium in a closed cycle through a cold head, the system can be cooled down to about 10 K. The $^4$He gas is compressed to $1.65 \times 10^4$ Pa in a compressor and then expanded in a controlled manner in the cold head. Intermediate temperatures can be achieved and stabilized through variations of the pumping speed and an integral heater.

The cold finger is equipped to take 12-pin TO8 headers on which the samples have been bonded with aluminium wires. A constant magnetic field of 0.314 T is supplied by a solenoid which is powered by a 5 A current source. Reverse fields can be achieved by switching polarity of the contacts to the solenoid.

To summarize the characteristics of the different cryomagnetic systems, an overview of their features is given in Table 6.1. All experiments can generally be put into one of two categories, i.e. those at high temperatures and low magnetic fields or those at low temperatures and high magnetic fields. Whereas experiments involving the probing station and the closed cycle cryostat fit into the former group, all cryostats using a superconducting magnet provide the necessary conditions to perform measurements of the latter category. Regarding the analysis of low-temperature magnetotransport phenomena and the corresponding transport properties, the 15T dilution fridge is by far the most important cryomagnetic system since it allows temperature-dependent SdH and QHE experiments up to high magnetic fields in the crucial temperature range between 4.2 K and 30 mK.
6.2 Experimental Set-up

The experimental set-up for the high-temperature measurements, carried out in the probing station and the closed cycle cryostat, is relatively simple and consists mainly of a current source, a magnetic field perpendicular to the plane of the sample and a voltmeter. In the case of the closed cycle cryostat, a Proportional-Integral-Differential (PID) temperature controller is employed to stabilize the temperature. Additionally, an optional scanner card can be used in the digital voltmeter (DVM), enabling the recording of all three Hall voltages of the sample in a single step for every temperature point. In order to exploit the Hall effect, two measurements have to be
performed, one with a positive and one with the corresponding negative magnetic field at a given temperature. Due to the long temperature settling time of the system during the cooling process, performing those measurements can take a long time and can become impractical if too many temperature points are chosen. Using an alternative approach instead, the magnetic field should be left constant while cooling down the sample, scanning the Hall voltages whenever a given temperature is reached during the cooling process. Thus the temperature does not have to settle at a certain value. The second reading of the Hall voltages is taken at the same temperature points, with the magnetic field reversed, while warming up the sample. Consequently, several hours can be saved during a temperature sweep between 300 and 10 K with temperature steps of for instance 10 K. This solution requires the temperature sensor to be as close to the sample as possible to guarantee an accurate temperature reading of the sample. Furthermore, the whole process can be automated by controlling both the temperature sweep as well as the measurements by a computer. One problem associated with this approach is a potential temperature hysteresis within the system, leading to different effective temperatures at the same set points between the cooling-down and the warming-up process. Another possibility is the 2D electron system itself behaving differently when warmed up or cooled down, e.g. the freeze-out and the re-appearance of the parallel layer might occur at different temperatures and at different rates. Both problems were found to be negligible as neither of the effects could be observed in comparative experiments.

A more sophisticated approach concerning the experimental set-up has to be taken for the low-temperature experiments where electron heating effects can occur. At very low temperatures, the lattice might not be able to absorb all of the dissipated input power, leaving the rest to be absorbed by the electron system, which can lead to an increase of the electron temperature above that of the lattice. In this case, the actual temperature of the system is higher than measured by the temperature sensor, leading to a false interpretation of the experimental results. Since the input power is proportional to the square of the injected current \( P = IR \), the value of the current has to be chosen very carefully to avoid any electron heating effects. By choosing a current range spanning several orders of magnitude for each temperature point, the possible onset of electron heating can be detected by observing the SdH oscillations whereby an increase of the electron temperature is represented by a damping of the envelope. Generally, currents between 1 nA and 100 μA were applied to the sample, depending on the temperature and the quality of the sample.

A crucial adverse consequence of using small currents is an increase in the noise level of the signal when measuring the transverse and longitudinal voltages of the sample. Therefore, a lock-in amplifier has to be employed which increases the signal-to-noise ratio (SNR) dramatically. Additionally the signal has to be pre-amplified before being fed back into the lock-in amplifier to achieve the best resolution of the recorded data. To further enhance the precision of the measurements, the signal from the lock-in amplifier, which is limited to five digits, is fed into a DVM capable of displaying seven digits in the mantissa. As shown in Figure 6.6, the digitized signal from the DVM is then sent to a computer through an IEEE link. The voltage source of the lock-in in conjunction with a resistance in series is used as the current source. A
modification of the amplitude of the current can be achieved by either varying the value of the resistance, by changing the output voltage of the lock-in amplifier or by a combination of both. During the experiment, great care has to be taken to ensure that the value of the resistance is much bigger than the sample resistance to avoid any influence of the sample on the total resistance, which will lead to a different operational current than calculated. In order to operate the magnet power supply safely and to control the magnetic sweep of the solenoid accurately, a Hall sensor inside the solenoid is utilized by the magnet control unit to measure the strength of the magnetic field which subsequently is also transferred to the computer through the IEEE interface.

Due to the limitations of the number of lock-in amplifiers available for the measurements, only two channels of the Hall bar structure, usually one SdH and one Hall channel, can be recorded simultaneously. Since the quality of the contacts and the material of the sample might differ slightly within a single Hall bar structure, the selection of the two channels to be measured is crucial in order to obtain the optimal results. Hence, for the first low-temperature measurements in the 8T VTI, a DVM was used in conjunction with the scanner card to record all seven channels of the Hall bar (three transverse and four longitudinal voltages), thus providing an overview of the performance of each channel.

Before starting the measurements, the parameters of the lock-in amplifiers have to be calculated and set accordingly. Among the parameters which are set once and subsequently remain constant throughout most of the experiments are the magnet sweep time, the number of data points taken per sweep, the lock-in frequency and
the lock-in time constant. These parameters have to be chosen carefully to achieve the lowest level of noise without modifying the magnetoresistance curves due to capacitive or resolution limiting effects.

First, the sweep time has to be determined by taking into account a certain minimum accuracy of the measurements as well as the heating effects of the sample and the insert due to the change in the magnetic field, both leading to a minimum acceptable sweep time. At the same time, the power consumption of the superconducting solenoid and the helium boil-off should be limited to a minimum, as well, suggesting a sweep time as short as possible. Furthermore, longer sweep times limit either the number of possible sweeps per sample or the number of samples to be investigated. A reasonable compromise for the sweep time has been established to be 10 minutes. For the base-temperature measurements in the dilution refrigerator, it was necessary for the sweep time to be raised to 30 minutes due to the increased sensitivity of the system to the heating effects from the change in the magnetic field.

The accuracy of the measurements is also dependent on the number of points taken per sweep. To resolve all oscillations of the recorded curves sufficiently, a minimum number of points is required, usually determined by the oscillation with the shortest period. The maximum number of points, on the other hand, is determined by the amount of measurements the lock-in amplifier is capable of taking per second which in turn is determined by the lock-in time constant. Adding further points to the measurement above a certain resolution will not increase the information content of the trace. A good resolution of the oscillations can be achieved by taking a measurement every 25 mT, resulting in an overall number of data points per sweep of 600.

Thus, the time interval between the acquisition of two successive data points can be calculated to be 1 s, which influences directly the choice of the time constant. Although longer time constants increase the SNR, a time constant exceeding the chosen time interval is useless as signal changes on a scale smaller than the time constant are averaged out. Hence, the time constant should be set to 300 ms, leaving the system enough response time while guaranteeing a low noise level. For the longer sweep times of the base-temperature measurements in the dilution refrigerator, the values for the time interval and time constant are retained. Instead, the number of points per sweep have to be increased to 1800 and thus the resolution raised to one measurement per 8.3 mT.

The last important parameter to be determined is the lock-in frequency which is limited by the capacitive properties of the electrical set-up such as the contacts and the size of the Hall bar. Higher frequencies reduce the noise, but as a trade-off, the signal becomes more and more distorted. Hence, the frequency has to be balanced to achieve a maximum SNR at minimal or preferably no distortion. Another problem is the possible pick-up of unwanted signals from the mains if the lock-in frequency or any of its harmonics is close to the mains frequency, thus increasing the noise level. By testing several frequencies between 5 and 100 Hz, the onset of any distortion to the signal could be detected at around 70 Hz. Finding the lowest noise level experimentally is not so straightforward since the level of picking up noise from the
mains is not linearly dependent on the frequency, although it is possible to calculate frequency bands which should be less noisy. Apart from that, the noise level of the signal did not show the tendency to increase with frequency below 70 Hz. To guarantee a totally undistorted signal and to avoid the possibility of picking up any noise from the mains, the lock-in frequency was set to 10.66 Hz.

The two remaining parameters which have to be set individually for each sweep are the sensitivity of the lock-in amplifier and the gain of the pre-amplifier. Both parameters are closely linked as they together determine the resolution of the measurements whereby the accuracy of the input signal improves with increasing resolution. Both parameters have to be balanced in the way that the recorded data does not saturate the input of the lock-in amplifier, nor should the signal be so weak that small changes are not detected. The pre-amplifier has to be used for signals that are too faint to achieve a sufficient resolution even for the highest sensitivity range of the lock-in amplifier. Unfortunately, the additional amplifying stage also introduces additional noise and distortion to the signal. To keep these negative side effects as low as possible, the gain of the pre-amplifier should always be set to its lowest possible value to find the optimum resolution while using the highest possible sensitivity for the lock-in amplifier.

6.3 SAMPLE PREPARATION

The preparation of the samples can be divided into three stages, comprising the growth of the structure, the photolithographical processing and finally the bonding. For the 2DEG structures, p-doped high resistive (2 - 10 kΩcm) nominally singular Si(100) substrates are employed, whereas for the 2DHG structures, n-doped Si(100) substrates with a resistivity of 1 - 10 Ωcm are used. Prior to the growth of the two-dimensional structures, all substrates have to be cleaned thoroughly to remove any carbon and native oxide on the surface, and subsequently, the cleaned surface has to be passivated with hydrogen. An overview of the employed modified Shiraki process is given in Table 6.2. The process starts off with a chemical wet etch with the wafer being dipped into a diluted HF solution (~9%) to remove the surface oxide. In the next step, the wafer is intentionally oxidized by immersing it in a diluted HNO₃ solution (68%) which subsequently is etched away by another wet etch in a fresh HF solution. This step is repeated three to four times to guarantee the surface oxide has been totally removed. Finally, any carbon is removed from the surface by dipping the wafer into an ammonium solution with H₂O₂ being added to the solution at the end to oxidize the surface for a last time. Hence, a final HF dip is necessary to get a perfectly clean and oxide free hydrogen terminated silicon surface. To avoid any contamination or oxidization of the substrate after the cleaning process, the wafer has to be put into the GSMBE machine immediately after the final step of the modified Shiraki process.
Sample Preparation

<table>
<thead>
<tr>
<th>step</th>
<th>action</th>
<th>temp</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>blow clean Si wafer</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>dip wafer in HF solution</td>
<td>RT</td>
<td>20 sec</td>
</tr>
<tr>
<td>3</td>
<td>wash wafer in DI H₂O¹</td>
<td>RT</td>
<td>2 min</td>
</tr>
<tr>
<td>4</td>
<td>dip wafer in HNO₃ solution</td>
<td>70°C</td>
<td>5 min</td>
</tr>
<tr>
<td>5</td>
<td>wash wafer in DI H₂O¹</td>
<td>RT</td>
<td>2 min</td>
</tr>
<tr>
<td>6</td>
<td>dip wafer in HF solution</td>
<td>RT</td>
<td>20 sec</td>
</tr>
<tr>
<td>7</td>
<td>wash wafer in DI H₂O¹</td>
<td>RT</td>
<td>2 min</td>
</tr>
<tr>
<td>8</td>
<td>repeat step no 4 to 7</td>
<td>–</td>
<td>9 min 20 sec</td>
</tr>
<tr>
<td>9</td>
<td>repeat step no 4 to 7</td>
<td>–</td>
<td>9 min 20 sec</td>
</tr>
<tr>
<td>10</td>
<td>repeat step no 4 to 7 (optional)</td>
<td>–</td>
<td>9 min 20 sec</td>
</tr>
<tr>
<td>11</td>
<td>dip wafer in NH₄OH solution</td>
<td>70°C</td>
<td>10 min</td>
</tr>
<tr>
<td>12</td>
<td>wash wafer in DI H₂O¹</td>
<td>RT</td>
<td>2 min</td>
</tr>
<tr>
<td>13</td>
<td>dip wafer in HF solution just before load²</td>
<td>RT</td>
<td>20 sec</td>
</tr>
</tbody>
</table>

¹with deionized water running, ²clean excess solution off with filter paper

HF solution: HF:DI H₂O = 1:10
NH₄OH solution: NH₄OH:H₂O₂:DI H₂O = 1:4:20 (H₂O₂ added to the solution at the end)

Table 6.2 Overview of the modified Shiraki process for cleaning a wafer before growth.

A thorough implementation of the cleaning process is essential to guarantee optimum conditions of the substrate for homogeneous and isotropic two dimensional crystalline growth of the initial Si buffer layer and all subsequent epilayers.

**SAMPLE GROWTH**

All structures were grown in a GSMBE system using disilane (Si₂H₆) and germane (GeH₄) as matrix material sources, 0.1% arsine (AsH₃) diluted in argon as the n-type dopant source and 0.1% B₂H₆ diluted in argon as the p-type dopant source. A background pressure of about 10⁻⁴ Pa was maintained by cryopumping while during growth, typical pressures were in the 10⁻⁵ Pa range, obtained by turbo molecular pumping. As shown in Figure 6.7, the system is designed to allow the introduction of four distinct gases into the chamber via four highly collimated beams whose inlet pressure is computer controlled by capacitance manometers. The substrates are mounted on high purity poly-silicon substrate holders designed to be heated radiatively by a SiC-coated graphite meander-path heater element. Due to the pre-cleaning technique which resulted in a hydrogen terminated silicon surface, the growth could commence at ~500 °C with the hydrogen coming off the surface at around 450 - 550 °C. After being introduced into the preparation chamber through
the load lock, the wafer has to undergo an outgassing process for at least 3 hours at \(-270^\circ C\) to get rid of all residual moisture on the surface before being mounted on the substrate holder in the growth chamber. For every structure, an initial 200 - 500 nm Si buffer layer was grown on top of the Si wafer consisting of a low-temperature Si buffer grown at 500 - 550 °C followed by a high-temperature Si buffer grown at 700 - 750 °C. In order to check the integrity of the surface prior to and after the buffer growth, \textit{in-situ} RHEED measurements are carried out with a 15 keV electron beam skimming the surface at an incidence angle of \(-0.5^\circ\). This technique allows the detection of any potential contamination of the surface through residual carbon or oxide. Thus, it is possible to establish the appropriate growth conditions to maintain two-dimensional growth throughout the growth process. Cooling panels are mounted close to the chamber walls to protect the metal works of the growth chamber from the heat inside the chamber during growth.

**Growth Chamber**

- Base Pressure \(\leq 10^{-4}\)Pa
- Growth Pressure \(\leq 5 \times 10^{-4}\)Pa

**Preparation Chamber**

- Base Pressure \(\geq 10^{-4}\)Pa
- Outgassing stage
- Load lock

- Cassette
- Chilled water cooling panel
- Isolation valve
- Wafer transport

- RHEED gun
- RHEED screen

**Source gases:**
- \(\text{Si}_2\text{H}_6\)
- \(\text{GeH}_4\)

**Dopant gases:**
- \(0.1\% \text{B}_2\text{H}_6\) in Ar
- \(0.1\% \text{AsH}_3\) in Ar

![Schematic diagram of a Gas Source Molecular Beam Epitaxy machine.](image)

Although the structural design of 2DEG and 2DHG structures are very different, all samples were grown in the temperature range between 500 - 550 °C where the growth rate of Si and \(\text{Si}_x\text{Ge}_{0.3}\) are \(\sim 0.2\) and \(\sim 1.2\) nm/min, respectively. The growth of a 2DEG structure is complicated by the fact that in order to confine the electrons in a quantum well, a Si layer has to be grown under tensile strain on top of a relaxed SiGe
layer. Since there are no SiGe substrates commercially available, a virtual SiGe substrate has to be established during the growth process as shown in the schematic layer diagram of a 2DEG structure in Figure 6.8. After the growth of the high-temperature Si buffer layer on top of the substrate, a SiGe buffer layer is grown with its Ge content graded linearly from 5% to 35% over 1.5 - 2 μm. Its principal function is to relieve the strain in the alloy and to confine dislocations near the substrate-epilayer interface. Near the top of the compositionally graded layer, the lattice constant relaxes approximately to that corresponding to a 30% Ge alloy. A strain-relaxed 0.5 - 1 μm Si0.7Ge0.3 buffer layer is subsequently grown which serves as a virtual substrate for a 11 nm Si channel grown under tensile strain. An undoped 15 nm Si0.7Ge0.3 spacer layer separates the Si channel from a 30 - 50 nm As doped (> 10¹⁸ cm⁻³) Si0.7Ge0.3 supply layer which is capped by 6 nm undoped Si0.7Ge0.3 grown to suppress excessive As segregation into the 2 nm Si cap. After the growth, a protective oxide will be formed at the surface protecting the whole structure from the environment. The composition of these layers and degree of relaxation is verified by symmetric and asymmetric X-ray diffraction measurements (XRD), the As doping profile is checked by secondary-ion mass spectroscopy (SIMS) and dislocation densities are determined by transmission electron microscopy (TEM).

Figure 6.8  Schematic layer diagram of a modulation-doped 2DEG structure (not to scale).

Due to the UHV conditions and low growth temperatures of the GSMBE system, the incorporation of background impurities in the epilayers is kept at a minimum with SIMS profiles of the structures showing oxygen background concentration levels well below those detected in the substrate material and carbon levels close to the detection limit of SIMS. Yet, donor like background impurity concentration levels in the order of 10¹⁵ cm⁻³ are detected in all samples, and photoluminescence (PL) studies identify these impurities as phosphorous, thought to be incorporated from the disilane gas source. Another significant advantage of the low growth temperature is...
the lowering of the Gibbs energy of segregation for Ge in SiGe alloys due to the presence of hydrogen on the surface of the substrate, consequently allowing for more abrupt interfaces in GSMBE than in SSMBE. Accordingly, SIMS measurements of the structures indicate characteristic decay lengths of 2.4 nm/decade for the Si/Si$_{0.7}$Ge$_{0.3}$ interfaces. Although phosphine (PH$_3$) has been established as the preferred $n$-type dopant source of both Si and SiGe in CVD and GSMBE, arsine (AsH$_3$) has been used in these structures due to a higher solid solubility and lower diffusion coefficient of As in Si when compared to P in Si. Both dopant sources are known to reduce the layer growth rate by as much as a factor of two. A further negative effect is the strong segregation of As during the growth of Si and SiGe. The Gibbs energy of segregation for As in SiGe, however, is lowered considerably compared with that for As in Si, allowing for a wide range of doping concentrations in the supply layer.

With regard to the strain relaxation of the SiGe buffer layers, the slow growth rate favours dislocation glide relative to dislocation nucleation in the linearly graded buffer layer, which is characterized by dislocations with relatively long misfit segments and propagation into the substrate, resulting in relaxed layers with low threading dislocation densities. XRD measurements of the subsequent Si$_{0.7}$Ge$_{0.3}$ layer show a relaxation by more than 95%. Cross Section TEM studies indicate threading dislocation densities below 10$^7$ cm$^{-2}$ in the relaxed Si$_{0.7}$Ge$_{0.3}$ epilayers.

The design of a 2DHG structure is much simpler since it only requires a SiGe layer grown under compressive strain on top of a relaxed Si layer to create a quantum well for holes. In Figure 6.9 the schematic layer diagram of such a 2DHG structure is shown. Directly after the growth of the high-temperature Si buffer on top of the substrate, a 300 nm relaxed Si buffer is grown followed by a 40 nm SiGe channel grown under compressive strain. The Ge content of the channel varies between 10% and 30% in the different structures. An undoped 6 nm Si spacer separates the SiGe channel from a 20 - 30 nm B doped (>$10^{18}$ cm$^{-3}$) Si supply layer. Unlike in 2DEG structures, no capping layers are necessary since the Si supply layer itself allows the formation of a protective oxide.

![Figure 6.9 Schematic layer diagram of a modulation-doped 2DHG structure (not to scale).]
From the description of the layer sequence of both structure types it is evident that most research so far has been concentrating on growing 2DEG structures because of their much more complex layer sequence and the need for a high-quality virtual substrate. However, despite being relatively easy to grow, 2DHG structures have a major drawback inherent in the growth process. Due to the segregation of Ge in SiGe, the quality of the interface between Si and SiGe (which represents the upper interface of the Si channel in 2DEG structures) is much better than that between SiGe and Si which is the upper interface of the SiGe channel in 2DHG structures. Since in normally doped samples with uniform quantum wells, the wavefunction resides in the vicinity of the upper interface, the transport behaviour in 2DHG samples is strongly influenced by the proximity of the inferior of the two interfaces of the SiGe channel. Possible solutions are the use of inverted structures where the doping occurs before the growth of the SiGe channel in order to pull the wavefunction towards the superior lower interface. The limitation of this approach, however, is the segmentation of the dopant into the spacer layer and the channel itself. An alternative, more promising solution is the growth of normally doped structures with graded quantum wells. A more detailed description of all samples with respect to the layer thickness, the Ge concentration and the doping density is given in chapter 7.

**PHOTOLITHOGRAPHY**

The photolithographic processing used for all samples essentially consists of two stages, the mesa definition and subsequently the deposition of the contacts. Both stages themselves comprise three main phases, the spinning of photoresist, the exposure to ultraviolet radiation through a mask and finally either the mesa etch or the metal deposition. An overview of all individual steps is given in Table 6.3. All steps of the photolithographic processing were performed in a Class 100 cleanroom.

Two different sets of pattern masks were used for the processing of the samples, the main difference being the length, the width and the length-to-width ratio of the Hall bars. The orientation of the Hall bars was chosen to be along the two main axes of the Si(100) wafer with one set of masks having additional Hall bars oriented diagonally to the main axes. Thus, it is possible to check the isotropy of the grown structure, whereas the homogeneity of the material can be analysed by placing Hall bars at different positions on the wafer, e.g. near the edge and close to the centre of the wafer. Proper wafer markings are a vital requirement for a precise alignment of the masks and therefore an optimal congruity between the positioning of the mesa and the contact area. All pattern masks used were light field mask, requiring the application of positive photoresist for the photolithography.

Prior to the photolithographic processing, all samples have to be cleaned carefully to obtain a clean surface free of any organic or inorganic pollution. First, trichloromethane or acetone is used as a solvent to remove any organic contamination from the surface of the sample followed by either methanol or isopropanol (IPA) to get rid of any inorganic contamination and the residues of the organic solvent. Afterwards, the cleaned sample is dried in nitrogen gas.
Positive photoresist is spun onto the surface of the sample at 4000 r.p.m. for 40 sec to ensure an even coating of suitable thickness. The sample is then baked at 90 °C for 30 min to dry the photoresist.

In the next step, the sample is mounted in a mask aligner where the pattern mask has to be aligned according to the wafer markings on the mask. After having been exposed to UV light for 10 sec, the sample is dipped in developer solution for about 45 sec where the exposed areas of the resist are softened. The concealed areas of the resist are then further hardened by again baking the sample for 30 min at 90 °C, giving these areas a better protection during the etching process.

Plasma etching of the mesa structure takes place in a reactive ion etching (RIE) machine with the plasma comprising 1% O2 and 50% CHF3. Etching rates vary depending not only on the input power and the magnetic field of the plasma but also on the structures being etched. All existing data for the etching rate is empirical and was derived exclusively from earlier experiments with SiGe 2DEG structures. Values for the etching rates of SiGe 2DHG structures, on the other hand, were not available prior to the etching process. For an input power of 160 W and a magnetic field of 50 mT, etching rates of around 50 nm/min were thus assumed for both structures. Presuming only minor variations in the etching rate for Si and SiGe, the error due to the different materials is negligible compared to the empirical error. Etching times of 5-8 min guarantee the mesas being etched deep into the respective buffer layers. Finally, the remaining photoresist has to be removed with acetone.

<table>
<thead>
<tr>
<th>step</th>
<th>definition of mesa</th>
<th>deposition of contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stage 1</td>
<td>stage 2</td>
</tr>
<tr>
<td>1</td>
<td>clean sample:</td>
<td>clean sample:</td>
</tr>
<tr>
<td></td>
<td>i) CHCl3 or acetone</td>
<td>i) CHCl3 or acetone</td>
</tr>
<tr>
<td></td>
<td>ii) CH3OH or IPA</td>
<td>ii) CH3OH or IPA</td>
</tr>
<tr>
<td>2</td>
<td>spin positive photoresist @ 4000 r.p.m. for 40 sec</td>
<td>spin positive photoresist @ 4000 r.p.m. for 40 sec</td>
</tr>
<tr>
<td>3</td>
<td>bake @ 90°C for 30 min</td>
<td>bake @ 90°C for 30 min</td>
</tr>
<tr>
<td>4</td>
<td>photolithography</td>
<td>photolithography</td>
</tr>
<tr>
<td>5</td>
<td>develop for 45 sec</td>
<td>develop for 45 sec</td>
</tr>
<tr>
<td>6</td>
<td>bake (optional) @ 90°C for 30 min</td>
<td>remove oxide with HF:H2O = 1:50</td>
</tr>
<tr>
<td>7</td>
<td>RIE</td>
<td>evaporation</td>
</tr>
<tr>
<td>8</td>
<td>remove resist with acetone</td>
<td>remove resist with acetone</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>annealing @ 500°C for 1 min</td>
</tr>
</tbody>
</table>

Table 6.3 Overview of the photolithographic processing.
For the deposition of the contacts, which constitutes the second stage, all steps from the first stage have to be repeated until the photolithography and the subsequent development of the photoresist. Just before being mounted inside the evaporation chamber, the sample is immersed in a diluted HF solution (~2%) for 30 sec to remove the surface oxide from the contact regions. The metals used to form the contacts of the 2DEG and 2DHG samples are Au:Sb and Al (with 1% Si to prevent spiking), respectively. Removing the remaining photoresist afterwards with acetone takes much longer than in the first stage because of the metal coating on top of the resist from the evaporation process. Finally, the contacts are annealed at approximately 500 °C for 1 min in a furnace or a rapid thermal annealer.

 Whereas there were no obvious problems concerning the contacts of the 2DEG samples, the deposition of the Al for the 2DHG samples created some complications resulting in Schottky-like rather than ohmic contacts in some cases. In other cases, the Al spiked through to the substrate forming pn-junctions due to the relatively low resistive substrate (1 - 10 Ωcm).

**Bonding**

Finally, each sample has to be mounted on a sample header where all contacts of the Hall bar are bonded to the pins of the header. Since various cryomagnetic systems with different inserts were used, it was necessary to employ more than one header type. The different header types and the corresponding type of bonding have been explained in the section describing the cryomagnetic system they are used in. Similar to the deposition of the contacts in the photolithographic processing, it is essential to ensure that the connection between the wires and the contacts of the samples are ohmic in nature.
7 LOW-TEMPERATURE MAGNETOTRANSPORT PHENOMENA

Magnetotransport phenomena have been investigated at high magnetic fields up to 15 T and low temperatures down to 30 mK on a variety of two-dimensional electron and hole gas systems in high-quality n- and p-type modulation-doped Si/Si$_1-x$Ge$_x$ heterostructures, respectively. Providing the basis for the subsequent analysis, the Shubnikov-de Haas (SdH) oscillations as well as the integer quantum Hall effect (IQHE) have been recorded in each sample at different temperatures and under various electric fields. Before the experimental data can be analyzed, it is necessary to convert the measured voltages into corresponding sheet resistivities which have a unit of Ω per square. To calculate the longitudinal resistivity $\rho_{xx}$, the voltage along the Hall bar has to be divided by the current and the length-to-width ratio of the Hall bar geometry where the length is equal to the distance between the voltage probes. The transverse resistivity $\rho_{xy}$ is simply given by the voltage across the Hall bar divided by the current along the Hall bar.

All of the information necessary for the calculation of the relevant parameters which fully describe the 2D systems in terms of their transport behaviour can be extracted from the low-temperature magnetotransport phenomena. Among these parameters is the effective mass which is expected to be constant in 2DEG samples but is anticipated to vary in 2DHG structures with the Ge content in the channel. Other vital parameters are the mobility, the sheet density and the transport and quantum relaxation times which are strongly correlated to the dominant scattering mechanism in the corresponding structure. Detailed information about the strength of the phonon processes in each sample can also be obtained from the magnetotransport experiments by establishing the deformation potential as well as the momentum and energy relaxation times associated with acoustic-phonon scattering.

At first, however, it is useful to analyse the SdH oscillations and the IQHE with respect to the distribution of the quantum states within the system. To provide an adequate quantum mechanical description of each system, the individual Landau levels and the corresponding splitting of the spin and valley states have to be identified. The filling order of the sub-levels as they move successively through the Fermi energy with decreasing energy is usually accomplished by assigning a quantum number to each peak and a corresponding filling factor to each minimum of the SdH oscillations (or plateau of the IQHE). By describing the quantum states of the 2D systems in this manner, potential anomalies like coincidences between sub-levels, ferromagnetic polarization or transitions into a Hall insulating phase can be detected, which may help to explain possible peculiarities in the present structures.

A few fundamental parameters can also be deduced directly form the curves of $\rho_{xx}$ and $\rho_{xy}$. The mobility, for example, can be estimated from the magnetic field at which the longitudinal resistivity starts to oscillate ($\mu B \geq 1$). Some limited qualitative
information about the quantum relaxation time can be extracted from the FWHM of the SdH oscillations ($\Gamma = h/2\tau_q$), and the sheet density can be calculated from the fundamental field ($n_s = g_\text{gs}eB_f/h$). When comparing the results of the analysis, exact knowledge of the structural composition in each sample is important as it provides the necessary feedback regarding the quality of the samples to further improve the transport behaviour of $n$- and $p$-type modulation-doped Si/\text{Si}_{1-x}\text{Ge}_x$ heterostructures.

### 7.1 Sample Structures

Of the whole range of samples investigated, a total of 14 samples display well defined SdH oscillations that are good enough to use for further analysis. Ten of these samples are $n$-type structures and four are $p$-type structures. According to their structural properties, all samples can be divided roughly into three different groups, two sets of 2DEG samples and one set of 2DHG samples. Although the layer sequence for both types of samples is essentially as shown in the schematic layer diagrams of chapter 6 (Figure 6.8 for $n$-type structures and Figure 6.9 for $p$-type structures), specific modifications have been introduced to find superior layer structures with respect to their transport behaviour by comparing the transport properties of the different samples.

The first batch of samples consists of four 2DEG samples with corresponding layer structures as shown in Table 7.1. In all four samples, the Ge content in the SiGe buffer is linearly graded from 5 to 35%, and the subsequent uniform SiGe buffer layer has a final Ge composition of 30%. In order to produce a virtual substrate with a minimum dislocation density at a minimum thickness, the thickness of both SiGe buffer layers has been varied while maintaining the grading profile. During growth of the supply layer, strong surface segregation of arsenic has to be taken into account resulting in an increasing As concentration towards the surface in a nominally uniformly doped supply layer. One possible way to avoid the ramping characteristic of this effect is to deposit enough arsenic directly after growth of the spacer layer but before the supply layer is grown. This technique helps to establish a more uniform doping profile as the delta deposition of the arsenic segregates to the surface during the growth of the supply layer. Consequently, the two doping techniques produce opposite doping profiles, the former having an increasing Ge content toward the surface, whereas the latter has a slightly decreasing Ge concentration. To compare both types of doping techniques, two samples are doped uniformly and the other two samples are delta-doped which therefore do not need a SiGe cap layer.

The transfer rate of carriers from the supply layer into the quantum well depends not only on the thickness of the spacer layer but also on the doping density and the thickness of the supply layer. As the doping density in a uniformly doped supply layer might vary slightly for a given growth temperature and pressure, it is more convenient to modify the transfer rate by changing the thickness of the supply layer. The sheet density of the delta deposition, on the other hand, can be determined quite
accurately and can easily be translated into an average bulk density for a given supply layer thickness. For a supply layer thickness of 50 nm, the given sheet densities of $2 \times 10^{12}$ cm$^{-2}$ and $1.5 \times 10^{12}$ cm$^{-2}$ calculate into average bulk densities of $4 \times 10^{15}$ cm$^{-3}$ and $3 \times 10^{15}$ cm$^{-3}$, respectively. Despite the average bulk densities of the nominally uniformly doped supply layers being an order of magnitude larger than those of the delta-doped supply layers, the densities close to the spacer layer are comparable as the effect of surface segregation reduces the doping density considerably towards the spacer layer.

### Table 7.1 Layer structure of all 2DEG samples in the first set.

<table>
<thead>
<tr>
<th>2DEG</th>
<th>BF197</th>
<th>BF213</th>
<th>BF313</th>
<th>BF314</th>
</tr>
</thead>
<tbody>
<tr>
<td>growth temperature ($^\circ$C)</td>
<td>520</td>
<td>520</td>
<td>510</td>
<td>510</td>
</tr>
<tr>
<td>linearly graded SiGe buffer (5 - 35% Ge) (nm)</td>
<td>2000</td>
<td>1500</td>
<td>1500</td>
<td>1500</td>
</tr>
<tr>
<td>uniform SiGe buffer (30% Ge) (nm)</td>
<td>500</td>
<td>500</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Si channel (nm)</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>SiGe spacer (nm)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>delta deposition (Torr)</td>
<td>–</td>
<td>–</td>
<td>27 min @ 2</td>
<td>20 min @ 2</td>
</tr>
<tr>
<td>SiGe supply (nm)</td>
<td>50</td>
<td>30</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>AsH$_3$ (Torr)</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>doping density</td>
<td>$1-5 \times 10^{18}$ cm$^{-3}$</td>
<td>$1-5 \times 10^{18}$ cm$^{-3}$</td>
<td>$2 \times 10^{15}$ cm$^{-2}$</td>
<td>$1.5 \times 10^{12}$ cm$^{-2}$</td>
</tr>
<tr>
<td>SiGe cap (nm)</td>
<td>6</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Si cap (nm)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Although the virtual substrates of the 2DEG structures in the first batch already have very low threading dislocation densities of below $10^7$ cm$^{-2}$ in the vicinity of the strained Si channel, a further reduction in dislocation densities is likely to reduce the large-angle scattering and thus increase the electron mobility in the quantum well. To explore this possibility, a second set of altogether six 2DEG samples has been investigated. Their layer structures, as shown in Table 7.2, differ only in the grading profile of the linearly graded SiGe buffer layer and the final Ge composition of the subsequent virtual substrate. To enable a direct comparison of the quality of the various virtual substrates, all growth parameters of the other layers have been kept unchanged.
A slight modification of the growth conditions to that of the first set is the introduction of growth interrupts of 30 seconds before and after the Si channel to adjust the growth temperature where necessary and the pressure in the gas lines to the appropriate values. As a result, more abrupt interfaces are expected and hence not only interfaces of higher quality but also samples with structures that are closer to the intended layer sequence. In all but one sample, Ge was deposited for ten seconds after the Si channel and before the SiGe spacer layer to increase the abruptness of the band alignment in the conduction band. Both modifications, the growth interrupt and the Ge pre-deposition, are expected to yield a higher interface quality, especially for the upper quantum well interface which is closest to the position of the wavefunction.

<table>
<thead>
<tr>
<th>2DEG</th>
<th>BF741</th>
<th>BF743</th>
<th>BF834</th>
<th>BF835</th>
<th>BF851</th>
<th>BF852</th>
</tr>
</thead>
<tbody>
<tr>
<td>growth temperature (°C)</td>
<td>520</td>
<td>520</td>
<td>725/550</td>
<td>725/550</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>linearly graded SiGe buffer (1000 nm) (%)</td>
<td>5 - 20 20 - 35</td>
<td>5 - 20 20 - 35</td>
<td>0 - 10 10 - 26</td>
<td>0 - 6.25 6.25 - 13 13 - 19.5 19.5 - 26</td>
<td>5 - 27</td>
<td>5 - 27</td>
</tr>
<tr>
<td>Uniform SiGe buffer (500 nm) (%)</td>
<td>30</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Si channel (nm)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>SiGe spacer (nm)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>SiGe supply (nm)</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>AsH₃ (Torr)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>doping density (cm⁻³)</td>
<td>1-5 × 10¹⁸</td>
<td>1-5 × 10¹⁸</td>
<td>1-5 × 10¹⁸</td>
<td>1-5 × 10¹⁸</td>
<td>1-5 × 10¹⁸</td>
<td>1-5 × 10¹⁸</td>
</tr>
<tr>
<td>SiGe cap (nm)</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Si cap (nm)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

¹Ge pre-deposition before the spacer layer

Table 7.2  Layer structure of all 2DEG samples in the second set. To improve the quality of the SiGe buffer layer, the grading steps and the final Ge composition of the virtual substrates have been varied. All other layers are nominally identical.

In two samples, the SiGe buffer layer and virtual substrate were grown at a much higher temperature (725°C) than the Si channel and the subsequent layers (550°C) to increase the growth rate and hence reduce the growth time of the virtual substrate considerably. One possible downside, however, could be the increased nucleation of dislocations in the buffer layer, leading to a higher density of threading dislocations in the virtual substrate. No delta doping of the supply layer has been employed in this set of samples to maintain the possibility of direct comparison although the
doping density is still subject to small variations. The last two samples of Table 7.2 are nominally grown identical in direct succession of each other to verify the capability of exactly reproducing a certain layer sequence, which should consequently display identical electrical properties.

The division of the 2DEG samples into two separate sets is intended to improve the structures according to the influence of long-range scattering (by remote ionized impurities) and short-range scattering (by threading dislocations and interface roughness). In the first set, the growth parameters in the supply layer are varied, whereas in the second set, the growth conditions of the virtual substrate are changed. It is thus possible to check for potential reductions of the small- and large-angle scattering by analysing the quantum and transport life times within as well as between the two sets. Hence, information can be obtained independently about the possible improvement of the virtual substrate, the quality of the interfaces and the effect of modulation doping.

The third and final set of samples consists of four 2DHG structures for which the layer sequence is shown in Table 7.3. Three of the four structures have uniform SiGe quantum wells, each with a different Ge concentration (10%, 20% and 30%), but otherwise, they are identical. By changing the Ge content in the SiGe channel it is possible to alter the hole effective mass which should be kept as low as possible to achieve the best transport behaviour. Furthermore, changing the Ge composition in the channel influences all of the four principal scattering mechanisms in a 2DHG structure. Alloy disorder and local strain fluctuations are directly connected to the Ge concentration, whereas interface charge and interface roughness depend strongly on the quality of the interface which in turn is dependent on the channel composition.

Displaying an altogether different structure, the fourth sample has a unique feature with a step-graded Ge content in the SiGe channel as simulated by the Poisson-Schrödinger solver of the band structure model in chapter 5. Since large-angle scattering is dominant in 2DHG structures, mainly from interface roughness and interface charge, utilizing a step-graded SiGe channel helps to pull the wavefunction away from the upper interface into the middle of the channel. Due to surface segregation of Ge in Si, the upper interface is of lesser quality than the much more abrupt lower interface. Given that the Ge content is ~25% at the position of the wavefunction, this structure should not only be compared to the uniform channels in general but also more specifically to the corresponding samples with a uniform channel of 20 and 30% Ge. In addition, the doping density and the supply layer thickness of the step-graded sample are reduced when compared to the uniform samples to decrease the carrier density within the channel and hence increase the mobility. Employing a back-gate bias or using inverted 2DHG structures to move the wavefunction away from the upper interface has not been taken into consideration because of difficulties in the fabrication of such structures. Instead, this novel approach of a step-graded SiGe channel to reduce large-angle scattering has been investigated. Ideally, all of the three different approaches should be compared to establish which technique is providing the best and most consistent results.
Low-Temperature Magnetotransport Phenomena

**Table 7.3** Layer structure of all 2DHG samples in the third set. The first three samples have a uniform SiGe channel with different Ge contents, ranging from 10 - 30%. The last sample has a step-graded SiGe channel to pull the wavefunction away from the upper interface.

Although more 2DHG samples with different layer structures have been grown, especially with different step-graded channels, all other samples displayed high resistances at low temperatures, most likely due to contact problems, or did not show any SdH oscillations at the available magnetic fields. The fact that even some 2DEG samples did not show any quantum transport phenomena at low temperatures is indicative of the problem to consistently reproduce high-quality samples with identical transport behaviour.

Two sets of masks have been used to process the samples photolithographically into standard Hall bars. The distance between two adjacent probes along the Hall bar constitutes the length of the Hall bar geometry and at 375 μm is identical in both masks. Corresponding values for the Hall bar width of 75 μm and 50 μm result in length to width ratios of 5 and 7.5, respectively.

### 7.2 SdH Oscillations and QHE

The emphasis of this section lies on the quantum mechanical description of the SdH oscillation, the integer quantum Hall effect and the assignment of quantum numbers \((N, \uparrow, \pm)\) to the corresponding electronic states. A thorough analysis with regard to the effective mass, the electronic transport properties, the dominant scattering mechanisms and the deformation potential is presented in the subsequent chapters. Due to space restrictions, only the longitudinal and transverse magnetoresistivities of the four 2DEG samples of the first set and those of the four 2DHG structures of the third set are explained in detail. Despite being calculated at a later stage, the
corresponding carrier mobilities and sheet densities are used in this section to explain the differences between the SdH oscillations of individual samples. Their values are presented in Table 9.1 (n-type samples) and in Table 9.2 (p-type samples). Since the SdH oscillations and Hall plateaux are best resolved at low temperatures, the magnetotransport traces from all but one sample are taken from base-temperature measurements in the dilution fridge. Depending on each individual cool-down cycle, the base temperature fluctuated slightly between 30 and 100 mK. At such low temperatures, the heating effect from the variation of the magnetic field with time on a metal is directly proportional to the gradient of the change \( \frac{dB}{dt} \), which may influence the temperature of the sample holder and thus the temperature of the sample itself. In order to minimize this unwanted heating effect, the sweep time at base temperature has to be increased from 10 to 30 min, which corresponds to a magnetic field change of 0.5 T/min. Unfortunately, the temperature measurement during a magnetic sweep is ineffective as the voltage of the diode used for the thermometry is magnetic field dependent. Hence, it is important to establish a stable temperature prior to the start of the magnetic sweep to guarantee a constant temperature during the measurements. An additional verification of the temperature is required each time the magnetic field is lowered back to zero field for a definitive confirmation of the required temperature invariance of the measurements.

### 7.2.1 Magnetotransport Studies of 2DEG Samples

In general, the experimental curves of the SdH oscillations are governed by the effective mass (Landau level splitting), the temperature (thermal broadening), scattering processes (collision broadening) and the sheet density (fundamental field). Any modifications of the SdH oscillations in different n-type SiGe structures at the same temperature are therefore caused by variations in the scattering processes and the sheet density. To assign the appropriate quantum number to each energy level, it is important to have an understanding of the magnitude of Landau level, spin and valley splitting. In particular, the spin splitting \( \Delta E_s = g^* \mu_B B \) is determined by the total magnetic field \( B \), whereas the Landau level splitting \( \Delta E_L = \hbar \omega_c = \hbar eB/m^* \) only depends on the field component \( B_n \) normal to the 2DEG. Hence, it is possible to increase the spin splitting with respect to the Landau level splitting by tilting the magnetic field, thus increasing the ratio \( B/B_n \). Compared to both spin and Landau level splitting, the valley splitting \( \Delta E_v \) is much smaller, and although its field dependence is not yet fully understood, it is assumed to be governed mainly by the normal field \( B_n \).

As illustrated in Figure 7.1, sub-levels of different Landau levels can overlap whenever the spin splitting (b) or the sum of spin and valley splitting (a) is an integer multiple of the Landau level splitting. These coincidences change the periodicity of the SdH oscillations as the corresponding peaks are shifted, merge and eventually split up again with rising tilting angle. A special coincidence with a ferromagnetic polarization of the spin states (c) is achieved whenever the spin splitting is an integer multiple of the sum of Landau level and valley splitting.
The occurrence of such coincidences can be used to determine certain quantum parameters. At lower magnetic fields where the value of the valley splitting is negligible, for example, the effective Landé factor $g^*$ can be deduced from the corresponding tilting angle $\theta$ at which a coincidence between successive valley-split Landau levels occurs. Since Landau level and spin splitting are identical in this situation ($g^* = \mu_B B = eB\cos\theta/m^*$), it is possible to calculate the value of the effective $g^*$ factor as the only unknown variable in this equation [118]. The effective Landé factor has been shown to be affected by exchange interactions which cause the value of $g^*$ to decrease with increasing filling factor and also with increasing sheet density [118-120]. In two-dimensional heterosystems with a very large $g^*$ factor, weak effects associated with the occurrence of coincidences might be observed without tilting the magnetic field.

**Sample BF197**

Due to experimental difficulties, sample BF197 could not be mounted in the dilution fridge and could therefore not be cooled down into the milli-Kelvin regime to obtain the best resolution of the SdH oscillations and the Hall plateaux. Hence, Figure 7.2 shows the magnetoresistivities of BF197 at 1.7 K, the lowest temperature reached in the VTI during this specific cool-down cycle. Well-pronounced SdH oscillations and coinciding Hall plateaux can be seen for the magnetotransport measurements of sample BF197 with clear signs of spin and valley splitting, as indicated by the arrows ($\uparrow, \downarrow$) and signs (+, -), respectively. Assigning filling factors to the minima of the SdH oscillations and the equivalent Hall plateaux is facilitated by the fact that the Hall resistivity has plateaux with values $h/e^2 \Omega$ at the corresponding filling factor $v$. Additionally, due to the $1/B$ periodicity of the SdH oscillations, a doubling of the filling factor requires the value of the correlated magnetic field to be halved. The value of the magnetic field at which the first Landau level is completely filled is called the fundamental field $B_y$. In the case of an $n$-type SiGe heterostructure with
spin-split and valley-split electronic states, $B_f$ is consequently the value of the magnetic field at filling factor $v = 4$ (equivalent to a four-fold degeneracy). In absence of any valley degeneracy, the sole spin degeneracy of the electronic states in a $p$-type SiGe system (and equally in $n$-type GaAs structures) results in the fundamental field occurring at filling factor $v = 2$ (equivalent to a two-fold degeneracy).

As expected for a 2DEG structure, sample BF197 has a fundamental field occurring at filling factor $v = 4$ with a magnetic field value of 5.8 T. Further main filling factors belonging to completely filled Landau levels appear at $v = 4i$, where $i$ as an integer. Lifting of the spin degeneracy occurs at even filling factors $v = 4i-2$ and splitting of the valley states can be seen at odd filling factors $v = 2i-1$. A subsequent allocation of the electronic states to the peaks of the SdH oscillations according to their quantum numbers simply follows the Landau level filling order of Figure 3.4. With decreasing magnetic field, the $(N = 0, \uparrow, -)$ state is occupied first, followed by the $(N = 0, \uparrow, +)$, the $(N = 0, \downarrow, -)$ and finally the $(N = 0, \downarrow, +)$ state, thus completely filling the first Landau level. As the magnetic field continues to decrease, the quantum states of all higher Landau levels are filled in identical manner. Due to the relatively high temperature of 1.7 K, the longitudinal resistivity vanishes only at 11.6 T or filling factor $v = 2$. Lifting of the valley degeneracy can be seen clearly at $v = 3$ and less plainly at $v = 5$. Splitting of the ferromagnetic spin occurs at $v = 2$, $v = 6$ and very faintly at $v = 10$. The main filling factors are visible up to $v = 24$ at a magnetic field of 0.97 T where the SdH oscillations of the longitudinal resistivity set in. First signs of Hall plateaux, however, start to emerge at 2.9 T or $v = 8$, with the only distinct plateaux in the transverse resistivity appearing at filling factors $v = 4$ and $v = 2$. Below 0.5 T, weak
localization gives rise to a negative longitudinal magnetoresistance, whereas at higher magnetic fields, electron-electron interactions contribute a positive magnetoresistance to the background of the SdH oscillations.

**SAMPLE BF213**

Having being carried out at the much lower temperature of 100 mK, which is lower by a factor of 17 compared to BF197, the magnetotransport measurements of sample BF213 predictably yield far better resolved SdH oscillations and Hall plateaux with visibly stronger spin and valley splitting than displayed in Figure 7.3. The SdH oscillations start at 0.51 T, which is equivalent to a filling factor of $v = 56$, and Hall plateaux can be seen as early as filling factor $v = 36$ at 0.79 T. Spin splitting sets in for half that filling factor at 1.58 T, whereas the valley degeneracy is first resolved at 4 T or $v = 7$. Initially, the longitudinal magnetoresistance vanishes at the main filling factor $v = 8$, which indicates a completely filled second Landau level. With increasing magnetic field, the magnetoresistance tends to zero for all subsequent even filling factors. The fundamental field of these measurements can be found at 7.1 T.

![Figure 7.3](image.png)

**Figure 7.3**  SdH oscillations and quantum Hall effect of sample BF213 at 100 mK and 20 nA. Lifting of the spin and valley degeneracy are clearly resolved at higher magnetic fields. Distinct Hall plateaux and narrow peaks of the SdH oscillations indicate minimal thermal broadening of the Landau level states.

A distinct feature of the Hall resistance traces of all measured 2DEG samples is the presence of a small peak just before each plateau in direction of increasing magnetic fields. Coinciding with the maxima of the SdH oscillations, these peaks can be ascribed to inhomogeneities in the material or alternatively to the size of the contacts of the Hall probes which can be circled by the current if they are too small. In both cases, the incorporation of a small longitudinal component into the measurement of the transverse magnetoresistivity is very likely, leading to a somewhat modified Hall
resistance. Although being another possible cause for the peaks in $\rho_{xy}$, a slightly asymmetric geometry of the Hall bar structure and thus a potential misalignment of opposite Hall probes can be disregarded due to the photolithographic processing of the samples, which leaves no margin for error in that respect. The most plausible explanation for the local peaks in $\rho_{xy}$, however, are exchange enhancement effects of spin and valley splitting associated with a slight tilt of the sample with respect to the direction of the magnetic field $B$ [119]. The resulting reduction of the normal component of $B$ causes the cyclotron energy to decrease, which induces level crossings of different Landau indices as a consequence of exchange enhancement.

One aspect of an entirely different nature is the uncertainty about the structural integrity of the strained layers in a 2D heterosystem over time. In particular, 2DEG structures with their complex layer arrangement and their thick virtual substrate are subject to a potential structural degradation with elapsing time, thus influencing the electronic transport properties and subsequently the magnetotransport experiments of such a system. As part of the programme, all samples were investigated for a second time after a period of 18 months to assess the almost certainly detrimental effect such a partial deterioration of the structure would have on the performance of the sample. In Figure 7.4, the magnetotransport response of sample BF213 is shown for measurements performed 18 months after the initial measurements. From this point onwards, the initial measurements are also referred to as BF213 (a), whereas BF213 (b) is used to identify the second set of measurements.

![Figure 7.4 SdH oscillations and quantum Hall effect of sample BF213 at 50 mK and 100 nA (18 months later). Fewer oscillations, a lower fundamental field, a higher zero-field resistivity and weaker spin and valley splitting suggest a decline in sheet density and more important a lower electron mobility due to broadening of the Landau level states.](image)

Although taken at an even lower temperature of 50 mK, the longitudinal and transverse magnetoresistance traces are clearly modified with visibly less
pronounced spin and valley splitting. Most prominent is the shift of the fundamental field from 7.1 down to 6.1 T, indicating a decrease in sheet density within the quantum well. Linked to the lower fundamental field is a reduction of the frequency in the $1/B$ periodicity and hence a noticeably lower number of oscillations in the longitudinal magnetoresistance. Furthermore, the oscillations set in at a much lower filling factor $v = 32$ and a higher magnetic field of 0.77 T, implying a drop in mobility of the electrons in the channel. Correspondingly, the Hall resistance starts to form plateaux at the lower filling factor $v = 20$ and the higher magnetic field of 1.22 T compared to the first measurements. Another indicator for a lower electron mobility is the raised value of the longitudinal resistivity at zero magnetic field, offsetting the otherwise beneficial influence of the reduction in sheet density. With increasing magnetic field, the onset of spin splitting can be seen earliest at $v = 14$ rather than at $v = 18$ for the initial measurements and the valley degeneracy is first lifted at $v = 5$ instead of $v = 7$. Because of the lower fundamental field, the $(N = 0, 7, +)$ state starts to be resolved within the limits of the magnetic field of the dilution fridge.

Comparing both sets of measurements, it is obvious that the structure of the 2DEG sample has noticeably degraded over time, affecting the transport behaviour of the sample significantly. Whereas the modification of the electron density does not necessarily represent a disadvantage regarding the transport properties, as a lower sheet density alone would almost certainly improve the electron mobility, it is essential for 2DEG structures that were specially designed in accordance with a band structure model not to vary too much from their given parameters in any direction. Otherwise, the expected behaviour for which they were intended could not be guaranteed any more in those devices. One reason for the decrease in density could be the partial relaxation of the strained Si channel which would result in a weaker confinement potential, thus reducing the transfer of charged carriers into the channel. Alternatively, if the surface potential is strong enough, it might be capable of depleting the carrier population and accordingly decreasing the sheet density significantly. Assuming a certain relaxation across the whole structure, another possible cause could be the introduction of threading dislocations in the supply layer which can act as additional traps for charged carrier. A far more serious problem, however, is the apparent decline in electron mobility, most likely attributed to the progressive relaxation of the graded buffer layer through the nucleation of dislocations. Eventually, two slightly different scenarios are plausible to emerge as a result of the relaxation, both capable of increasing large-angle scattering considerably. It is possible for the threading dislocations to nucleate through the previously relaxed buffer layer into the strained channel to establish a very potent scattering centre right in the path of the freely moving electrons, capable of trapping charged carriers. Additionally, the threading dislocations could form misfit dislocations at the interface to the quantum well, thus degrading the quality of the interface and increase the short-range scattering potential. In both cases, the ratio of transport and quantum relaxation time would be expected to decrease, whereby the threading dislocations in the channel would force a bigger drop in the ratio than the misfit dislocations at the interface. Considering that the decline in the mobility in sample BF213 is not overly critical, it is reasonable to assign this decline to a deterioration of the quality of the lower interface due to misfit dislocations.
SAMPLE BF313

Looking at the Hall resistance of sample BF313 in Figure 7.5, it is immediately obvious that within the magnetic field limit of 15 T, only Landau level states of filling factor $\nu = 4$ and higher are resolved. Consequently, the fundamental field lies at a much higher magnetic field of 12.6 T, about twice the value of the fundamental field of sample BF213, indicating a doubling of the sheet density. Together with a slightly higher zero-field resistivity of the SdH oscillations, the electron mobility of sample BF313 can be expected to be around half of the mobility of BF213 (a). The resolution of the well-pronounced SdH oscillations and the width of their peaks suggest a minimal broadening of the Landau level states similar to those in sample BF213 (a). The highest Landau level to be resolved has a filling factor of $\nu = 60$ at a magnetic field of 0.84 T, and plateaux in the Hall resistance start to appear at $\nu = 40$ or 1.26 T. The higher value of the magnetic field for the onset of the SdH oscillations compared to BF213 (a) is another indicator for a lower electron mobility in this structure. Spin splitting sets in at $\nu = 22$ and valley splitting can first be seen at $\nu = 7$.

![SdH oscillations and quantum Hall effect of sample BF313 at 50 mK and 100 nA. Landau levels are resolved up to a filling factor of $\nu = 60$. A longitudinal resistance vanishing at $\nu = 6$ (spin splitting) and almost vanishing at $\nu = 5$ (valley splitting) indicates a strong Zeeman effect and a distinct separation of the energy levels between degenerate valleys.](image)

As samples BF213 and BF313 mainly differ in the doping technique of the supply layer and hence the doping profile, yet not the average doping density, the higher sheet density of sample BF313 suggests a much closer proximity of the donor atoms to the spacer layer. Although an increase in sheet density might not be desirable, the comparison of both doping techniques shows that the delta doping technique used in BF313 promises a much more precise method of doping the supply layer to the required specification. This can be explained by the uncertainty of the width of the
effective spacer layer when employing the nominally uniform doping technique, as the surface segregation of the dopant creates a gradually increasing and rather uncontrolled doping profile. Delta doping, by contrast, produces a better defined distribution of the donors with a more accurate density close to the spacer, thus maintaining the intended effective width of the spacer layer.

Regarding the potential structural degradation of the sample over time, both sets of measurements, carried out within the referred time interval of 18 months, produced identical longitudinal and transverse magnetoresistance curves, demonstrating that this specific sample has a stable layer structure that is preserved over a substantial period of time. An obvious explanation for the preservation of the electron mobility can be found in the thickness of the uniform SiGe buffer layer which is twice as wide in BF313 than in BF213. A thicker totally relaxed buffer layer is better capable of compensating the propagation of threading dislocations towards the strained Si channel. Assuming a continued relaxation of the virtual substrate, however, even the wider buffer layer would ultimately be penetrated by the threading dislocations if given enough time. Hence, it might only be a matter of time before the electronic transport properties of sample BF313 start to degrade, unless the relaxation process stops before misfit dislocations are introduced at the interface to the quantum well. Apart from the partial relaxation of the strained Si channel, possible variations of the sheet density simply depend on the strength of the surface potential and its capacity to trap free carriers at the Si/SiO₂ interface where they are localized in space. Obviously, the protective oxide in sample BF313 is relatively stable, and its quality did not degrade within the given time span to seriously increase the surface potential. Again, increasing the time period could eventually result in a drop in sheet density, depending on the variation of the surface potential and thus the quality of the oxide. Another vital point of interest concerning the transfer of charged carriers from the supply layer into the channel is the percentage of electrically activated donors, i.e. the amount of donor atoms in the supply layer that are actually ionized.

**SAMPLE BF314**

Having virtually the same layer structure as BF313, magnetotransport measurements of sample BF314 yield almost identical traces for the longitudinal and transverse resistance, as shown in Figure 7.6. Due to a slightly lower doping density in the supply layer, the subsequently reduced sheet density causes a lower fundamental field of 11.5 T. Assuming the same type and strength of scattering mechanism in both samples, a lower sheet density is associated with a correspondingly higher electron mobility, which is confirmed by the fact that the oscillations set in at a marginally lower magnetic field of 0.77 T for the same filling factor \( v = 60 \). As in BF313, the first Hall plateau can be seen at \( v = 40 \) which lies at 1.15 T in sample BF314. Similarly, the spin degeneracy is faintly resolved at \( v = 22 \) for the first time whereas the lifting of the valley degeneracy starts at \( v = 7 \). Similar to BF313, magnetotransport measurements of sample BF314 after a time period of 18 months show no changes in the SdH oscillations or Hall resistance, demonstrating the absence of any detectable degradation of the 2DEG structure within the given time span. Yet again, it is possible for the deterioration of the sample to set in at a later point in time, but taking
into account that the samples have been grown long before they were first measured, the energy level of the threading dislocations and the relaxation energy level should have reached equilibrium within this time interval, thus suspending any further relaxation.

A prominent feature of samples BF313 and BF314 is the strong resolution of not only the spin splitting but also the valley splitting, once it has set in at high enough magnetic fields. This is indicative of Landau level states with minimal broadening due to scattering and hence a quantum relaxation time that is relatively high. Additionally, it suggests a rather strong energy separation of the spin-split and valley-split quantum states. In the region of the same cyclotron energy, the spin-split states of BF213 (a) exhibit comparable broadening and energy separation. However, although the valley-split states display a similar energy separation, the broadening of their energy levels seems to be much stronger, leading to less resolved peaks in the SdH oscillations at odd filling factors.

Another very distinctive characteristic present in all investigated n-type structures is the apparent spin dependence of the magnitude of the SdH maxima where the (↑) peaks seem to be enhanced compared to the surrounding (↓) maxima. This peculiar feature might be caused by a spin dependent scattering process associated with electron-electron scattering [120]. However, the analysis of the SdH oscillations is not affected by this phenomenon since the observed spin dependence vanishes when the background of the longitudinal resistivity is removed with a digital filter.

Figure 7.6  SdH oscillations and quantum Hall effect of sample BF314 at 50 mK and 100 nA. Similar curves to those of BF313 imply approximately the same electronic transport properties for both samples. Just like BF313, sample BF314 displays no signs of discernible structural degradation over time.
7.2.2 Magnetotransport Studies of 2DHG Samples

Numerous principal features distinguish the magnetotransport response of a 2DHG system from that of a 2DEG structure. Most importantly, the energy minimum in the valence band is located at the zone centre, due to which the quantum states in the Landau levels exhibit no valley degeneracy. However, rather complex interactions of the valence subbands produce a much more complicated electronic system to analyze. Strong spin-orbit coupling, the strong curvature of the populated HH subband, the proximity of HH and LH band and the eventual anti-crossing of the subbands away from the zone centre suggest slight modifications of the system's response as the hole effective mass and interaction effects vary with changing cyclotron energy. Having a significant influence on the 2DHG system, the effect of the Landé factor on the energy distribution of the quantum states is demonstrated in Figure 7.7 in which the \( g^* \) factor increases from left to right. Calculations of the effective \( g^* \) factor from tilted magnetic field measurements are problematic due to the highly anisotropic nature of \( g^* \), which requires very large tilt angles and consequently very high magnetic fields to achieve only a small change in the spin splitting. This behaviour is caused by the compressive strain which strongly decouples the HH and LH bands. According to the appropriately generalized Luttinger-Kohn Hamiltonian, the splitting between the HH states depends only on the perpendicular component of the magnetic field under these conditions [121, 122]. At large tilt angles, however, the spin splitting is expected to decrease slightly relative to the cyclotron splitting.

![Diagram](image)

Figure 7.7 Schematic representation of the spin splitting for \( \Delta E_z < \hbar \omega_c \) (a), for \( \Delta E_z = \hbar \omega_c \) (b) and for \( \Delta E_z > \hbar \omega_c \) (c) in which case the system is ferromagnetically polarized at \( v = 2N \).

In p-type SiGe structures, the Landé factor is magnetic field dependent and tends to be much larger (\( g^* = 6-8 \) [123]) than in a comparable n-type SiGe system where \( g^* \) typically has a value between three and four [118-120]. When Landau level and spin splitting are identical, the resulting spin degeneracy shown in (b) causes the minima in \( \rho_{xy} \) to vanish at even filling factors. Depending on the broadening of the spin-split Landau levels, the adjacent spin states of successive Landau levels in (a) and (c) might also be resolved as a single peak in the SdH oscillations. In these
circumstances, the SdH oscillations of the 2DHG system are dominated by minima associated with odd filling factors.

**SAMPLE BF216**

Looking at the longitudinal and transverse magnetoresistance traces of sample BF216 in Figure 7.8, a completely different characteristic to those of the investigated n-type samples is immediately noticeable. One of the most obvious differences in $\rho_{xx}$ is the zero-field resistivity which is in the region of 3 k$\Omega$ and therefore much higher when compared to the corresponding value of around 100 $\Omega$ in the n-type samples. Such a big discrepancy is not entirely unexpected as it indicates a far lower hole mobility due to an anticipated increase of short-range scattering from the interface. Further evidence of an enhanced scattering potential is the absence of dissipationless minima in $\rho_{xx}$ caused by a relatively small quantum relaxation time and thus considerable broadening compared to the energy separation of the Landau levels. The negative longitudinal magnetoresistance for magnetic fields below 0.3 T and the positive contribution to the background for higher magnetic fields can be attributed to a combination of weak localization and other interaction effects.

![Figure 7.8 SdH oscillations and quantum Hall effect of sample BF216 at 30 mK and 10 nA. Minima of $\rho_{xx}$ appear at odd filling factors and at higher magnetic fields a phase transition to an insulating phase can be seen. Landau level indices are shown to explain the dominance of odd filling factors.](image)

When assigning filling factors to the minima of the SdH oscillations, however, the most prominent feature emerges to be the dominance of minima associated with odd filling factors. This distinctive behaviour is a well-established property of p-type SiGe systems in which the effective Landé factor is sufficiently large to cause the spin splitting to be comparable to the Landau level spacing. In such a system, the $(N,\uparrow)$ state is very close in energy to the $(N+1,\uparrow)$ state, forcing a spin degeneracy of up and
down spin states of successive Landau levels at even filling factors. The splitting between the two adjacent spin states of successive Landau levels is only fully resolved for magnetic fields above 9 T where the $v = 2$ integer quantum Hall state is clearly defined. The strong background contribution in $\rho_{xx}$ causes the corresponding minimum to be raised considerably above its expected value of zero resistivity. Due to the strong spin-orbit coupling, hole-hole effects cause an enhancement of the spin splitting and an oscillating effective $g^*$ factor as the magnetic field increases [82]. At high magnetic fields, the spin splitting can thus become sufficiently large that it exceeds the cyclotron energy and consequently, the system might become ferromagnetically polarized at filling factor $v = 2$. For even higher fields, the system experiences a phase transition to a Hall insulating phase as indicated by the anomalously large peak of $\rho_{xx}$ in the region $2 > v > 1$. The occurrence of a Hall insulating phase at filling factors $v \leq 2$ is a well-established phenomenon in $p$-type SiGe structures which can be controlled by varying the sheet density and also by tilting the magnetic field [121, 122, 124-129]. It has been suggested that the existence of a Hall insulating phase in $p$-type SiGe is strongly correlated to many-body interactions which are also responsible for the ferromagnetic polarization of the system. However, the Hall insulating phase is suppressed when this ferromagnetic polarization persists into the $v = 1$ integer quantum Hall state, possibly because a triplet exciton instability may be energetically favourable under these conditions [122, 130].

Since $p$-type SiGe systems exhibit only spin degeneracy and no valley degeneracy, the fundamental field of each sample lies at filling factor $v = 2$, where the first Landau level is completely filled. At 9.95 T, the fundamental field of BF216 implies an even lower sheet density than in sample BF213, and consequently, only few oscillations are resolved. The onset of the oscillations at 1.17 T coincides with the first appearance of a Hall plateau at a comparatively low filling factor $v = 17$. The Hall resistance displays an unusual feature at filling factor $v = 3$ where it deviates slightly from its expected value of $h/ve^2 \Omega$, which can be attributed to localization effects [121]. Furthermore, anomalous local maxima in $\rho_{xy}$ can be seen between the odd integer plateaux where localization effects are much less important. This behaviour might be caused by oscillations in the Fermi level due to Landau level mixing which could be particularly important in $p$-type SiGe structures because of the large $g^*$ factor [121, 126, 127]. As spin splitting and cyclotron energy are comparable, the mixing occurs between spin states of successive Landau levels which consequently have different cyclotron radii. Another anomaly in the magnetotransport can be seen at $v = 8/3$ where $\rho_{xx}$ exhibits a dip and $\rho_{xy}$ develops a plateau like feature. This is not likely to be associated with a feature of the fractional quantum Hall effect but is rather thought to be caused by Landau level mixing effects in conjunction with exchange effects [121, 126, 127]. Interestingly, the value of $\rho_{xy}$ at $v = 2$ is much higher than $h/2e^2$ and does not remain constant during the Hall insulating phase as would be expected, which is mainly attributed to the fact that $\rho_{xy}$ is difficult to measure and susceptible to sample imperfections. Due to $\rho_{xx}$ admixture effects, increasing contributions of the longitudinal component in $\rho_{xy}$ might also play a vital role in the observed deviations of the Hall resistance.
SAMPLE BF217

A much more pronounced dominance of the longitudinal component in the Hall resistance can be seen for the magnetotransport curves of sample BF217 in Figure 7.9, this time, however, over the whole magnetic field range. Such a strong mixture of $\rho_{xx}$ and $\rho_{xy}$ components implies either a very inhomogeneous sample or extremely untidy contact regions and makes it impossible for the Hall resistance to be analyzed correctly. Strong inhomogeneities in the sample are the more plausible cause for this phenomenon, since different samples from the same wafer exhibit similar behaviour. Yet, the contacts of all p-type and n-type samples were made using the same two mask sets, and the process of evaporating Al as the contact material was identical for all p-type samples.

A sheet density approximately half of that of sample BF216 can be deduced from the fundamental field lying at 5.6 T, resulting in even less Landau levels being resolved and hence fewer oscillations in the longitudinal magnetoresistance. Consequently, the oscillations of $\rho_{xx}$ set in at a very low filling factor of $v = 9$, again about half the filling factor at which the longitudinal magnetoresistance of sample BF216 starts to oscillate. Assuming identical hole effective masses, a magnetic field value of 1.23 T for this filling factor suggests a similar hole mobility for both samples, however, almost identical zero-field values of $\rho_{xx}$ and a sheet density about half of that of BF216 suggest a hole mobility roughly twice as high.

As for sample BF216, the oscillations in $\rho_{xx}$ are governed by minima linked with odd filling factors, whereby splitting between the two spin states of successive Landau levels is partially resolved at filling factor $v = 2$, and signs of incipient removal of the
spin degeneracy can be seen at \( v = 4 \) in the form of a shoulder on side of the \( \rho_{xx} \) peak. Although the background component of \( \rho_{xx} \) seems to be of the same order in both samples, the resolution of the peaks of the oscillations in BF217 are comparatively faint and the value of the minima barely reaches the zero-field value but follows the positive contribution of the background. Paradoxically, a possible explanation for such a bad resolution can be found in a presumed enhancement of the Landau level spacing, which under normal circumstances is associated with a better resolution of the peaks. Adopting a similar effective \( g^* \) factor for sample BF217 and thus the same spin splitting as in sample BF216, an increase in the separation energy between adjacent Landau levels would pull the degenerate spin states further apart. As a consequence, the extended states of two successive spin states whose degeneracy has just been lifted are still too close together to be clearly resolved but not close enough to be truly degenerate and thus be resolved as a single peak. Additionally, spin states of the same Landau level are much closer in energy, impeding the formation of pronounced minima at odd filling factors. However, an enlarged Landau level spacing leads to the minima at even filling factors starting to be resolved, although faintly, at lower magnetic fields. As the Landau level spacing is directly proportional to the cyclotron energy, \( \omega_c \) has to be raised to enhance the energy separation. An increase of the cyclotron frequency for a given magnetic field can be achieved by a reduction in the hole effective mass. Consequently, the hole effective mass of sample BF217 should be lower than the effective mass of BF216. Of course, the effective \( g^* \) factor itself and hence the spin splitting is dependent on the effective mass and the sheet density, rendering the whole problem much more complex.

**Sample BF218**

The larger number of well-pronounced SdH oscillations of sample BF218 in Figure 7.10 indicates a much higher sheet density, in fact the highest of all investigated \( p \)-type samples. This assumption is supported by a very large fundamental field of 22.2 T which is about twice that of sample BF216 and four times that of BF217. As established before, all minima coincide with odd filling factors, consistent with a \( p \)-type SiGe system with a large effective \( g^* \) factor. Even filling factors are only visible at \( v = 4 \) where the hint of a shoulder on the side of the peak represents the incipient lifting of the spin degeneracy. Clearly, the background contribution in \( \rho_{xx} \) is much smaller than in the previous two samples and the minimum at \( v = 3 \) tends to zero. However, a stronger increase of the background at higher magnetic fields as in BF216 is still possible.

The longitudinal magnetoresistance starts to oscillate at a comparatively high magnetic field of 1.98 T which corresponds to a filling factor of \( v = 23 \) in this sample. Together with a higher sheet density and a similar zero-field resistivity, the late onset of the oscillations in \( \rho_{xx} \) suggest a much lower hole mobility than in both previous \( p \)-type samples. Plateaux in the Hall resistance are first resolved at filling factor \( v = 22 \) or 2.02 T, which represents a big paradox as the plateaux do not coincide with the minima of \( \rho_{xx} \) at odd filling factors but rather with its peaks at even filling factors. The first Hall plateau associated with an odd filling factor is faintly resolved at \( v = 7 \) with a better resolution at \( v = 5 \), and eventually, \( \rho_{xy} \) displays a proper plateau at the
odd filling factor $v = 3$. This peculiar behaviour of $\rho_{xx}$ at lower magnetic fields where the occurrence of the plateaux is essentially in antiphase to the oscillations in $\rho_{xx}$ has also been reported for a similar $p$-type SiGe structure [121, 122].

![Graph showing SdH oscillations and quantum Hall effect](image)

Figure 7.10 SdH oscillations and quantum Hall effect of sample BF218 at 30 mK and 50 nA. The oscillations of $\rho_{xx}$ are dominated by minima associated with odd filling factors whereas the Hall plateaux coincide with even filling factors.

As the only $p$-type sample to be studied twice within the referred time period of 18 months, it is important to evaluate the structural integrity of sample BF218 carefully since it is studied representatively of all the 2DHG structures. Although it would have been preferable if additional magnetotransport measurements could have been carried out on all samples after a period of time, BF218 has the highest Ge content in the channel and therefore should have the biggest potential for structural degradation. Except for a slightly higher zero-field resistance in the second set of measurements, the trace of the longitudinal resistance is almost identical to that of the initial measurements. This result suggests a certain decline in the hole mobility, whereas the sheet density remains unchanged over time. Consequently, the lower hole mobility has to be explained by an increase of the large-angle scattering potential, concurring with a small deterioration of the compound in the channel through partial relaxation. Only a small degree of relaxation would create enough threading dislocations to disrupt the flow of holes severely and thus reduce their mobility as the nucleation of threading dislocations occurs directly in the active region of the structure. However, the width of the peaks in $\rho_{xx}$ are unaffected by the larger scattering potential, implying that the quantum relaxation time of the system has stayed constant over time. Having a lower Ge concentration in the channel, the other two $p$-type samples are expected to experience less relaxation and thus should exhibit an even lower degree of structural degradation over time.
SAMPLE BF638

Despite the low temperature at which the measurements were carried out, the oscillations of the longitudinal magnetoresistance in Figure 7.11 are barely resolved sufficiently to analyze sample BF638 with respect to its electronic transport properties. Being the only 2DHG structure with a graded Ge content in the channel, BF638 displays a zero-field resistance that is nearly four times as high as in the previous p-type samples with a constant Ge concentration. This indication of a comparatively low hole mobility is further supported by profoundly broadened peaks in the oscillations of $\rho_{xx}$ suggesting a rather low quantum relaxation time and hence a strong scattering potential. Furthermore, the considerable broadening of the Landau levels reduces the mobility gap between the extended states of successive Landau levels to a minimum, leading to a lower resolution of the SdH oscillations and less distinct Hall plateaux at odd filling factors.

A fundamental field at 18.2 T indicates a sheet density only marginally lower than that of BF218 but almost twice as high as BF216. However, the broadened Landau levels do not allow for many oscillations, in fact, BF638 has even fewer oscillations than BF217, the 2DHG sample with the lowest sheet density. Because of the low mobility, the SdH oscillations and the formation of Hall plateaux start at the relatively high magnetic field of 5.4 T, which corresponds to a filling factor of $\nu = 7$ due to the low resolution of the oscillations. As in sample BF216, the system enters the insulation phase between the filling factors $\nu = 1$ and $\nu = 3$. 

Figure 7.11 SdH oscillations and quantum Hall effect of sample BF638 at 30 mK and 5 nA. A high zero-field resistance and a late onset of the oscillations indicate a very low hole mobility. Relatively broad peaks and a poor resolution of the oscillations suggest a low quantum relaxation time.
7.3 SUMMARY

Magnetotransport measurements have been performed at high magnetic fields up to 15 T and low temperatures down to 30 mK on a variety of high-quality \( n \)- and \( p \)-type Si/Si\(_{1-x}\)Ge\(_x\) heterostructures grown by GSMBE. Within this framework, longitudinal and transverse resistivities have been recorded in each sample at different temperatures and under various electric fields in order to characterize the corresponding two-dimensional electron and hole gas systems. The SdH oscillations in the longitudinal resistivity of the \( n \)-type structures are well pronounced and exhibit dissipationless minima at filling factors corresponding to the main Landau level states. In addition, the longitudinal resistivity also vanishes at the spin-split and valley-split states, indicating a 2DEG of high quality in the quantum well. Coinciding with the minima of the SdH oscillations, the quantized Hall plateaux in the transverse resistivity are clearly resolved and display the expected value of \( h/ev^2 \). Characteristically, the SdH oscillations of the 2DHG samples are dominated by minima associated with odd filling factors. This well-established property of a \( p \)-type SiGe system is caused by the fact that in such a system, spin splitting is comparable to the cyclotron energy due to a sufficiently large effective Landé factor. Hence, the up and down spin states of successive Landau levels at even filling factor are forced into spin degeneracy. Other types of anomalies, like a phase transition to the Hall insulating phase at filling factors \( v \leq 2 \) and an antiphase behaviour of SdH effect and IQHE at lower magnetic fields, have also been observed within this set of \( p \)-type samples. The process of structural degradation of both \( n \)- and \( p \)-type samples over time is strongly dependent on the respective layer structure of each sample, and in most cases, it does not change the magnetotransport phenomena to a large extent.
8 Effective Mass

Apart from the rudimentary quantum mechanical description of the modulation-doped n- and p-type Si$_{1-x}$Ge$_x$ heterostructures in the previous chapter, it is important to analyse the present samples with respect to their electronic transport properties and thus establish a qualitative and quantitative characterization in order to improve the quality of the samples by varying their structural composition. Of the relevant parameters, the effective mass $m^*$ has to be determined first because the calculation of all other transport properties requires the knowledge of the exact value of $m^*$.

The presence of uniaxial strain in SiGe heterostructures and its effect on the band structure is one of the most important factors to determine the properties of a two-dimensional heterosystem. One particular aspect is the dependence of the effective mass on the existing strain as the electron and hole effective mass is governed by the characteristic of the conduction band and valence band states in $k$-space, respectively. Being an important parameter influencing the oscillations in $\rho_{xx}$ through the cyclotron frequency, any further analysis of the SdH oscillations therefore requires the knowledge of the electron or hole effective mass. According to eqn. (3.66), the cyclotron frequency influences the thermal damping factor as well as the exponential envelop function of the oscillations. Consequently, the electron or hole effective mass of each sample has to be determined before the electronic transport properties of the corresponding structures can be analyzed. In n-type structures, the electron effective mass is relatively insensitive towards uniaxial strain since the conduction band edge states are merely shifted in energy but maintain their spherical character with an unchanged parabolic curvature of the bands. Therefore, the electron effective mass can be assumed to remain at a constant value of $m^* = 0.19 m_0$ in all n-type SiGe systems, irrespective of the Ge content in the virtual substrate. The valence band structure, however, is altered considerably in the presence of uniaxial strain. Together with a shift in energy, a large change in the curvature of the valence bands is a typical feature associated with the effects caused by uniaxial strain in p-type SiGe structures. The dependence of the hole effective mass on the Ge content in the channel is thus not only due to the different values of $m^*$ in pure Si and Ge but also because the magnitude of the strain is correlated to the Ge composition.

8.1 Temperature Dependence of SdH Oscillations

In order to extract an effective mass from the oscillations in $\rho_{xx}$ it is useful to derive a term which describes only the oscillatory component of the longitudinal resistivity. Such a term, purely relating to the amplitude of the SdH oscillations, is given by the following expression [96, 98]
\[ \Delta \rho_{xx} = 4 \rho_0 \frac{\chi}{\sinh \chi} \exp \left( - \frac{\pi}{\omega_c \tau_q} \right), \quad (8.1) \]

where \( \rho_0 \) is the zero-field resistivity, \( \omega_c \) the cyclotron frequency and \( \chi / \sinh \chi \) the thermal damping factor with

\[ \chi = \frac{2 \pi^2 k T}{\hbar \omega_c}. \quad (8.2) \]

From eqns (8.1) and (8.2) it is obvious that the effective mass can be easily calculated from the cyclotron frequency in the thermal damping factor for SdH oscillations measured at different temperatures. Hence, measurements of the longitudinal resistance of each sample have to be carried out at various temperatures while the current and thus the electric field across the Hall bar structure has to be kept constant. To illustrate this principle, Figure 8.1 shows simulated SdH oscillations of a SiGe 2DEG system in a temperature range between 0.03 K and 4 K at a constant electric field. The resistivities are calculated in agreement with eqn. (3.66), thus omitting any effects of the lifting of the spin or valley degeneracy on the oscillations.

To guarantee a good comparability to real 2D systems, the generated oscillations should have a close similarity to those of the investigated samples. Accordingly, the parameters used in the simulation represent a typical SiGe 2DEG structure with an electron effective mass \( m^* = 0.19 m_0 \), a quantum relaxation time \( \tau_q = 1 \) ps, a sheet density \( n_s = 6 \times 10^{11} \text{ cm}^{-2} \) and a zero-field resistivity \( \rho_0 = 100 \m\Omega \). As expected from
eqn. (8.1), the amplitude of the oscillations increases continuously with decreasing temperature and in accordance with the influence of the thermal damping factor. Consequently, the effective mass at the Fermi level can be extracted from the variation of the amplitude of the SdH oscillations with temperature.

Expeditiously, the longitudinal resistivity at the lowest temperature is chosen to act as the reference to the SdH oscillation at all other temperatures. The amplitudes of the oscillations at each temperature can then be correlated to those at the same magnetic field at the reference temperature according to eqn. (8.1). Taking the ratio of the amplitudes at a given temperature and those at the reference temperature leads to the expression

$$\frac{\Delta \rho_{xx}(T)}{\Delta \rho_{xx}(T_0)} = \frac{\frac{\xi T}{\sinh(\xi T)}}{\frac{\xi T_0}{\sinh(\xi T_0)}} = \frac{T}{T_0} \frac{\sinh(\xi T)}{\sinh(\xi T_0)}$$

where $T_0$ is the reference temperature and $\xi T = \chi$ to emphasize the importance of the temperature dependence. The exponential envelope function does not appear in eqn. (8.3) since it is independent of temperature.

Subsequently, the effective mass can be obtained by plotting the ratio of the amplitudes against the temperature as shown in Figure 8.2 for the simulated 2DEG system. Because of the perfect symmetry of the oscillations, values for both minima and maxima can be selected to calculate the ratio. In real 2D structures, the oscillations might not be sufficiently symmetric due to interaction effects so that only values at the minima should be used. Fitting the expression of eqn. (8.3) to the ratios of the amplitudes at constant magnetic field for different temperatures yields the value for the effective mass as the only adjustable parameter in this expression.

Figure 8.2 Amplitude ratio of minima and maxima values at temperatures between 0.03 K and 4 K. The effective mass can be obtained by fitting eqn. (8.3) to the data points at constant magnetic field.
Having thus determined the electron effective mass at different magnetic fields, it is useful to plot the effective mass against the magnetic field, as can be seen in Figure 8.3, to establish a possible dependence of the effective mass on the magnetic field. Due to the lack of any built-in nonparabolicity effects in the model, the fitting process reproduces the initial value of $m^*$ used for the simulation with no dependence on the magnetic field. However, the fitting process also generates a very small calculation error of $m^*$ which tends to increase with decreasing magnetic field. The presence of a fitting error for a simulated ideal 2D system demonstrates the extreme sensitivity of this process to the slightest of variations in the extrema of the SdH oscillations, which suggests much bigger error margins when calculating $m^*$ in real heterostructures. Hence, a detailed error estimation is required for the calculation of the effective mass in the present 2DEG and 2DHG samples.

When analysing the effective mass of a real 2D system, it is imperative to only use minima or maxima at magnetic fields below the onset of spin splitting to guarantee the validity of eqn. (8.3). Another important aspect is the choice of the temperature range. A larger temperature range produces more data points which is beneficial for the accuracy of the fitting process, but at the same time, it also reduces the number of maxima or minima for which the ratio can be calculated because the resolution decreases with increasing temperature, rendering amplitudes at lower magnetic fields useless. Hence, an appropriate choice of the maximum temperature has to be made for each set of measurements individually by establishing a suitable balance between the accuracy of the effective mass and the magnetic field range for which an effective mass is being calculated.

Alternatively, the effective mass can be calculated directly from eqn. (8.1) by plotting $\ln(\Delta \rho_{\text{osc}}/\rho_0)$ vs $\ln(\chi/\sinh \chi)$ in the given temperature range for values of the magnetic field corresponding to the minima and maxima of the oscillations. Such a plot is shown in Figure 8.4 where the linear relationship between all data points at the same minimum or maximum and thus identical magnetic field is clearly visible. Modifying the value of the effective mass as the only adjustable parameter changes the gradient of the straight line, whereby the correct effective mass of the system is obtained for a
gradient of unity. Since each straight line crosses the ordinate at an offset of \(-\pi/\omega_0 \tau_q\), it is also possible to calculate the quantum relaxation time from this type of plot.

![Figure 8.4 Plot of ln(\(\Delta \rho_{xx}/\rho_0\)) vs ln(\(\chi/\sinh \chi\)) in the temperature range between 0.03 K - 4 K for various magnetic fields. Fitting eqn. (8.1) to the data points at constant magnetic field yields the effective mass for a gradient of unity.]

When evaluating the effective mass of both the 2DEG and the 2DHG samples, the first method is used to calculate the actual value the effective mass, whereas the second approach allows the results to be verified by checking the divergence of the gradient of the curves from unity.

### 8.2 Digital Filtering

In the preceding treatment of the temperature dependence of the longitudinal resistivity, any influence of electron-electron interaction effects on the amplitude of the oscillations has been neglected. The zero-field resistivity constitutes the only background contribution to the oscillations which is independent of \(B\) and can therefore easily be removed by subtracting \(\rho_0\) from the measured dataset. In real 2D systems, however, electron-electron interactions contribute an additional \(B^2\) component to the background. Figure 8.5 shows the effects of an added background contribution that is quadratic in \(B\) on the oscillations in \(\rho_{xx}\) for the same 2DEG system as analyzed before with otherwise identical parameters.

For that reason, the data has to be sent through a high-pass filter to eliminate all non-linear background components before the oscillations can be analyzed. In practice, either an analog filter or alternatively a digital filter can be used to subtract the low
frequency background from the actual oscillations at higher frequencies. However, employing an analog filter has several weaknesses concerning the practicality and the quality of the filter response. As the measured signal has to be fed directly through an RC filter component for every individual measurement of $\rho_{xx}$, the corresponding band edge frequency of the high-pass filter has to be estimated before the power spectrum of the oscillations is known. Without exact knowledge of the oscillations prior to the measurement, it is difficult to set up the analog filter correctly, and post processing of the data is not possible. In addition, filter-induced phase lags are common in analog filters, and potential distortions of the signal are likely to occur as a result of a limited frequency response of the RC component.

**Figure 8.5** Simulated SdH oscillations of the same 2DEG system as shown in Figure 8.1 with identical parameters but with an additional $B^2$ background component.

All these shortcomings of an analog filter can be avoided by employing a digital filter which can be used on the data set at any time after the initial measurement without data loss of the original signal. Due to the ability to acquire detailed information about the power spectrum of the measured signal prior to the filtering process, the frequency of the filter can be chosen precisely to pass only the desired fundamental signal. Digital filters do not introduce phase lags or distort the signal, and the quality of the filter response is only determined by the resolution of the input signal and hence by the sampling rate of the measurement. Depending on the background component of the measured signal, two different types of digital filter can be used. If the background signal and the fundamental signal are clearly distinguishable in the power spectrum, a digital filter can be applied in the frequency domain by simply cutting of the power spectrum of the background component. For a more complex background, however, it is necessary to apply a digital filter that works in the time domain by using a filter algorithm that weights the original input data with a
number of filter coefficients. Although it takes a lot longer to compute the output data, this method is much more accurate and thus delivers far better results.

8.2.1 Digital Filtering in Frequency Domain

In order to transform the oscillations of $\rho_{s}$ into the frequency domain, the resistivity has to be plotted against the reciprocal magnetic field as $\rho_{s}$ is periodic in $1/B$. Subsequently, the data points of the oscillations have to be interpolated at evenly spaced intervals in $1/B$, whereby the total number of data points has to be an integer power of two to allow for a fast Fourier transform (FFT) to create the power spectrum of the SdH oscillations which provides a variety of information about the investigated 2D system. Assuming only one occupied subband, no background component and no spin or valley splitting, an ideal 2D system exhibits exactly one pronounced peak with its maximum being located at the fundamental field. In Figure 8.6, the power spectrum of the longitudinal resistivity of the simulated 2DEG system of Figure 8.5 is shown at positive frequencies for values of $\rho_{s}$ before the onset of spin splitting with the contribution of a non-linear background component.

![Image of power spectrum](image)

*Figure 8.6* Power spectrum of SdH oscillations of a 2D system with a single occupied subband, a non-linear background component and no spin or valley splitting (left). A digital high-pass filter cuts off all lower frequencies, leaving only the fundamental field and higher frequencies (right). Due to the symmetry of the power spectrum with respect to the ordinate, only positive frequencies are shown.

In this specific case, in which the spectrum of the background is visibly separated from that of the fundamental field and the amplitude falls close to zero at a cut-off frequency between the two peaks, a straightforward digital filter can be applied in the frequency domain by setting the amplitude of all frequencies associated with the background below the cut-off frequency to zero. Finally, the filtered power spectrum has to be transformed back into the time domain using an inverse FFT to obtain the pure oscillatory component of $\rho_{s}$ as a function of the reciprocal magnetic field. Figure 8.7 shows the oscillations of the simulated 2DEG system of Figure 8.5 after the data has been sent through a digital filter in the frequency domain.
When preparing the original data for the transformation into the frequency domain, the sampling rate of $\rho_{xx}$ as a function of the inverse magnetic field has to be high enough to resolve each period of the oscillations at sufficient accuracy. At the same time, a high resolution of the peaks in the frequency domain is necessary for the identification of the spectrum of the fundamental field and the background to apply the digital filter. As the resolutions in the two domains are inversely proportional, a high resolution in the time domain yields a low resolution in the frequency domain, i.e. doubling the resolution of a function in the time domain causes the resolution of its Fourier transform to halve. Additionally, the frequency of the fundamental field should be well below the Nyquist critical frequency $f_s$ which is exactly half the sampling frequency. Hence, the number of data points and the data range have to be chosen carefully when calculating $\rho_{xx}$ as a function of $1/B$ in order to establish a sampling rate high enough to guarantee an adequate resolution of the oscillations and a sufficiently high Nyquist critical frequency, yet low enough to sustain an acceptable resolution of the power spectrum.

### 8.2.2 DIGITAL FILTERING IN TIME DOMAIN

The conditions for the application of a digital filter in the frequency domain require the oscillations in the time domain to be well pronounced and to have a clearly higher frequency in the reciprocal magnetic field than the background component. In real 2D systems, however, these conditions are rarely met because the range of the spectrum associated with the background often extents into the higher frequencies, thereby overlapping with the spectrum of the fundamental field as shown in Figure
8.8. Applying a digital filter in the frequency domain at the minimum between the two peaks of the background and the fundamental field would introduce additional frequencies in the time domain corresponding to the sharp vertical edge of the digital filter. Apart from choosing a smoother filter function which would have to be specially defined for each individual case, the most convenient solution to this problem is the application of a digital filter directly in the time domain [131, 132].

The algorithm for a such a filter is assumed to be that of a linear weighting of the input data values where the weighting factors are the coefficients of the filter. In its basic version, the output data \( y_n \) are calculated from the original input data \( x_n \) by

\[
y_n = \sum_{k=-N_f}^{N_f} b_k x_{n-k}
\]

where \( b_k \) are the filter coefficients, \( 2N_f \) is the span of the filter in samples and \( 2N_f+1 \) is the total number of terms in the filter. As the output \( y_n \) is simply the convolution of the input \( x_n \) with the filter coefficient sequence \( \{ b_k \} \), the filter design problem is reduced to the determination of a set of weighting factors \( \{ b_k \} \) that fulfils the desired filter characteristic. Figure 8.8 demonstrates the principle of how a digital low-pass filter in the time domain can be applied by displaying the power spectrum of the input (SdH oscillations) on the left and that of the output (background) on the right. The pure oscillatory component, whose power spectrum is also shown on the right, can then be calculated by subtracting the background from the original oscillations.

The weighting coefficients for an ideal filter are directly determined from the Fourier series expansion of the ideal filter characteristic, whereby an infinite number of terms is required in the expansion. Practically, however, a finite number of terms may be sufficient, allowing the series to be truncated after \( N \) terms, which gives a filter with \( 2N+1 \) terms. The net effect of this truncation is the appearance of a certain disturbance of the approximation through the introduction of the Gibbs phenomenon about the band edge frequency \( \beta \). Increasing the number of terms does not reduce this unwanted rippling effect but rather tends to concentrate the variation in the region about the band edge. In order to reduce the rippling effect, an adjustable weighting function is used that alters the filter weighting coefficient in a
certain way as to spread the large approximation error at the band edge over the whole of the passbands and stopbands, effectively reducing the error to within $\pm \varepsilon$ at the expense of a slightly broadened transition band. This method is therefore called the nearly equal ripple approximation (NER), and the definitions of its significant filter parameters are shown in Figure 8.9 for a low-pass filter (left) and a bandpass filter (right) where $\varepsilon$ is the approximation error, $\delta$ is the transition band width and $\beta$ is the band edge. The NER filter design is characterized by approximately equal passband and stopband magnitude errors.

![Figure 8.9](image-url)  
*Figure 8.9  NER filter parameters and tolerance definitions. A description of the band edge $\beta$, the transition band width $\delta$ and the passband and stopband tolerance $\varepsilon$ is given as used in the design of an NER low-pass filter (left) and an NER bandpass filter (right).*

The complexity of the NER filter as measured by the number of filter coefficients is inversely proportional to the width of the transition region with the implication that a greater number of terms is required for a narrower transition region of the filter. As the number of terms is also linearly related to the logarithm of the permissible approximation error, it is convenient to define the passband and stopband tolerance $\varepsilon$ on a logarithmic scale in terms of decibels (dB) as

$$\lambda = -20 \log_{10} \varepsilon. \quad (8.5)$$

A consequence of this definition of the filter design is the requirement for all three filter parameters $\beta$, $\delta$ and $\varepsilon$ to be normalized with respect to the critical Nyquist frequency of the chosen filter. With only minor modifications in the algorithm of the filter coefficients, three types of filters can then be realized, a low-pass, a bandpass and a bandstop filter. A pseudo high-pass filter as a fourth filter type can be accomplished by subtracting the output data values of a low-pass filter from the original input data values. Before the filter coefficients can be calculated, the number of pairs of coefficients $N_f$ has to be determined for the chosen values of $\lambda$ and $\delta$ by defining

$$K_f = \begin{cases} 
0.13927(\lambda - 7.95) & \lambda > 21 \\
1.8445 & \lambda < 21 
\end{cases} \quad (8.6)$$

and subsequently setting
For the design of an NER bandpass or bandstop filter, the only modification of the algorithm for the filter coefficients that is required is the necessity to define a second band edge. Depending on the relative location of the two band edges with respect to each other, either a bandpass or a bandstop filter will be realized, whereby $\beta_1 < \beta_2$ results in a bandpass filter and $\beta_1 > \beta_2$ defines a bandstop filter. After the number of pairs of coefficients $N_p$ and the window parameter $\eta$ have been determined using the same equations as given for the low-pass filter, the filter coefficients $b_k$ of the NER bandpass or bandstop filter can be calculated as
By using a bandpass filter instead of a pseudo high-pass filter, it is possible to not only eliminate the background from the SdH oscillations but also the oscillations at higher frequencies that correspond to the spin and valley splitting, thus facilitating any subsequent analysis of the filtered oscillations. In addition, the bandpass filter provides an ideal method for smoothing the experimental data since any possible high-frequency noise in the measured signal will be suppressed. The application of an NER bandpass filter on SdH oscillations including spin and valley splitting and with a strong background is shown in Figure 8.10. Using a bandpass filter, the output yields directly the pure oscillatory component, whereas the background component can be calculated by subtracting the output data from the original input data.

![Figure 8.10](image_url)  
*Figure 8.10 Power spectrum of SdH oscillations including spin and valley splitting and with a strong background (left). The frequency response of the NER bandpass filter yields directly the spectrum of the fundamental field (right).*

Although it is still necessary to perform an FFT on the original data set to determine the filter parameters and despite the need for lengthy calculations that increase the overall computational time considerably, the use of NER filters in the preparation of $\rho_{xx}$ for further analysis is generally justified due to a much better ability to separate the background from the oscillations with a far higher accuracy. Since the power spectrum does not have to be transformed back into the time domain, potential additional errors introduced by an inverse FFT do not have to be considered. Apart from minor inaccuracies due to truncations in the inverse FFT itself, the biggest potential contribution to such errors is the aliasing of frequencies above the Nyquist critical frequency. An additional advantage of the NER filter is the availability of a bandpass filter which can be applied to separate oscillations associated with main Landau levels and those belonging to spin and valley-split levels or simply to smooth the experimental data by eliminating high-frequency noise.
8.3 ELECTRON EFFECTIVE MASS

Due to the limited temperature ranges of the measurement equipment employed in the experiments, it is not possible to cover the whole intended temperature range between roughly 0.03 K and 4 K with the use of only one apparatus and hence within a single cool-down cycle. Whereas the VTI is capable of covering the higher temperature range from around 1.3 K to 4 K, the lower temperatures between 0.03 K and approximately 1.7 K can only be achieved in the dilution fridge. As the resolution of the oscillations increases with decreasing temperature, there is a trade-off concerning the size of the exploitable temperature range and the number of oscillations that can be analyzed. For a high accuracy of the value for the effective mass, eqn. (8.3) should be fitted to the ratio of the amplitudes along a wide temperature range, while the desire to analyze as many oscillations as possible requires lower temperatures and therefore a much narrower temperature range. The temperature dependence of the SdH oscillations in the temperature range of 0.03 - 4 K is shown in Figure 8.11 for samples BF213, BF313 and BF314 and in the temperature range of 1.7 - 4 K for sample BF197 as it has not been measured in the dilution fridge.

Figure 8.11 Temperature dependence of the SdH oscillations for the 2DEG structures of Table 7.1 at constant electric field.

An additional disadvantage when using oscillations at low temperatures arises from the relatively weak change in amplitude due to comparatively small changes in temperature, thus introducing a higher percentage error from the limited resolution.
of the electrical recording equipment than at higher temperatures where the change in amplitude is much more pronounced. Consequently, the temperature range for analysing the temperature dependence of the oscillations has to be chosen individually for each sample to incorporate temperatures high enough to achieve a fit of sufficient accuracy, yet limit the highest temperatures to provide an adequate number of well-resolved oscillations.

When taking measurements of $\rho_{xx}$ at various temperatures it is important to maintain a constant electric field to guarantee absolute comparability among the different sets of measurements. This requirement might represent a big problem for a wide temperature range insofar as the electric field has to be chosen small enough to avoid electron heating effects at lower temperatures and high enough to keep the noise as low as possible at higher temperatures. For sample BF197, a relatively high current of 1 $\mu$A was chosen for all temperatures which corresponds to an electric field of 168 mV/m. For the measurement of $\rho_{xx}$ in samples BF213 and BF313, a constant current of 100 nA was injected into the Hall bars which is equivalent to an electric field of 32 mV/m and 26 mV/m, respectively. In sample BF314, a current of 50 nA was used to generate an electric field of 13 mV/m. A thorough investigation of the amplitude of the oscillations at low temperatures shows no discernible heating effect in the samples at the given electric fields, and the noise pollution at high temperatures is sufficiently small not to compromise the resolution of the oscillations in the limit where the amplitude of the peaks is small.

Another point of concern is the choice of the magnetic field range in which the peaks of the oscillations are analyzed. At low temperatures, the value of $\rho_{xx}$ starts to vanish for higher magnetic fields when the Fermi level lies in the mobility gap, rendering this part of the oscillations useless for the analysis. At higher temperatures, where the oscillations are much less pronounced, the choice of a wider magnetic field range may enhance the definition of the peak of the fundamental field in the frequency domain which is beneficial for the application of the digital filter. In Figure 8.12, the difference of the power spectra for different magnetic field ranges is shown in both the low temperature and the high-temperature limit. Increasing the magnetic field range yields better defined peaks corresponding to the fundamental field and spin splitting, and an additional peak for valley splitting appears at low temperatures.

![Figure 8.12 Power spectrum of sample BF313 for different magnetic field ranges at 0.057 K (left) and at 4.2 K (right).](image)
Having established both the temperature range and the magnetic field range that promise to deliver the best results for the analysis of the SdH oscillations, the longitudinal resistance at each temperature has to be transformed into a function of the inverse magnetic field, and the data points have to be interpolated at evenly spaced intervals in $1/B$. Before a digital NER filter can be applied, the data has to be transformed into the frequency domain to determine the band edge frequencies of the filter. Figure 8.13 shows the power spectrum of the input data (measured $\rho_{xx}$), the output data (oscillations of the fundamental field) and the discarded data (high- and low-frequency background) of an NER bandpass filter for the four 2DEG samples at low temperatures. Here, the precise separation of fundamental field and superfluous background is demonstrated quite clearly, especially for BF197 where due to the higher temperature the peaks corresponding to the fundamental field and spin splitting lie very close together.

![Figure 8.13 Power spectrum of $\rho_{xx}$ at temperatures below 0.1 K (only BF197 at 1.7 K), the frequency response of the bandpass filter and the power spectrum of the background component.](image)

After the application of the digital filter, the amplitudes of the filtered oscillations can be normalized to those taken at the lowest temperature and plotted as a function of temperature. Finally, it is possible to calculate the electron effective mass as a function of the magnetic field by fitting eqn. (8.3) to the data points. A subsequent verification of the validity and accuracy of the results can be implemented by evaluating the gradient of the data points when plotting $\ln(\Delta \rho_{xx}/\rho_0)$ vs $\ln(\chi/\sinh\chi)$. Both types of plots are displayed in Figure 8.14 for the $n$-type structures, whereby the plots on the left hand side are used to determine the effective mass as described and those on the right hand side are used to provide confirmation of the results.
Figure 8.14 Fitting eqn. (8.3) to the ratio of the amplitudes as a function of temperature at various magnetic fields yields the effective mass (left). According to eqn. (8.1), plotting $\ln(\Delta \rho_\perp/\rho_\parallel)$ vs $\ln(\chi/sinh\chi)$ should generate a straight line with a gradient of unity for the correct effective mass (right).
Although it is feasible to extract the quantum relaxation time from the intercept of the straight line with the ordinate of the $\ln(\Delta \rho_{\text{eff}}/\rho_0)$ vs $\ln(\chi/\sinh \chi)$ plot, possible deviations of the gradient from unity due to accumulated errors from the filtering and fitting process might result in a slightly different offset. Hence, any deduced values for the quantum relaxation time can only serve as a guide to the range in which the actual value of $\tau_q$ can be found. As this plot is used exclusively to verify the validity of the calculated value for the effective mass, the data points are fitted with a fixed rather than with a variable effective mass, and the divergence of the gradient from unity is an indicator of the accuracy of the calculated effective mass. In order to ascertain the potential error in $m^*$, it is necessary to establish the relationship between the effective mass and the gradient. Obviously, $m^*$ is simply proportional to $\chi$, but $\ln(\chi/\sinh \chi)$ has a more intricate relation to $\chi$ within the limit of interest to the fitting procedure. For samples BF197, BF313 and BF314, a change in gradient of 10% is equivalent to a maximum change in effective mass of about 8%. The same variation of the gradient for sample BF213 results in a much smaller modification of $m^*$ which halves to around 4%. This complex interdependence between the effective mass and the gradient is reflected in the deviation of the gradient obtained for the different samples. Whereas the highest divergence for BF213 is approximately 22%, all other samples have maximum deviations of around 11%, all resulting in a maximum error of roughly 9% for the calculated value of the effective mass. A close look at the electron effective masses in Figure 8.15 confirms that all data points lie indeed within the estimated error margin around the expected value of $0.19 m_0$. The error bars include the fitting error as well as the experimental uncertainty which is given by an estimated experimental error of about 3% for all meter readings and a calculated frequency error between 1.6% (BF314) and 5.5% (BF197) as deduced from the ratio of the FWHM and the position of the fundamental field in the power spectrum of $\rho_{\text{ext}}$.

![Figure 8.15](image-url)  
*Figure 8.15* Electron effective mass of the four n-type structures as a function of magnetic field. The dotted lines indicate the anticipated value for the effective mass of $0.19 m_0$ (black) and the corresponding maximum error margins (grey).
Except for sample BF314, for which the electron effective mass is in very good agreement with the value for relaxed Si over the chosen magnetic field range, all other samples display deviations of $m^*$ from this value by at least 2.5%. Similar variations of $m^*$ have been found in other $n$-type SiGe structures for a variety of sheet densities (0.191 - 0.210 $m_e$ [133, 134]) and strain (0.190 - 0.204 $m_e$ [135, 136]). Samples BF213 and BF313 even show a tendency of $m^*$ to increase with increasing magnetic field. Such a dependence on the magnetic field could be explained by a possible nonparabolicity of the electron potential in the conduction band, which would raise $m^*$ as the electron wavevector $k_i$ increases in the plane perpendicular to the magnetic field with increasing magnetic field. Apart from a few other SiGe samples [135, 136], a magnetic field dependent effective mass has also been observed in GaN structures [137], which is thought to be caused by the nonparabolicity of the conduction band. Although nonparabolicity can account for variations of $m^*$ in the present samples, it is expected to be strongest in samples with a high electron density and narrow quantum wells, which is contrary to the observations. Despite having a much lower sheet density, BF213 has a much stronger $B$-dependence of $m^*$ than BF313. Furthermore, $m^*$ is not dependent on the magnetic field in BF197 and BF314 which have sheet densities comparable to BF213 and BF313, respectively.

An alternative explanation for the observed $B$-dependence of $m^*$ is the possible confinement of the electrons by a potential created through disorder. This type of localization may occur when electron-electron interactions or potential fluctuations are sufficiently strong that electrons confined in their local minimum in the disorder potential couple together and effectively experience a single average parabolic potential for all the electrons [138, 139]. Consequently, the system displays no dependence on carrier concentration or interaction effects. Using a zero-dimensional confinement model to account for localization effects, an electron effective mass between 0.1954 $m_e$ and 0.1966 $m_e$ was extracted for similar $n$-type SiGe structures by cyclotron resonance measurements with only little variations of $m^*$ as a function of strain and no dependence on the magnetic field [135, 136]. Without compensating for the disorder potential, $m^*$ displays a significant dependence on both the strain and the magnetic field, comparable to the one observed in Figure 8.15. This specific approach assumes the disorder to be created by the partial relaxation of the strain in the epitaxial layers. However, incorporating the model for zero-dimensional confinement into the analysis of the SdH oscillations presents an insurmountable problem as the zero-field frequency arising from the confinement is unknown. Another type of dependence has been established by cyclotron resonance measurements of $n$-type SiGe samples with a slightly lower strain, which produced oscillating effective masses as a function of the magnetic field with maxima appearing at integer filling factors [140]. The observed filling factor dependence of $m^*$ in those samples can be ascribed to the carrier screening effect on the potential fluctuations due to impurities.
8.4 Hole Effective Mass

Due to a far more complex structure of the valence band, it is much harder to predict the hole effective mass theoretically for a given SiGe p-type structure. Strong nonparabolicity of the subbands at \( k \neq 0 \), the appearance of anti-crossing for increasing \( k \), the warping effect of the hole bands in the \( xy \)-plane and strong spin orbit coupling introduce additional terms not present in the conduction band structure and which influence the value of the effective mass considerably. Therefore, the hole effective mass is expected to experience a strong dependence on the strain, the sheet density and possibly the magnetic field. As a consequence of so many parameters governing \( m^* \), extracting the hole effective mass from both magnetotransport and cyclotron resonance measurements have produced various contrary results, mostly not in agreement with band structure calculations of zone centre effective masses. In order to determine the hole effective mass of the four \( p \)-type structures of Table 7.3, their SdH oscillations are shown in Figure 8.16 for a variety of temperatures. Except for sample BF217 which does not exhibit any oscillations above 2 K, \( \rho_{\alpha\alpha} \) was measured in a temperature range of approximately 0.03 - 4 K for all other 2DHG samples. Apart from having the fewest oscillations and therefore making the analysis more difficult, sample BF638 exhibits a rather strong variation of the background component between the measurements in the VTI and the dilution fridge.

![Graphs of SdH oscillations for BF216, BF217, BF218, and BF638 at various temperatures.](Image)

**Figure 8.16** Temperature dependence of \( \rho_{\alpha\alpha} \) for all \( p \)-type samples at constant electric field.

Like in the analysis of the \( n \)-type structures, the electric field has to be kept constant over the entire temperature range for each sample. Yet, despite using comparable
currents for the measurements, the generated electric field is relatively larger due to a higher resistivity. Sample BF216 and BF638 both use a current of 10 nA, producing an electric field of 77 mV/m and 197 mV/m, respectively. An even higher electric field of 403 mV/m is realized by a current of 50 nA in sample BF218, but because sample BF217 has the lowest sheet density, it needs the highest current (100 nA) to generate well-pronounced oscillations in $\rho_{xx}$ at higher temperatures which is equivalent to an electric field of 795 mV/m. Systematic examinations of the oscillations concerning heating effects at low temperatures and the SNR at higher temperatures reveal no negative effects and thus confirm the validity of the chosen electric fields.

Generally, different magnetic field ranges have to be selected than for the $n$-type samples to achieve the best results because the oscillations start at higher magnetic fields and $\rho_{xx}$ does not vanish at all, allowing for all oscillations up to the maximum magnetic field of 15 T to be analyzed. Again, frequencies corresponding to spin splitting (there is no valley splitting in $p$-type structures) can be eliminated together with the low-frequency background and any potential high-frequency noise by applying a digital bandpass filter whose stop-to-pass and pass-to-stop frequencies can be set exactly as to allow only the frequencies of the fundamental field to pass through the filter as shown in Figure 8.17.

![Figure 8.17](image)

*Figure 8.17 Power spectrum of $\rho_{xx}$, the fundamental field and the discarded background for all four $p$-type samples at temperatures below 0.1 K.*

Except for sample BF638 which does not exhibit any spin splitting at all, the peaks of the 2DHG samples corresponding to the fundamental field and spin splitting overlap...
strongly due to a bigger FWHM compared to those of the n-type samples. Representing the worst case concerning the filtering process, the two merged peaks of sample BF217 are distinctly separated by the NER bandpass filter, clearly demonstrating its precision and effectiveness in removing the background of the SdH oscillations.

The plots to calculate the hole effective mass for the 2DHG structures are illustrated in Figure 8.18 in which the normalized damping factor is fitted to the temperature-dependent normalized amplitudes of the oscillations at various magnetic fields. Also shown is the confirmation of the extracted values by first plotting \( \ln(\Delta \rho_x/\rho_0) \) vs \( \ln(\chi/\sinh\chi) \) and then checking the gradients for possible deviations from unity. The quantum relaxation times as calculated from the intercepts of the straight lines can also be used to assess the validity of the extracted effective mass by comparing the value of \( \tau_q \) to the results from the Dingle plot in chapter 10.

Both types of fitting procedures display a remarkable congruence between the fitted curves and the measured data points for all four samples. However, the error estimation associated with the calculation of the hole effective mass also has to include the experimental uncertainty. Consequently, the error bars of the hole effective mass shown in Figure 8.19 as a function of the magnetic field are considerably larger than the error range of 1 - 3% generated by the fitting process and the error estimation of 3 - 9% as deduced from the divergence of the gradient from unity. Since there is no expected value for the hole effective mass as in the case of the electron effective mass, the error estimation of the verification procedure is included in the display of the error bars. Again, the experimental error related to the meter readings is estimated to be about 3%. However, the calculation of the experimental error associated with the accuracy of the SdH oscillations generates much larger values of 3.6% (BF218), 4.5% (BF216), 8.2% (BF217) and 8.4% (BF638). The relatively large frequency errors seen in samples BF217 and BF638 can be explained by their poor SdH oscillations. Furthermore, an additional fitting error has to be introduced in both samples as the accuracy of the fit diminishes with decreasing number of temperature points.

The most prominent feature of Figure 8.19 is the strong magnetic field dependence of the hole effective mass in all samples as demonstrated by the increase of the effective mass with the magnetic field. This behaviour can be explained by the nonparabolicity of the subbands due to strong valence band mixing which, despite being negligible near the zone centre, becomes more and more significant as the hole energy and thus the in-plane wavevector \( k_x \) increases. Theoretical models to calculate the valence band nonparabolicity and the corresponding hole effective mass are presented for example in [141-143]. Although all three types of subbands (HH, LH and SO) are affected by the band mixing and therefore display nonparabolicity, only the lowest subband, i.e. the HH1 subband, is generally populated in a uniaxially strained QW at low temperatures which therefore determines the hole effective mass. Higher subbands (beyond the HH2 band edge) experience a much stronger nonparabolicity because of enhanced intersubband interactions.
Figure 8.18 The hole effective mass of each sample can be extracted from the $\Delta \rho_{xx}/\rho_{xx0}$ vs $T$ plot on the left and the results can subsequently be verified by the $\ln(\Delta \rho_{xx}/\rho_0)$ vs $\ln(x/\sinh x)$ plot on the right.
As the wavevector $k_0$ grows larger with increasing magnetic field, interaction effects of the HH1 subband with higher subbands and steadily growing anti-crossing effects with the LH1 subband lead to a progressive flattening of the corresponding $E(k)$ curve. This continued reduction of the subband curvature with increasing hole energy is the cause for the observed magnetic field dependence of the hole effective mass. To explain why all four samples exhibit a different magnetic field dependence of the effective mass, it is vital to understand the correlation of the principal factors that determine the nonparabolicity. Apart from the width of the QW, the nonparabolicity of a system is predominantly dependent on the splitting of the HH1 and LH1 subbands, $\Delta$, which is effectively governed by the Ge content. It is also dependent on the hole energy which is strongly influenced by the sheet density via the Fermi energy. Raising the Ge content in the channel increases the heavy hole-light hole splitting, which reduces the band mixing with the LH1 subband and hence effectively extends the region in which the HH1 subband is parabolic to higher values of the wavevector. However, increasing the sheet density also increases the Fermi energy and thus the hole energy, and consequently, larger wavevectors are required, which in turn enhances the nonparabolicity of the system.

![Figure 8.19](image)

**Figure 8.19** Hole effective mass for all p-type samples as a function of the magnetic field. The error bars represent the inaccuracies of the fitting process and the error estimation of the verification procedure.

It has been shown that nonparabolicity is generally insignificant for hole energies smaller than $\Delta/4$ [144]. Therefore, it is possible to establish the significance of the nonparabolicity in a system by assessing the Fermi energy ($E_F$), the cyclotron energy ($\hbar\omega_c$) and the thermal energy ($kT$) as the three energies that constitute the hole energy of the system. For the wavevector to be small enough to access the parabolic part of the HH1 subband, the three relevant energies have to be minimized, and hence, the experiments should be performed on low-density samples in low
magnetic fields at low temperatures. This evaluation of the nonparabolicity effect is valid for SdH as well as for cyclotron resonance (CR) measurements. Hole effective masses that are independent of the magnetic field have been reported for a range of low-density samples with \( n_s \) between \( 1 \times 10^{11} \) and \( 6 \times 10^{11} \) cm\(^{-2} \) and a Ge content between 5% and 30% [145, 146]. Due to very high hole mobilities in these samples, the temperature dependence of the SdH oscillations could be analyzed at relatively low magnetic fields (\( B < 3 \) T), thus keeping the cyclotron energy at a minimum. Assuming a subband splitting of at least 25 meV for the samples with the lowest Ge content, a negligible thermal energy of below 0.1 meV, a maximum Fermi energy of 3 meV for the samples with the highest sheet densities and finally a maximum cyclotron energy (\( 1/2 \ h \omega_c \) for the lowest subband) of 1 meV for the highest magnetic fields, this assessment supports the experimental results that the nonparabolicity of the subbands is negligible in these samples. However, a strong magnetic field dependence of the hole effective mass has been found for high as well as low-density samples with a Ge content of 20% and a sheet density ranging from \( 2 \times 10^{11} \) to \( 1 \times 10^{12} \) cm\(^{-2} \) [147]. A similar evaluation of the relevant energies implies the hole energies of all samples to be low enough to access the parabolic part of the lowest subband, which is contrary to the observed magnetic field dependence of the effective mass in these samples. Establishing the hole effective mass by CR measurements generally neglects or simplifies a possible magnetic field dependence as long as the linear fit of the resonance energy as a function of the magnetic field passes through the origin [144, 148-151].

To allow for a qualified comparison of the samples, the zero-field values of the hole effective mass are extracted in Figure 8.20 by extrapolating the corresponding linear fit of each sample. As a very important parameter of a 2D system, the zero-field effective masses of the four \( p \)-type structures are summarized in Table 8.1 together with the gradient of the linear fit which represents the strength of the magnetic field dependence. For an assessment of the magnitude of the nonparabolicity, the relevant energies of the system are also listed in the table. The subband splitting is calculated by the SCPS programme for nominally identical structures as investigated in the experiments. The theoretical Fermi energy of the idealized parabolic subbands and the experimental Fermi energy as calculated from the measured effective mass are essential for a qualified analysis of the nonparabolicity. Assuming the thermal energy to be lower than 0.35 meV for temperatures below 4 K, the hole energies of all four \( p \)-type structures are well below the nonparabolicity limit of \( \Delta/4 \). Hence, the wavevector of all samples should lie within the parabolic part of the valence subband, and the hole effective mass should display no dependence on the magnetic field. As this approach again fails to explain the observed magnetic field dependence of the effective mass, it is doubtful whether this particular technique of establishing the limit of the nonparabolicity is in general applicable to \( p \)-type SiGe systems.

As all four samples display nonparabolicity effects, it is not surprising that the zero-field effective masses deviate considerably from their theoretical band edge values. In order to analyse the experimental data with respect to the nonparabolicity effects, it is necessary to correct for the fact that the effective masses have been measured at finite sheet densities, and the extrapolated values correspond to energies at the Fermi
wavevector instead of the zone centre. An efficient method to describe the nonparabolicity uses a simple equation derived from the tight binding calculation [152, 153]

\[
\frac{\hbar^2 k_\parallel^2}{2m_f} = \varepsilon \left(1 + c \frac{\varepsilon}{\Delta}\right) \quad \text{and} \quad m^* = \hbar^2 \left(\frac{\partial^2 \varepsilon}{\partial k_\parallel^2}\right)^{-1}
\]  

(8.11)

where \( k_\parallel \) is the 2D wavevector, \( \varepsilon \) is the hole energy, \( m_f \) is the band edge mass and \( c \) is the nonparabolicity factor. While the values for the HH1-LH1 splitting \( \Delta \) are taken from the band structure calculations of the SCPS programme, the Fermi energy has been calculated for the band edge mass assuming a parabolic subband and also for the measured hole effective mass. Hence, it is possible to deduce the nonparabolicity factor for the Fermi wavevector of each sample.

![Graph showing magnetic field dependence of the hole effective mass. The zero-field effective mass of each sample is extrapolated by the corresponding linear fit.](image)

As a preliminary method to evaluate the magnitude of the nonparabolicity, the ratio \( E_f/\Delta \) should provide a trend for its influence on the system since it has been established that both the Fermi energy and the subband splitting are the defining factors that govern the nonparabolicity effects [145]. For low ratios, the extrapolated value of the effective mass is expected to be closer to the band edge mass, whereas higher ratios indicate stronger anticrossing of the subbands and therefore a greater deviation between the hole effective masses at the Fermi wavevector and the zone centre. For the single-step structures, the divergence of the measured effective mass from the band edge mass should thus decrease slightly with declining Ge content, but in reality, the exact opposite trend can be observed. Consequently, this approach is not suitable to predict the magnitude of the nonparabolicity effects of individual
structures due to relatively strong scattering of the data, but it rather provides a certain noticeable trend within a wide range of different structures.

<table>
<thead>
<tr>
<th></th>
<th>BF216</th>
<th>BF217</th>
<th>BF218</th>
<th>BF638</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge content (%)</td>
<td>20</td>
<td>10</td>
<td>30</td>
<td>25/20/15</td>
</tr>
<tr>
<td>sheet density $n_s (10^{11} \text{ cm}^{-2})$</td>
<td>4.8</td>
<td>2.7</td>
<td>11</td>
<td>9.1</td>
</tr>
<tr>
<td>splitting $\Delta (\text{meV})^1$</td>
<td>48.3</td>
<td>32.3</td>
<td>50.3</td>
<td>56.7</td>
</tr>
<tr>
<td>Fermi energy $E_F (\text{meV})^2$</td>
<td>6.2</td>
<td>3.2</td>
<td>15.6</td>
<td>12.3</td>
</tr>
<tr>
<td>$E_F / \Delta$</td>
<td>0.13</td>
<td>0.10</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>cyclotron energy (meV)</td>
<td>0.4 - 1.0</td>
<td>0.4 - 0.7</td>
<td>0.6 - 1.5</td>
<td>1.4 - 4.0</td>
</tr>
<tr>
<td>Fermi energy $E_F (\text{meV})^3$</td>
<td>4.8</td>
<td>2.4</td>
<td>12.0</td>
<td>12.5</td>
</tr>
<tr>
<td>nonparabolicity factor $c$</td>
<td>2.9</td>
<td>4.5</td>
<td>1.3</td>
<td>-0.07</td>
</tr>
<tr>
<td>gradient of magnetic field dependence</td>
<td>0.0499</td>
<td>0.0190</td>
<td>0.0313</td>
<td>0.0168</td>
</tr>
<tr>
<td>zero-field $m^* (m_0)$</td>
<td>0.29±0.05</td>
<td>0.33±0.09</td>
<td>0.27±0.04</td>
<td>0.17±0.09</td>
</tr>
<tr>
<td>band edge $m^* (m_0)$</td>
<td>0.185</td>
<td>0.201</td>
<td>0.169</td>
<td>0.177</td>
</tr>
</tbody>
</table>

$^1$ from SCPS model  $^2$ for a parabolic subband  $^3$ as measured

Table 8.1 Extrapolated hole effective mass (at the Fermi energy) as calculated from the temperature dependence of the SdH oscillations. Also shown are the significant energies of the system that determine the nonparabolicity effects of the HH1 subband.

A much more accurate method to account for the nonparabolicity of the valence subbands is the evaluation of the nonparabolicity factor $c$ in eqn. (8.11). Although this approach is successful at providing a nonparabolicity factor for each structure at a fixed value of the hole energy, it is not suitable to explain the entire nonparabolic behaviour of the subbands over the whole energy range, and hence, it cannot describe the magnetic field dependence of the effective mass. It is obvious from the correlation between the theoretical and the corrected measured hole energies displayed in Figure 8.21 that the nonparabolicity factor calculated at the Fermi energy cannot account for the deviations at higher magnetic fields. From the magnetic field dependence of the measured effective masses, this divergence is expected as the straight line represents a parabolic subband with an effective mass corresponding to that of the Fermi energy, whereas the HH1 subband continues to be nonparabolic for higher wavevectors. If the experimental hole energies were to lie on the straight line for a constant nonparabolicity factor, the subband would become parabolic again for values of $k_f$ above the Fermi wavevector, and the effective mass
would be independent of the magnetic field. The corrected experimental hole energies are each fitted by a square fit which ultimately corresponds to a linear fit of the measured effective masses. Any divergence of the experimental hole energies from the straight line signifies a change of the nonparabolicity factor and thus demonstrates ongoing nonparabolicity of the subbands.

Figure 8.21 Effect of the nonparabolicity of the subbands on the hole energies. Filled symbols indicate hole energies of parabolic subbands, whereas open symbols signify the experimental hole energies corrected for nonparabolicity effects with a constant factor $c$ calculated at the Fermi wavevector. The dotted and dashed lines are guides to the eye.

In the single-step QW structures of BF216 and BF218 and the graded QW structure of BF638, $c$ can be seen to grow continuously with increasing hole energy, indicating a further flattening of the HH1 subband for higher values of the wavevector, as expected. The progression of the hole energies of sample BF217 displays a different character where $c$ falls slightly with increasing magnetic field, suggesting the subband curvature to decrease marginally compared to that at the Fermi energy. Yet, such a behaviour of the nonparabolicity factor is directly opposed to the previous results as it implies a negative gradient for the magnetic field dependence of the effective mass. The most plausible cause for this apparent contradiction is an overestimated value of the effective mass at the Fermi energy. Assuming a lower zero-field effective mass of $0.31 m_0$, the progression of the nonparabolicity factor changes to the expected increase with magnetic field, which is also found in the other samples. Although the rather small magnitude of the fitting error justifies the use of a linear fit to extrapolate the zero-field effective mass in all samples, it is important to point out that this is an idealized case as it presumes the HH1 subband to change its curvature linearly with energy. For small changes in the hole energy close to the Fermi energy, this linear approach can be considered to be fairly accurate, but for a larger range of energy, bigger deviations from the idealized linear dependence are
expected to occur. Along this contention, it could be argued that the effective masses of all samples display two linear regions of different gradient, but it is not clear whether this is a genuine feature of the structures or whether the corresponding values should be ascribed to experimental errors. Since the fitting error of the linear fit over the whole energy range is relatively small and due to a lack of exact knowledge of the dispersion of the hole subbands, it is well within reason to assume the latter case and treat the magnetic field dependence to be linear.

However, for further comparison of the experimental results, only the nonparabolicity factors of the hole energies at the Fermi wavevector are of importance since they correspond to the extrapolated zero-field effective masses. Keeping in mind that the nonparabolicity factor is not only dependent on the Fermi energy but is also directly proportional to the subband splitting, predicting a tendency of \( c \) within the three single-step QW structures is not quite straightforward as the sheet density and the Ge content vary significantly. Hence, it is not immediately obvious why \( c \) decreases with increasing Ge content despite a considerable increase in \( \Delta \) with strain. An explanation of this trend has to include the difference between the Fermi energy of the parabolic and the actual nonparabolic subband which by definition represents the strength of the nonparabolicity and is also a defining factor to determine \( c \). Presuming equal sheet densities, the difference between the two energies is governed by the theoretical band edge mass and the extrapolated zero-field effective mass. A full description of the nonparabolicity factor therefore requires knowledge of the subband splitting to characterize the heavy hole-light hole mixing and thus the overall nonparabolicity of the system, the Fermi energy to determine the part of the subband the wavevector is lying in and the experimental effective mass to establish the actual magnitude of the nonparabolicity. For all three single-step QW structures, the nonparabolicity factor listed in Table 8.1 is much higher than has been reported previously for samples with comparable layer structures. For a similar set of samples with Ge concentrations of 13, 25 and 37%, the corresponding nonparabolicity factors were calculated to be 1.9, 1.3 and 1.0, respectively [150]. Furthermore, it is not possible to establish an average value for \( c \) to correct for the nonparabolicity effects in all samples as has been suggested for a \( \text{Si}_{0.8}\text{Ge}_{0.2} \) structure with varying carrier concentration (\( c = 0.4 \)) [147] and for a set of SiGe samples where both strain and sheet density were changed (\( c = 0.5 \)) [145]. An exceptional behaviour can be seen in the graded QW structure of BF638 where \( c \) has a small negative value at the Fermi energy because the extrapolated effective mass is slightly lower than the theoretical band edge mass, but the nonparabolicity factor turns positive at higher wavevectors as its value grows with increasing magnetic field.

Due to the strong dependence of the hole effective mass on both the Ge content and the sheet density, measurements of the effective mass by SdH and CR experiments have so far yielded a wide variety of different values. The graph on the left hand side of Figure 8.22 displays the dependence of the effective mass on the Ge concentration with data taken from [6, 144, 145, 150, 154-156]. The fairly strong scatter of the data is partly due to the additional variation of the sheet density across the whole range of Ge content. Hence, two categories of samples have to be distinguished to explain the
data, the low-density samples with sheet densities in the low $10^{11}$ cm$^{-2}$ region or below and the high-density structures with carrier concentrations in the high $10^{11}$ cm$^{-2}$ region and above.

Despite being effectively dependent on two parameters in this graph, the effective mass can be seen to decrease with increasing strain, whereby two distinct regions can be identified. The low-density samples have lower effective masses than the high-density samples, especially for lower Ge concentrations. Consequently, two different gradients can be classified, a shallow gradient for the low-density samples and a steeper gradient for the high-density structures. The difference in gradient indicates a much stronger dependence of the effective mass on strain in the high-density samples, whereas the small wavevectors accessed in the low-density structures involve lower magnitudes of nonparabolicity, effectively reducing the influence of the subband splitting on the effective mass.

A far clearer picture emerges in the graph on the right hand side of Figure 8.22 where the variation of the effective mass with the carrier concentration is shown for a Si$_{0.4}$Ge$_{0.2}$ structure [147, 151]. Since the Ge concentration is constant in all samples, the effective mass is only dependent on a single parameter, and hence, the data is much less scattered around the visible trend. As expected, the effective mass can be seen to increase with the carrier concentration since the experiment accesses parts of the subband with a continuing flattened curvature as the wavevector grows bigger.

Concerning the comparability of the data obtained by SdH and by CR measurements, it should be noted that the cyclotron experiments measure the bare effective mass because of translation symmetry, whereas the transport experiments measure the many-body effective mass. An additional potential problem concerning the SdH experiments is the possibility that the small- and the large-angle scattering time which both govern the amplitude of the oscillations may not have the same temperature dependence.

Clearly, the effective masses of the three single-step QW structures have to be analyzed in this context and should also be compared to the values reported in the literature. Although the initial intention of the three different structures was to

![Figure 8.22](image-url)
establish a relation between the hole effective mass and the strain in the QW, the sheet density also increases inevitably with the Ge content due to a deepening of the potential well. As the influence exerted by both parameters on the effective mass are of opposite character, it has to be investigated which of the two parameters is dominant and whether a common trend can be established in relation to either one of the parameters. As can be seen in Figure 8.23, the hole effective mass decreases with increasing Ge concentration in agreement with theory [105] and experimental results from literature. This can be interpreted as confirmation that a change in strain has a stronger influence on the effective mass than any variation of the Fermi energy as long as the carrier concentration is not too high as for very high density samples, this dependency has been found to disappear, leaving the effective mass virtually independent of the Ge content. Despite the observed linear relation between the effective mass and the strain, the gradient of this dependency is expected to be slightly steeper for a constant \( n \), since \( m^* \) is still influenced by the sheet density. However, when taking into account the relatively large error bars associated with the calculated values of \( m^* \), the dependence of the hole effective mass on the strain can also be interpreted to be much weaker than expected from the corresponding band edge values, suggesting a much bigger influence of the sheet density and thus the Fermi energy.

![Figure 8.23](image)

Figure 8.23 Experimental effective mass values of the single-step QW structures (solid circles) and the graded QW structure (open circle) as a function of Ge composition compared with the calculated band edge mass (solid line) for strained SiGe alloys.

Although BF218 is clearly a high-density sample, the effective masses of all three single-step QW structures are well below those of the high-density samples of Figure 8.22 and should definitely be compared to the values of the low-density samples. Despite having higher sheet densities, the gradient of the present set of samples is almost identical to that of the low-density samples. Accounting for the differences in
the sheet density between both sets of samples, the values for the effective mass are very similar across the investigated Ge concentrations. While there is good agreement of the results in this respect, it is not quite as simple to explain the drop in effective mass from BF216 to BF218, considering that the subband splitting in BF218 is only slightly bigger than in BF216, whereas the Fermi energy more than doubles at the same time. As the subband splitting has been calculated by the SCPS programme, there might be variations between the model and the grown structure that can account for the difference in effective mass between the two samples. On the other hand, the decoupling of the subbands is not expected to increase linearly with Δ, and a small change in Δ at relatively high values might decouple the subbands to a larger extent than the same change at lower values.

Such an enhanced decoupling of the subbands at higher values of Δ could also partly explain the very low effective mass of BF638. Despite extending over the hole QW region, the main part of the wavefunction resides in the QW step with the highest Ge content. Hence, the corresponding band edge mass is best compared with that of a single-step QW of 25% Ge composition. Due to a higher subband splitting and a lower sheet density than found in BF218, the effective mass is expected to be the lowest of all four 2DHG samples. Considering the difference in Δ between BF216 and BF218 is only 2 meV, a further increase in Δ of over 6 meV should result in a substantially lower effective mass. However, even a totally decoupled HH1 subband cannot explain an experimental hole effective mass that lies almost 0.01 $m_0$ below the value of the corresponding band edge mass. Due to this inconsistency and the large error margin associated with the experimental error estimation, the actual value of the effective mass is believed to lie within the limit of the upper error bar. For the highest possible values, the effective mass would even lie within the range of the interpolated value of a 25% Ge uniform structure when including the error bars. An equally low value of 0.18 $m_0$ has been found by CR experiments for a low-density sample with a low Ge concentration of 12% [144]. To explain the results, it was argued that the hole energies accessed by the experiments were small enough to lie within the parabolic part of the subband. The band edge mass of this structure, however, can be calculated to be about 0.2 $m_0$ and is therefore higher than the experimental value. Lower values of $m^*$ between 0.12 $m_0$ and 0.20 $m_0$ have been observed in SiGe structures with a pure Ge channel [157, 158], which is substantially lower than $m^* = 0.29 m_0$ in bulk Ge. Depending on the carrier concentration, the lowest hole effective masses of below 0.1 $m_0$ have been reported for low-density samples, also with a pure Ge channel [159]. Although an extrapolation of the effective mass values in Figure 8.22 and Figure 8.23 predicts a value for $m^*$ in a pure Ge channel that agrees with these results, the lowest measured value of 0.04±0.01 $m_0$ again lies potentially below the corresponding band edge mass of about 0.05 $m_0$. 

194
8.5 SUMMARY

The temperature dependence of the SdH oscillations can be used to evaluate the effective mass of a given structure with high accuracy. Before the electron and hole effective mass of the corresponding $n$- and $p$-type system can be extracted from the cyclotron frequency of the thermal damping factor, the low-frequency background of the SdH oscillations has to be eliminated with a digital filter, whereby the best results can be achieved when the oscillations are filtered in the time domain. Since the parabolic curvature of the electronic states in the conduction band remains unchanged in the presence of uniaxial strain, the electron effective mass has a constant value of $0.19 \pm 0.02 \, m_e$ in all investigated 2DEG structures, irrespective of the structural layer design and largely independent of the magnetic field. Despite including the fitting error as well as the experimental uncertainty, the error margin of the electron effective mass is relatively small in all $n$-type samples. The hole effective mass, on the other hand, is sensitive to the degree of uniaxial strain, the Ge concentration and the sheet density due to the nonparabolicity of the valence bands, and hence, the present $p$-type structures are expected to display different hole effective mass values. As the nonparabolicity effects also cause the hole effective mass to increase with magnetic field, it is necessary to extrapolate the value of the hole effective mass at zero field from the experimentally acquired values. In agreement with theory, the hole effective mass decreases with increasing Ge concentration from $0.33 \pm 0.09 \, m_e$ for a Ge concentration of 10% to $0.29 \pm 0.05 \, m_e$ for a 20% Ge content and finally to $0.27 \pm 0.04 \, m_e$ for a 30% Ge composition. However, the observed tendency of $m^*$ diminishes when the error bars are incorporated into the interpretation, in which case the hole effective mass may even be interpreted to remain constant over the whole range of Ge compositions. This apparent contradiction to the behaviour of the band edge mass can be explained by the increase of the sheet density and thus the Fermi energy to which the zero-field effective mass values correspond. With a value of $0.17 \pm 0.09 \, m_e$, the lowest effective mass could be observed in the step-graded structure, whereby the drop in $m^*$ is thought to be caused by an increased splitting of the HH1 and LH1 subbands, effectively reducing the nonparabolicity effects in the HH1 band. Due to the large experimental uncertainty, however, the value is believed to lie realistically within the upper limit of the error margin where it might even fit the general tendency of the effective mass in the uniform structures.
Knowledge of the effective mass by itself, however, is not sufficient to describe the quality of a SiGe system adequately as optimum semiconductor device design is ultimately based on a full understanding and accurate modelling of charge-carrier transport in semiconductors. Hence, the most important transport parameter of a semiconductor material is the carrier mobility $\mu$ which describes the linear relation between the average carrier velocity $v$ and an external electrical field $E$ in the low-field limit as

$$v = \mu E$$

(9.1)

in the absence of external magnetic fields. The mobility itself is directly proportional to the transport relaxation time $\tau$, and inversely proportional to the effective mass $m^*$, viz.

$$\mu = \frac{e}{m^* \tau}.$$  

(9.2)

At the same time, the mobility of the charged carriers is also given by the sheet density $n_s$ and the longitudinal zero-field resistivity $\rho_{xx}$ via

$$\mu = \frac{1}{n_s \epsilon \rho_{xx}(B = 0)}.$$  

(9.3)

Because of its relevance to both basic understanding and device applications, there has always been a strong interest in precise model descriptions of the mobility as a function of strain, temperature and sheet density. All three parameters are of great importance to the mobility as they influence either the effective mass (strain), the transport relaxation time (temperature) or both (sheet density). As has been established in the previous section, especially the hole effective mass is strongly dependent on the strain as well as on the sheet density due to the profound nonparabolicity of the valence subbands. On the other hand, the transport relaxation time is dependent on the dominant scattering mechanisms which in turn are dependent on the temperature and the sheet density amongst other factors. Hence, measurements of the low-temperature mobilities are commonly employed to determine the overall quality of a heterostructure, and that knowledge can subsequently be used to optimize the growth parameters in order to produce superior 2D structures. Heterostructures of high quality are ultimately needed for advances in device applications, e.g. by improving the current drive capability and the speed performance of SiGe MODFET devices.
9.1 **FUNDAMENTAL FIELD**

Using eqn. (9.3) as the most convenient way to calculate the mobility requires prior knowledge of the sheet density which is given by

\[ n_s = \frac{g_s g_v e B_f}{\hbar} \]  

with \( B_f \) being the fundamental field. Hence, the calculation of the sheet density and subsequently of the carrier mobility is practically reduced to the evaluation of the fundamental field. This can be achieved either by analysing the extrema of the SdH oscillations or alternatively by assessing the peak of their power spectrum as has been done in the preceding chapter. Figure 9.1 shows the principle of both methods to determine the fundamental field.

The initial condition that the Fermi energy of a 2D system has to coincide with the Fermi energy in presence of a quantizing magnetic field sets the index of the extrema of the SdH oscillations and the corresponding value for the reciprocal magnetic field into the specific linear relation

\[ n = B_f \frac{1}{B_n}. \]  

Plotting the index \( n \) of the minima of the oscillations against the reciprocal magnetic field \( 1/B_n \) as shown in the left diagram of Figure 9.1 and performing a linear regression of the individual data points yields a straight line whose gradient is given by the fundamental field according to eqn. (9.5). It is possible to include both the minima and the maxima of the oscillations in this analysis to double the number of data points for a more precise regression, but it has to be taken into account that this approach also doubles the gradient of the linear fit which therefore represents twice the fundamental field. This method of assessing the fundamental field can only be employed if only the lowest subband is occupied. For structures with several populated subbands, the oscillatory component of \( \rho_{xx} \) is the sum of as many sine waves as there are occupied subbands, all with different frequencies. In this case, the data points in the \( n \) vs \( 1/B_n \) graph are not represented by a straight line, rendering this approach ineffective for the calculation of the fundamental field.

Alternatively, the fundamental field can be determined from the Fourier transform of the SdH oscillations by locating the main peak of the power spectrum as shown in the right diagram of Figure 9.1 where the corresponding frequency denotes the fundamental field in Tesla. The precision of the fundamental field value in this method is limited by the resolution of the FFT which is considerably lower than the resolution of the extrema of the original oscillations used in the first method. However, it is generally still possible to obtain a fairly accurate value for the fundamental field from the power spectrum of the oscillations, and since an FFT has to be performed in any case to determine the parameters of the digital filter, no additional calculation time has to be assigned for this method. Moreover, this
technique is still valid for structures with more than one populated subband. Increasing the number of occupied subbands generates more peaks in the power spectrum of $\rho_{xx}$, each representing the fundamental field of the corresponding populated subband, whereby the peak corresponding to the lowest subband is located at the highest frequency. Thus, the density of each subband can be calculated individually, and subsequently summing up over all densities provides the total sheet density of the system. Ideally, both methods should be used to establish the fundamental field because the validity of the result can be verified immediately by comparing both values. However, both techniques are only applicable for magnetic fields below the onset of spin splitting which introduces further oscillations.

![Graph](image)

**Figure 9.1** The two methods of evaluating the fundamental field. Plotting the index $n$ of the extrema of the oscillations vs $1/B_y$ creates a gradient that is equivalent to the fundamental field (left). The frequency corresponding to the main peak of the power spectrum of the oscillations is also equal to the fundamental field (right).

Having thus established and verified the value of the fundamental field, the sheet density can subsequently be calculated according to eqn. (9.4), and finally, it is possible to determine the carrier mobility from eqn. (9.3). Alternatively, the sheet density can also be calculated from the gradient of the Hall resistance which is defined by the simple relation

$$\text{Hall gradient} = \frac{1}{n_e e}. \quad (9.6)$$

Apart from using the magnetotransport measurements of the transverse instead of the longitudinal channel, the main difference of this approach to the first method is that the gradient of the Hall resistance is related to the average density of all parallel conducting layers in the structure, whereas the analysis of the SdH oscillations yields the sheet density exclusively in the QW. Hence, it is possible to investigate the existence of parasitic parallel conducting channels in the supply layer by comparing the results of both methods. An agreement between the sheet densities extracted from the SdH and the Hall measurements implies a total freeze out of all potential parallel conduction in the supply layer.

As the sheet density and the carrier mobility are directly linked, it is practical as well as useful to investigate both transport properties together and in close relation. A
principal aspect of such an analysis is the potential dependence of the transport properties on the temperature. According to the temperature range of the different experimental equipment used in the measurements, it is necessary to distinguish between two separate temperature regimes. All low-temperature magnetotransport measurements were carried out in the Heliox system (0.4 - 4.2 K), the VTI (1.3 - 4.2 K) or the dilution fridge (0.03 - 4.2 K). Additional transport measurements were performed at higher temperatures in the closed cycle cryostat between 10 K and 300 K. Due to space restrictions, the transport properties of the 2DEG samples are discussed only in the low-temperature regime (below 4.2 K), whereas the 2DHG structures are also examined in the high-temperature range (10 - 300 K). In the low-temperature regime, the transport properties of the 2DEG and 2DHG samples have been studied not only with respect to their dependence on the lattice temperature but also with regard to a possible dependence on the carrier temperature by keeping the lattice temperature constant and varying the applied electric field.

9.2 ELECTRON MOBILITY AND SHEET DENSITY

When extracting the transport properties of a QW structure from low-temperature magnetotransport measurements, it is important to establish that the investigated sample behaves like a true 2D system. Before the transport properties can be analyzed correctly, it is therefore necessary to find out whether any parallel conduction occurs in the supply layer and whether higher subbands in the QW are populated. In all investigated 2DEG samples, the two corresponding values of the sheet density as deduced from the Hall resistance and the SdH oscillations are identical to within a few percent, and the Fourier analysis indicates only the lowest subband to be occupied. The successful implementation of both criteria confirms the true two-dimensional character of the n-type samples in the low-temperature regime and permits the use of the previously defined analysis techniques to evaluate the transport properties. The influence of the ageing process on the transport properties of a 2DEG structure has been investigated for samples BF213 and BF314 by taking two sets of measurements that are separated by a time period of 18 months. Changes in the transport behaviour have been taken into account by reporting on both sets of results for sample BF213 whereby the first set of measurements is referred to as BF213 (a) and the second set is represented by BF213 (b).

The two graphs at the top of Figure 9.2 show the sheet density (left) and the electron mobility (right) of the 2DEG structures as a function of the lattice temperature below 4.2 K. In this temperature range, the sheet density of all samples is characteristically independent of temperature, whereas the electron mobility decreases almost linearly with increasing lattice temperature. As expected, the two delta-doped samples BF313 and BF314 have the highest sheet density, about twice as high as the density of the two uniformly doped samples BF197 and BF213. This relation supports the assumption that the resulting change in doping profile of the delta deposition improves the electron transfer from the supply layer into the channel significantly.
The slightly larger sheet density in BF313 is consistent with a marginally higher doping concentration in the supply layer due to a longer deposition time of the dopant material. Of the two low-density samples, BF197 has the lower electron density in the channel because of a lower doping concentration, a stronger surface segregation or a combination of both. Although the same growth parameters were chosen for the supply layer of both samples, the resulting doping concentration can vary appreciably as it is very difficult to establish identical growth conditions in the growth chamber for different runs. The influence of the ageing process on the structure of sample BF213 is clearly demonstrated by a noticeable drop in the sheet density due to a partial relaxation of the strained QW and a corresponding drop in the confining potential. In contrast, the transport properties of sample BF314 are not affected by the ageing process as the values of the sheet density and the electron mobility remain unchanged over time within the margin of experimental error.

In agreement with theory, the two high-density samples display the lowest mobilities with BF313 exhibiting the highest sheet density and the lowest mobility. Despite having almost identical sheet densities, BF197 has a much higher mobility than BF213 due to a significantly lower zero-field resistivity, indicating much reduced scattering in the channel. The appreciable drop in mobility for sample BF213 after a time period of 18 months can be explained by the partial relaxation of the strained Si channel and the associated increase of threading dislocations in the virtual substrate, raising the scattering rate in the well and thus the zero-field resistivity. To explain

---

Figure 9.2   Sheet density (left) and electron mobility (right) of the 2DEG samples as a function of temperature (top) and electric field (bottom, at base temperature). The dotted lines are guides to the eye.
the almost linear decrease of the mobility with increasing temperature and electric field in all samples, it is necessary to investigate the temperature dependence of both acoustic phonon and Coulomb scattering. Within the given temperature range, electron-phonon interactions are characterized by a $T^\gamma$-dependence with an exponent of $\gamma \geq 3$ (typically $\gamma = 5$ for unscreened deformation-potential scattering). At the same time, Coulomb scattering, e.g. by ionized impurities or interface roughness, is generally assumed to be independent of temperature at very low temperatures. However, this assumption is only valid in the long-wavelength limit where the screening parameter $q_s$ is constant but not at shorter wavelengths where the screening parameter decreases with increasing wavevector and also with increasing temperature \[99, 160\]. This wavevector and temperature dependence of the screening parameter is principally responsible for a substantial variation of the Coulomb scattering rate with temperature. Considering the full wavevector dependence at low temperatures, the Coulomb scattering rate and consequently the corresponding reciprocal mobility increases approximately linearly with temperature \[82, 161, 162\]. At this point, it is useful to establish the relation between the various reciprocal mobilities which is given by

$$
\mu^{-1}(T) = \mu_0^{-1}(T) + \mu_{ac}^{-1}(T)$$

(9.7)

where $\mu(T)$ is the $T$-dependent total mobility, $\mu_0(T)$ represents the $T$-dependent mobility limited by Coulomb scattering and $\mu_{ac}(T)$ denotes the $T$-dependent mobility limited by acoustic-phonon scattering. In samples with very high mobilities, as is the case in most III-V heterostructures, the $T$-dependence of $\mu_0^{-1}$ is negligible compared to that of $\mu_{ac}^{-1}$ because $\mu_0^{-1}$ is only a small fraction of $\mu^{-1}$. Accordingly, the power law of the combined $T$-dependence is roughly that given by the electron-phonon interactions. In Si inversion layers, on the other hand, the mobility is relatively low, and $\mu_0^{-1}$ constitutes a considerable part of $\mu^{-1}$. Consequently, the linear temperature dependence of $\mu_0^{-1}$ dominates the $T$-dependence of the total mobility in these samples, whereas the variations of $\mu_{ac}^{-1}$ with temperature are insignificant. Since the present SiGe samples display mobilities that are higher than those in Si inversion layers but well below those in GaAs structures, the $T$-dependence of Coulomb as well as acoustic-phonon scattering contribute to the $T$-dependence of the total mobility. Taking into account that the mobilities of the present samples lie in the lower part of the mobility range and $\mu_0^{-1}$ still represents a significant part of $\mu^{-1}$, the $T$-dependence of the total mobility is governed predominantly by Coulomb scattering and only weakly by the power law of acoustic-phonon scattering. Such a combination of the scattering rates results in an almost linear temperature dependence in the given temperature range, but the $T^\gamma$-dependence of the phonon processes cause the mobility to deviate from a straight line at higher temperatures roughly above 3 K.

At the bottom of Figure 9.2, the sheet density (left) and the electron mobility (right) are shown as a function of the electric field at a constant lattice temperature. The measurements were taken at base temperature which corresponds to the lowest temperature reached in each of the experiments. By applying high electric fields to the 2DEG structures, the electrons in the QW are heated above the lattice
temperature \( T_L \) and hence, the two graphs can also be interpreted as displaying the dependence of the transport properties on the electron temperature \( T_e \). Describing the most essential difference to a uniformly heated lattice, the phonon scattering process of hot electrons \( (T_e > T_L) \) is governed predominantly by phonon emission of the electrons as they lose energy to the cooler lattice, even though some phonon absorption may still be possible. Due to this net loss of energy to the lattice by the electron system, the phonon scattering rate of hot electrons is expected to be reduced compared to the case when the lattice is also heated and \( T_e = T_L \). By individually varying either the temperature or the electric field in separate experiments, the specific behaviour of the mobility and sheet density as a function of \( T_L \) and also as a function of \( T_e \) can be investigated. When using \( T_i \) as the variable, the injected current has to be small enough at each temperature for the resulting electric field not to heat the electrons above the respective lattice temperature.

To demonstrate the importance of growing high-quality virtual substrates with a low density of threading dislocations, the results of four additional samples from the second batch have been included in both diagrams at the bottom of Figure 9.2. Again, the sheet density does not vary with temperature for all samples and the initial results for the four samples of the first batch are replicated. Except for the growth temperature and the layer structure of the virtual substrate, all samples of the second batch are nominally identical to BF213, but their sheet densities are considerably higher, indicating an enhanced electron transfer due to a higher average doping concentration or otherwise an improved doping profile which, in fact, is less likely. The problem of exactly reproducing a desired layer structure is further demonstrated by the big variation in sheet density between BF851 and BF852, two samples with nominally identical layer structures and grown at the same temperature in immediate succession. Only BF835 and BF851 appear to have identical doping profiles and concentrations as both samples exhibit identical sheet densities.

Considering the lattice temperature is held constant at the respective base temperature in all experiments as the electric field is varied, the mobilities at the lowest fields where no electron heating occurs have to be identical to the highest mobilities of the corresponding \( T_L \)-dependence experiments. The validity of this condition can be seen for the samples of the first batch (BF197 – BF314) when comparing the mobilities as a function of the lattice temperature with those as a function of the electric field. Due to their high carrier densities, the mobilities of the samples from the second batch (BF741 – BF852) are expected to be comparable to those of the high-density samples BF313 and BF314. However, very poor mobilities much below those of BF313 and BF314 indicate very strong short-range scattering, most likely due to a greatly increased density of threading dislocations in the virtual substrate. In contrast to the samples of the first batch, there is no discernible relation between the sheet density and the mobility as the quality of the buffer layer seems to be the dominant factor to influence the mobility.

In order to allow for a direct comparison between the \( T_L \)- and the \( T_e \)-dependence of the mobilities, it is necessary to convert the values of the electric field into values of the electron temperature. This conversion can be completed indirectly by using the
$T_L$-dependence of the amplitude of $\rho_{\alpha\beta}$ as reference for the thermometry. In this approach, the amplitude is plotted as a function of $T_L$ and a polynomial fit of the data points represents the calibration for $T_c$. By comparing the amplitudes of the different electric fields with the calibration curve, a corresponding electron temperature can be obtained for each value of the electric field. Alternatively, the electron temperature can also be determined more directly by exploiting the temperature dependence of the amplitudes in eqn. (8.1) as has been done in chapter 8 to determine the effective mass. Rewriting the indices in eqn. (8.3) accordingly, the relation between the $T_L$- and the $T_c$-dependence of the amplitudes is given by

$$\frac{\Delta \rho_{\alpha\beta}(T_c)}{\Delta \rho_{\alpha\beta}(T_L)} = \frac{T_c}{T_L} \frac{\sinh(\xi T_c)}{\sinh(\xi T_L)},$$

(9.8)

where $T_c$ is the only adjustable parameter. By plotting the right hand side of this equation as a function of $T_c$, it is possible to extract the matching electron temperature for all ratios of the amplitudes since each individual ratio has to lie on the plotted curve. Following the procedures of this method, the extracted values of the electron temperature are shown in Figure 9.3 as a function of the electric field.

![Figure 9.3](image_url)  
Figure 9.3 Electron temperature $T_e$ as function of the electric field $E$ at a lattice temperature of 50 mK for the samples of the first batch (left) and at a higher lattice temperature of 0.4 K for the samples of the second batch (right). The dotted lines are guides to the eye.

Clearly visible in both graphs, two fundamental principles concerning the electron temperature in a 2DEG system can be identified. Most significantly, the value of $T_c$ as a function of the electric field is almost independent of the transport properties and thus the structure of the sample. Furthermore, the electrons start getting hot at lower electric fields the lower the lattice temperature, making the electron temperature a function of the electric field and the lattice temperature. These two characteristic features can be explained by looking at the theoretical concepts of the acoustic-phonon scattering process. Taking into account that all electronic transitions take place in a narrow region of width $\Delta k \approx m^* k_b T_c / \hbar^2 k_f$ near $k_f$ and the area of this region is $2\pi m^* k_b T_c / \hbar^2$, the available phase space for transitions is independent of both the Fermi wavevector and the carrier density. At low lattice temperatures and constant $T_c$, the total power emitted by the electron system is therefore independent of $n_e^{-1}$, which means that the power loss per electron is proportional to the sheet

203
density. In steady state, the electrons have reached equilibrium at temperature $T_e$ where the energy-loss rate $-(dE/dt)$ matches exactly the input power per electron $P_e = e\mu F^2$ (in this case, the electric field is symbolized by $F$ to avoid confusion with the energy $E$), thus fulfilling the energy balance

$$e\mu F^2 + \left(\frac{dE}{dt}\right) = 0.$$  \hspace{1cm} (9.9)

Apart from an additional weak $n_e$-dependence associated with various scattering processes, the mobility is also proportional to $n_e^{-1}$ and hence, the variation of the electron temperature with the applied electric field displays only a weak dependence on the carrier density. However, structures with divergent mobilities due to distinct scattering rates require slightly different electron temperatures to reach equilibrium for a given electric field. For increasing mobilities, the energy-loss rate has to rise to satisfy eqn. (9.9), and since $(dE/dt)$ increases with $T_e$, the electron temperature will also increase. The temperature dependence of the energy-loss rate is also the reason why the electrons get hot more readily the lower the lattice temperature. While the energy-loss rate is proportional to $T_e^\gamma$ ($\gamma > 3$), the supplied electrical power $P_e$ does not change much with $T_e$ as the mobility is dominated by temperature independent scattering. Hence, higher fields are required to heat the electrons above the lattice temperature with increasing $T_e$.

Having converted the values of the electric field into corresponding values of the electron temperature, the mobilities can be evaluated as a function of $T_e$, thus enabling a direct comparison to their equivalent variation with $T_e$. For a better understanding of the temperature dependence, it is convenient to separate the reciprocal mobility into a temperature-dependent term $\mu^{-1}_{\text{coul}}(T)$ and a term that is independent of temperature, ultimately modifying eqn. (9.7) into a linear relation of the form

$$\mu^{-1}(T) = \mu^{-1}_0 + \mu^{-1}_{\text{coul}}(T) + \mu^{-1}_{\text{ac}}(T) = \mu^{-1}_0 + \mu^{-1}_{0,\text{coul}}(T) = \mu^{-1}_0 + \alpha T.$$ \hspace{1cm} (9.10)

Here, $\mu^{-1}_0$ represents the Coulomb scattering rate at 0 K, and $\mu^{-1}_{0,\text{coul}}(T)$ comprises the linear $T$-dependence (in this temperature range) of both Coulomb and acoustic-phonon scattering which can also be expressed in terms of its associated gradient $\alpha$. As a critical restriction to eqn. (9.10), it should be noted that this linear relation is only applicable at low temperatures where $\mu^{-1}_0(T) >> \mu^{-1}_{\text{ac}}(T)$. Due to the temperature dependence of the screening parameter in the full wavevector approach which evidently relates to the temperature of the screening electrons, Coulomb scattering is presumed to be a function of the electron temperature as well and it should make no difference for the $T$-dependence of the corresponding reciprocal mobilities whether the whole lattice is heated or only the electrons ($\mu^{-1}_{0,\text{coul}}(T) = \mu^{-1}_{0,\text{coul}}(T_e)$). Acoustic-phonon scattering, on the other hand, is sensitive to this distinction because the phonon population increases much stronger with $T_e$ than with $T_e$. Due to additional scattering caused by phonon absorption, cold electrons that are at lattice temperature are expected to have a higher scattering rate and should therefore yield a larger value for $\alpha$ than hot electrons at equivalent temperature which scatter predominantly by
phonon emission alone ($\mu_{0+sc}^{-1}(T_l) > \mu_{ac}^{-1}(T_e)$). Any differences in $\mu_{0+sc}^{-1}(T)$ between individual samples and also between the $T_l$- and $T_e$-dependence can be described distinctively by the variation of the gradient $\alpha$ of the reciprocal mobilities. In Figure 9.4, the temperature-dependent part of the reciprocal mobilities $\mu_{0+sc}^{-1}(T)$ is shown as a function of both $T_l$ and $T_e$ for the $n$-type samples of the first set, which clearly demonstrates the validity of eqn. (9.10) in the temperature regime below $\sim 3$ K. At higher temperatures, the continuously growing contribution of acoustic-phonon scattering causes $\mu_{0+sc}^{-1}(T)$ to follow its characteristic $T^2$-dependence (for unscreened deformation-potential scattering). As a result, the experimental data points can be seen to deviate from the calculated straight line above $\sim 3$ K where they display a superlinear temperature dependence. Also shown in Figure 9.4 are the values of $\alpha$ as a function of the sheet density for the $T_l$-dependence of the first set and the $T_e$-dependence of both sets of $n$-type samples.

In all samples, the value of $\alpha$ is larger for the $T_l$-dependence than for the equivalent $T_e$-dependence, and assuming Coulomb scattering to have the same dependence on $T_l$ and $T_e$ in each sample, any variations of $\alpha$ can be ascribed exclusively to changes in acoustic-phonon scattering. From the linear behaviour of the reciprocal mobilities below $3$ K, it can be concluded that the contribution of $\mu_{0}^{-1}(T)$ to $\mu_{0+sc}^{-1}(T)$ is bigger by roughly two orders of magnitude compared to the corresponding contribution of $\mu_{ac}^{-1}(T)$. To fully describe the linear $T_e$-dependence in samples BF213 (b) and BF314,
for instance, acoustic-phonon scattering is calculated to constitute roughly 1% of\( \mu_{\text{phon}}^{-1}(T_e) \) in BF213 (b) and about 0.4% in BF314. These contributions are by far too small to account for the relatively large variations of \( \alpha \) between the \( T_e \)- and \( T_h \)-dependence which changes by almost 50% and 30%, respectively. Such a big difference in the gradient below 3 K can only be explained if Coulomb scattering is presumed to increase as well. However, this explanation directly contradicts the assumption made earlier and requires the screening parameter to depend not only on the screening effect of the electrons but apparently also on some other temperature-dependent property of the lattice. On the whole, \( \alpha \) can be seen to increase with sheet density, which emphasizes the importance of the screening effect from electrons on the screening parameter. It also suggests that the \( T \)-dependence of the screening effect per electron somehow increases to produce larger gradients for higher electron densities. A more detailed analysis of acoustic-phonon scattering from the gradient of the reciprocal mobilities thus appears to require a more comprehensive theoretical investigation of the Coulomb scattering dependences.

Finally, it is possible to calculate the mobilities \( \mu_{\text{phon}}(T) \) that are only limited by acoustic-phonon scattering and the temperature-dependent part of Coulomb scattering. Plotting \( \mu_{\text{phon}} \) against \( T_L \) and \( T_h \) on a double logarithmic scale as shown in Figure 9.5 results in a straight line with a negative gradient that is identical to the exponent of the \( T \)-dependence of the reciprocal mobilities. Consequently, the observed gradient of unity corresponds exactly to the linear \( T \)-dependence of Coulomb scattering. Due to the comparatively weak contribution of acoustic-phonon scattering, the expected increase of the gradient at higher temperatures to a value of five is only vaguely visible. The high values of \( \mu_{\text{phon}}(T) \) confirm that the temperature independent part of the Coulomb interactions (at \( T = 0 \) K) are the dominant scattering mechanism which limit the measured electron mobility.

The transport properties of all 2DEG samples and, where possible, the values of \( \alpha \) are summarized in Table 9.1, providing a suitable overview to compare the quality of the structures among each other but also to other 2DEG structures in literature. When comparing these results to the transport properties found in literature, it is important to distinguish between samples grown by MBE and those grown by CVD which constitute the two principal growth techniques. Depending on the type of virtual
substrate employed, the structures grown by CVD exhibit a wide range of mobilities. The lowest mobilities were found for virtual substrates that use a superlattice buffer layer to relax the strain [163] with initial values being as low as $1.9 \times 10^6 \text{ cm}^2/\text{Vs}$, although further improvements to this type of virtual substrate increased the mobilities more recently to $1.3 \times 10^5 \text{ cm}^2/\text{Vs}$ [164]. Slightly higher mobilities of $1.6 \times 10^6 \text{ cm}^2/\text{Vs}$ could be seen by employing a step-graded buffer layer instead of the superlattice [165], and most interesting for device applications, samples with a step-graded buffer layer grown on silicon-on-insulator (SOI) yielded mobilities of around $8.2 \times 10^4 \text{ cm}^2/\text{Vs}$ at a slightly higher temperature [166]. However, the best results were achieved when a virtual substrate was used in which the Ge composition was graded linearly, more than doubling the electron mobilities in the strained Si layer to $3.9 \times 10^5 \text{ cm}^2/\text{Vs}$ [167].

<table>
<thead>
<tr>
<th></th>
<th>base temperature (K)</th>
<th>sheet density ($10^{11} \text{ cm}^{-2}$)</th>
<th>electron mobility (cm$^2$/Vs)</th>
<th>$\alpha$ of $\mu^{-1}(T_L)$ ($10^7 \text{ Vs/cm}^2\text{K}$)</th>
<th>$\alpha$ of $\mu^{-1}(T_e)$ ($10^7 \text{ Vs/cm}^2\text{K}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF197</td>
<td>1.7</td>
<td>6.0</td>
<td>163,000</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>BF213 (a)</td>
<td>0.1</td>
<td>6.8</td>
<td>107,000</td>
<td>3.9</td>
<td>2.7</td>
</tr>
<tr>
<td>BF213 (b)</td>
<td>0.05</td>
<td>6.1</td>
<td>88,000</td>
<td>4.6</td>
<td>3.1</td>
</tr>
<tr>
<td>BF313</td>
<td>0.05</td>
<td>12.0</td>
<td>53,000</td>
<td>10.0</td>
<td>7.5</td>
</tr>
<tr>
<td>BF314</td>
<td>0.05</td>
<td>11.2</td>
<td>58,000</td>
<td>12.2</td>
<td>9.4</td>
</tr>
<tr>
<td>BF741</td>
<td>0.4</td>
<td>8.5</td>
<td>34,000</td>
<td>-</td>
<td>7.9</td>
</tr>
<tr>
<td>BF743</td>
<td>0.4</td>
<td>6.8</td>
<td>40,000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF834</td>
<td>0.4</td>
<td>6.3</td>
<td>17,000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF835</td>
<td>0.4</td>
<td>9.2</td>
<td>16,000</td>
<td>-</td>
<td>12.0</td>
</tr>
<tr>
<td>BF851</td>
<td>0.4</td>
<td>9.3</td>
<td>14,000</td>
<td>-</td>
<td>9.1</td>
</tr>
<tr>
<td>BF852</td>
<td>0.4</td>
<td>11.4</td>
<td>22,000</td>
<td>-</td>
<td>13.6</td>
</tr>
</tbody>
</table>

Table 9.1 Sheet density and electron mobility of 2DEG samples at base temperature and low electric fields (to avoid heating effects). Also shown are the slopes of the inverse mobilities as a function of both the lattice and the electron temperature.

A further improvement of these results could be realized by applying a positive bias to the back-gate of the sample, whereby the wavefunction is pulled into the middle of the QW and thus away from the upper interface. Depending on the back-gate bias, the mobility of this sample could be enhanced up to a value of $5.2 \times 10^5 \text{ cm}^2/\text{Vs}$ [103, 167], also due to an increase in effective spacer layer thickness without
compromising the sheet density, thereby decreasing the influence of the remote ionized impurities. More moderate mobilities of $2.7 \times 10^6$ cm$^2$/Vs were found for linearly graded virtual substrates grown at higher temperatures [168, 169]. Here, increasing the spacer layer thickness of the structure reduced the electron transfer from the supply layer into the Si channel, thus decreasing the sheet density and the mobilities considerably.

As for samples grown by MBE, the mobilities for structures with a linearly graded buffer layer tend to be significantly lower than their counterparts grown by CVD. This discrepancy might be caused by the marginally higher quality of the virtual substrate that can be realized by the CVD growth technique. Early measurements of MBE grown samples showed a substantial difference in low-temperature mobilities between structures with single-step buffer layers ($1.3 \times 10^6$ cm$^2$/Vs) and with virtual substrates that are linearly graded ($1.7 \times 10^6$ cm$^2$/Vs) [170]. Confirming their superior quality, similar results were found for similar structures with graded buffer layers, displaying mobilities of almost $1.3 \times 10^6$ cm$^2$/Vs at a slightly higher temperature [171]. Despite this apparent disadvantage compared with samples grown by CVD, the structure with the highest electron mobility to date has been grown by MBE combined with solid-phase epitaxy (SPE), whereby the SPE growth completely suppressed Ge segregation in the SiGe layer on top of the Si channel [172]. The resulting abrupt and very smooth interface allowed an extremely high electron mobility of $8.0 \times 10^5$ cm$^2$/Vs at 15 K. This important discovery implies a much bigger role of the interface quality influencing the electron mobility than first assumed when the mobility was increased by moving the wavefunction away from the upper interface [103, 167]. Indeed, it has been shown that interface roughness scattering becomes dominant in narrow quantum wells of widths below 5 nm [173, 174]. The structures of the previous examples were all grown by SSMBE which has been a well-established process since the early eighties. Being the most recent of the growth techniques, on the other hand, GSMBE has only been in use for about ten years, allowing for less experience and hence a relatively poorer understanding of the growth process itself. Consequently, the few samples grown by GSMBE so far exhibit even lower mobilities as the growth process keeps being improved. Despite this lack of experience, low-temperature mobilities just below those found in SSMBE structures were measured for similar structures as listed in Table 9.1, reaching values of $9.5 \times 10^4$ cm$^2$/Vs [175] and which certainly compare to the results of the samples of the first set (BF197 - BF314). However, the problem of reproducing high-quality structures with this process on a constant basis is demonstrated by recently reported poor mobilities of only $2.0 \times 10^4$ cm$^2$/Vs [176] that fall within the mobility range of the second set of samples (BF741 - BF852).

At room temperature, however, the difference in mobilities between the various growth techniques and virtual substrates is reduced considerably because acoustic and optical phonons are the dominant scatterers, and parallel conduction in the supply layer may become significant. For structures with a linearly graded buffer layer grown by CVD, the room-temperature mobilities vary between 1800 and 2800 cm$^2$/Vs [163, 177-179]. With a value of just above 2100 cm$^2$/Vs, the corresponding mobility for a step-graded virtual substrate [180] falls within this
range, showing no signs of a degenerating effect from the lower quality buffer layer on the mobility. This insignificance of the quality of the virtual substrate is further confirmed by the sample utilizing a step-graded buffer layer grown on SOI which displays the highest room-temperature mobility of 2900 cm²/Vs [166]. Hydrogenation of the heterostructures has been found to be an efficient process to improve the transport properties at room temperature by passivating the parallel conduction as well as some background impurities and dislocations in the channel, thereby enhancing the mobility dramatically [181]. Apart from an MBE-grown sample with an exceptionally high mobility of 2700 cm²/Vs [182] which rivals that of the best CVD-grown samples, other structures grown by SSMBE are located at the lower end of the mobility spectrum, displaying room-temperature mobilities of 1600 [171] and 1800 cm²/Vs [170]. The spread of results highlights the problem of this growth technique to reliably produce high-quality 2DEG structures. Compared to the lower values of the SSMBE samples, the results of the structures grown by GSMBE have only marginally inferior mobilities of 1300 (BF197), 1200 (BF213 and BF313) and 1100 cm²/Vs (BF314) for much higher sheet densities which reach values in the $10^{13}$ cm⁻² range. This is a clear indicator that these structures have not been optimized for room-temperature operation, but it also underlines the existing problems with this growth technique.

9.3 HOLE MOBILITY AND SHEET DENSITY

Prior to evaluating the transport properties from the low-temperature magneto-transport measurements of the $p$-type structures, it is essential to verify for each sample individually that the investigated system is indeed two-dimensional in nature. As in the case of the $n$-type samples, the Hall resistance and the SdH oscillations yield almost identical values for the sheet density, thus confirming the absence of any parallel conduction in the supply layer. Further evidence of their two-dimensional character is provided by a separate Fourier analysis of the oscillations, each displaying a single peak at the equivalent of the fundamental field reproducing the same sheet densities as calculated by the other two methods and corroborating the population of only the lowest subband in the channel.

The sheet densities and hole mobilities of all four 2DHG samples are shown in Figure 9.6 as a function of the lattice temperature below 4.2 K and also as a function of the applied electric field. By performing magnetotransport measurements on sample BF218 on two different dates separated by 18 months, it is possible to examine the degrading effect of the ageing process on the transport properties representative for all 2DHG structures. From this point onwards, the results from the initial assessment are symbolized by BF218 (a), whereas the corresponding results from the second set of measurements are referred to as BF218 (b).

Displaying a characteristic feature of 2D structures in the low-temperature regime, the sheet densities of all $p$-type samples are constant with temperature as well as...
with electric field. Despite having identical layer structures except for the Ge concentration in the channel, the densities of samples BF216 – BF218 span almost a decade, emphasising the importance of the alloy composition and thus the strength of the confining potential to the efficiency of the hole transfer from the supply layer into the QW. Consistent with having the lowest Ge content of all samples (10%), BF217 exhibits the lowest \( n_v \), whereas BF218 displays the highest density as the sample with the highest Ge concentration (30%). Interestingly, the values of \( n_v \) of BF216 (20%) are much closer to those of BF217, suggesting a non-linear relation between the Ge content in the channel and the resulting confining potential. Having only a slightly modified supply layer, the sheet density of BF638 lies between that of BF216 and BF218, according to a Ge concentration of 25% at the lower interface of the step-graded channel. Apparent form both graphs in Figure 9.6, it is sensible to categorize the p-type structures into a set of two low-density samples (BF216 and BF217) and another set of two high-density samples (BF218 and BF638). Concerning the ageing process of BF218, it is remarkable that no difference in \( n_s \) can be seen between the two sets of measurements, indicating no discernible degradation of the 2DHG structure with time and no significant relaxation of the strained alloy in the QW as this would imply a drop in the confining potential and subsequently a decline in the sheet density.

![Figure 9.6](image)

Figure 9.6 Sheet density (left) and hole mobility (right) of the 2DHG samples as a function of temperature (top) and electric field (bottom, at base temperature). The dotted lines are guides to the eye.

In general, it is not surprising for the low-density samples to display the highest mobilities and accordingly for the lowest mobilities to be found in the high-density
structures. As the sample with the lowest sheet density, BF217 exhibits by far the highest hole mobility of all four \( p \)-type structures. Having the highest density, on the other hand, agrees with BF218 being the sample with the lowest mobility of the three single-step structures, and consistent with the ratio of the sheet densities, the mobility of BF216 lies in the middle of these two values. The fact that the ratio of the three mobilities can be fully accounted for by the proportion of the sheet densities alone (to within a few percent) is an indicator for the scattering rates in those samples to be very similar in spite of changing alloy compositions and associated interface qualities that are anticipated to diverge. Although alloy scattering is known to increase with the Ge content, it is not a dominant scattering process in those samples as the measured mobilities are still below those limited by alloy scattering, even if screening is neglected due to the short-range nature of the scatterers. In low-density samples, interface charge scattering is dominant, and considering that the scattering rate decreases with increasing sheet density but increases with the Ge concentration in the alloy, it is possible that the effect of a higher \( n_s \) counterbalances the effect of a higher Ge content to generate similar scattering rates in samples BF216 and BF217. In the high-density sample BF218, the dominant scattering process is expected to be by interface roughness which increases with \( n_x \). Even for a comparable Ge content, the scattering rate should be slightly higher than in the low-density samples, and a higher Ge concentration is likely to increase it yet further. The lowest mobility, however, can be found in the step-graded sample BF638 due to the hole wavefunction extending across all three steps of the channel. Despite its peak being positioned in the part of the 3-step QW with the highest Ge content, the wavefunction encompasses two interfaces of unknown quality, and although both interfaces should have less scattering centres than the upper interface, their influence on the mobility is much stronger as they are placed directly in the path of the holes. Assuming the scattering rate in uniform channels with a Ge content between 10% and 30% to be primarily dependent on \( n_x \), the presence of the two interfaces within the channel reduce the mobility by a factor of approximately 2.7 compared to a uniform channel with a similar sheet density. While there has been no discernable modification of \( n_x \) in connection with the ageing process, a slight drop in mobility suggests an increased scattering rate due to either a very weak relaxation of the QW or a certain degradation of the quality of the upper interface. As a partial relaxation of the strained channel has been ruled out (although a minimal relaxation that hardly alters the confining potential but produces a small number of threading dislocations in the active layer is conceivable), an increase of the interface roughness scattering rate may be the most plausible explanation for the drop in mobility over time.

Analysing the temperature dependence of the transport properties requires the prior conversion of the electric field into corresponding values of the hole temperature \( T_{h\nu} \) which can be accomplished by using the same technique as described in the previous section to extract the electron temperature. An interesting aspect of \( T_e \) can be seen in Figure 9.7 which clearly shows that compared with \( T_e(E) \), the holes start heating up at higher electric fields for roughly the same lattice temperature. In fact, the hole temperature of BF638 at 30 mK follows exactly the same curve as \( T_e(E) \) at 0.4 K, and despite having a smaller gradient at higher electric fields, the holes of the single-step samples start to heat up above the lattice temperature at similar electric fields. An
interpretation of this characteristic can be given by considering the much lower mobilities of the holes which result in a significantly reduced input power per hole $P_h$ at a given electric field. This drop in $P_h$ partially offsets the reduction in the energy-loss rate at lower temperatures, and consequently, higher electric fields are necessary to achieve an input power that is high enough to heat the holes above the lattice temperature. The obvious divergence of $T_h(E)$ between the single-step and the step-graded samples can be explained by the difference in effective mass. Despite having higher hole mobilities and hence a higher energy gain at a given field, the larger hole effective masses of the single-step samples generate increased rates of energy loss that slow the heating of the holes. Under these conditions, the holes require a higher electric field at a given hole temperature to gain enough energy to fulfil eqn. (9.9) and remain in steady state. At low electric fields, this difference is hardly noticeable as the hole temperatures are infinitesimally small. With increasing field, however, the holes start to be heated significantly, and the discrepancy between the hole temperatures becomes visible. Taking these considerations into account, the exact match of $T_h(E)$ of BF638 at 0.03 K and $T_s(E)$ of the second set of $n$-type samples at 0.4 K is not considered to be a matter of principle but seems to be caused by a coincidental balance of the critical parameters ($\mu$, $m^*$ and $T_L$).

![Figure 9.7](image)

Figure 9.7  Hole temperature $T_h$ as a function of the electric field $E$ at a lattice temperature of 30 mK.

After the individual hole temperatures have been established in this manner, it is possible for the hole mobilities to be compared with respect to their dependence on $T_L$ and $T_h$. Following the approach given by eqn. (9.10), the temperature-dependent part of the reciprocal mobility $\mu_{\text{ned}}^{-1}(T)$ is illustrated in Figure 9.8 as a function of the lattice and hole temperature. In the temperature range below 3 K, $\mu_{\text{ned}}^{-1}(T)$ displays the expected linear behaviour in all $p$-type samples with the exception of BF638 for which the experimental data starts to deviate from a straight line at the slightly lower temperature of ~2.5 K. Due to the relatively high hole temperatures, BF638 is also the only $p$-type sample for which the contribution of acoustic-phonon scattering is clearly visible in the $T_h$-dependence of the reciprocal mobilities. To provide an adequate overview of the gradients in all samples, the values of $\alpha(T_L)$ and $\alpha(T_h)$ are also shown in Figure 9.8 as a function of the sheet density where $\alpha$ can be seen to increase in general with $n_s$, and $\alpha(T_L)$ is clearly bigger than $\alpha(T_h)$ in all $p$-type samples.
Similar to the situation in the \( n \)-type samples, these results confirm that the temperature dependence of Coulomb scattering is not only dependent on the screening effect of the holes but is also influenced by other temperature-dependent factors of the lattice. Having otherwise identical sheet densities, the different values of \( \alpha \) for BF218 (a) and BF218 (b) are thought to be caused by differences in the Coulomb scattering rate as a result of relaxation.

As illustrated in Figure 9.9, the double logarithmic plot of \( \mu_{\text{ph}} \) against \( T_L \) and \( T_T \) yields the expected straight line with a gradient of unity for each sample. The influence of acoustic-phonon scattering is only clearly visible in sample BF638 for which the gradient can be seen to increase towards higher temperatures. The dominance of Coulomb scattering at \( T = 0 \) K (which constitutes the temperature independent part of the Coulomb interactions) in limiting the actual hole mobilities is demonstrated by the much higher values of \( \mu_{\text{ph}} \) compared to the experimentally obtained values.

Presenting a general overview of the investigated 2DHG samples, the low-temperature transport properties and the temperature dependences are summarized in Table 9.2, facilitating a qualified comparison between the different structures but also to other results reported in literature. As most 2DHG samples are grown by MBE which is known to produce interfaces of very high quality due to its monolayer growth capability, it is not useful to distinguish between samples grown by CVD and
MBE. Instead, it is more meaningful to categorize the structures according to their Ge concentration in the channel into low and high Ge content samples. Low Ge content structures generally have a maximum Ge concentration of 30%, allowing for the strained channel to be grown directly on top of the Si substrate or, more precisely, a Si buffer layer. Samples that are considered to be high in Ge content, on the other hand, employ channels with a Ge concentration in excess of 70% and therefore require virtual substrates with a maximum Ge content varying between 50% and 80% to reduce the strain. In fact, most high Ge content structures have a pure Ge channel to fully exploit its high hole mobility and to completely avoid alloy scattering.

The first successfully grown 2DHG sample was a double heterostructure with supply layers positioned on both sides of the strained channel [6]. Due to the absence of a virtual substrate, the quality of the channel was already exceptionally high and low-temperature mobilities of 3300 cm²/Vs were achieved for a Ge content of 20% in the channel. Decreasing the Ge concentration in the channel of a similar structure to 12% increased the mobility to 3700 cm²/Vs [183], clearly laying emphasis on the importance of the Ge content. However, most 2DHG structures grown more recently have only one supply layer, and due to surface segregation of the dopant, a normal structure [114, 115, 155, 184-189] with the supply layer on top of the SiGe channel is preferable to an inverted structure [185, 190, 191] where the supply layer is grown before the alloy channel. Although high mobilities of 4550 cm²/Vs could be realized in an inverted Si₀.₈Ge₀.₂ structure [191], a direct investigation regarding the influence of a normal and inverted structure on the mobility yielded the much higher value of 6000 cm²/Vs for the normal structure with the supply layer on top of the Si₀.₈Ge₀.₂ channel [185]. Further improvements of otherwise relatively low mobilities to up to 4600 cm²/Vs could be accomplished in gated structures by moving the hole gas away from the upper interface and by varying the sheet density [114, 115, 190]. However, in order to achieve even higher mobilities in low Ge content structures, it is vital to increase the quality of the upper interface and simultaneously reduce the phonon scattering rate. Following the concept of this idea, low Ge concentrations in conjunction with low sheet densities have so far produced the highest low-temperature mobilities, reaching 9300 cm²/Vs [186] and 12500 cm²/Vs [188] for a Ge

![Figure 9.9](image_url)  
**Figure 9.9** Hole mobilities limited by acoustic-phonon scattering and the temperature-dependent part of Coulomb scattering as a function of $T_1$ (left) and $T_h$ (right).
concentration of 12% in the channel. By employing an even lower Ge content of only 6.5% in the channel, the mobility could be raised still further to an extremely high value of $2 \times 10^4 \text{ cm}^2/\text{Vs}$ [187]. Growing a strained QW of 13% Ge composition on top of an SOI substrate yielded a comparable high value for the mobility of almost $1.7 \times 10^4 \text{ cm}^2/\text{Vs}$ [189].

![Table 9.2](image)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Base Temperature (K)</th>
<th>Sheet Density ($10^{11} \text{ cm}^2$)</th>
<th>Hole Mobility (cm$^2$/Vs)</th>
<th>$\alpha$ of $\mu^{-1}(T_L)$ ($10^{-5}$ Vs/cm$^3$K)</th>
<th>$\alpha$ of $\mu^{-1}(T_h)$ ($10^{-5}$ Vs/cm$^3$K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF216</td>
<td>0.03</td>
<td>4.8</td>
<td>4,500</td>
<td>2.5</td>
<td>1.3</td>
</tr>
<tr>
<td>BF217</td>
<td>0.03</td>
<td>2.7</td>
<td>7,300</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>BF218 (a)</td>
<td>0.06</td>
<td>11.0</td>
<td>2,000</td>
<td>4.7</td>
<td>2.4</td>
</tr>
<tr>
<td>BF218 (b)</td>
<td>0.03</td>
<td>11.0</td>
<td>1,700</td>
<td>5.9</td>
<td>3.5</td>
</tr>
<tr>
<td>BF638</td>
<td>0.03</td>
<td>9.1</td>
<td>900</td>
<td>4.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Table 9.2 Sheet density and hole mobility of 2DHG samples at base temperature and low electric fields (to avoid heating effects). Also shown are the gradients of the reciprocal mobilities as a function of both the lattice and the hole temperature.

Despite being the only structures to be grown by GSMBE, BF216 - BF218 compare well with all the other samples in literature, and their results suggest a very similar quality of the SiGe channel and the upper interface. To produce structures with even higher mobilities, virtual substrates have to be employed to allow for the growth of high Ge content channels, preferably with pure Ge. Depending strongly on the quality and the final Ge composition of the virtual substrate, the mobilities vary immensely. The first structure grown with a pure Ge channel used a low-quality virtual substrate with a final Ge composition of 50% achieving hole mobilities of only 3200 cm$^2$/Vs [192]. This value could be increased to 15500 cm$^2$/Vs by improving the quality of the virtual substrate and by raising the final Ge composition to 70% [157]. Surpassing these results, the highest values to date have been found in structures with a pure Ge channel and a virtual substrate of very high quality. For a final Ge composition of 78% and 70% in the virtual substrate, exceptionally high hole mobilities of $2.7 \times 10^4 \text{ cm}^2/\text{Vs}$ [158] and $5.5 \times 10^4 \text{ cm}^2/\text{Vs}$ [159] have been realized, respectively.

As for the transport properties of p-type structures at room temperature, the study of most low Ge content samples show similar mobilities, ranging from 200 cm$^2$/Vs for a double heterostructure with 20% Ge in the channel [193] to just over 300 cm$^2$/Vs for normal structures with a Ge content of 36% [194] and 20% [195] in the QW. Moving the wavefunction into the middle of the channel by grading the Ge composition in the channel from 15% to 25% does not seem to improve the mobility at room temperature as it yields a maximum value of only 220 cm$^2$/Vs [196]. By comparison, BF217 exhibits a room-temperature mobility of 190 cm$^2$/Vs despite a relatively high...
sheet density of $7 \times 10^{12} \text{ cm}^{-2}$. Samples BF216 and BF218, however, compare less favourably as they have even higher sheet densities of $3 \times 10^{13} \text{ cm}^{-2}$ and $4.5 \times 10^{13} \text{ cm}^{-2}$, respectively, restricting their mobilities to a mere 100 cm$^2$/Vs. The room-temperature performance of these samples could certainly be improved by reducing the doping concentration in the supply layer. Much higher mobilities were achieved by high Ge content structures reaching values as high as 1050 cm$^2$/Vs for a Ge concentration of 80% in the channel [197]. Equivalent to the low-temperature regime, structures with a pure Ge channel display also the highest room-temperature mobilities with reported values of 830 cm$^2$/Vs [158] and 1700 cm$^2$/Vs [182, 198].

As illustrated in Figure 9.10, the hole mobility and the sheet density of all three single-step samples display the typical $T$-dependence of 2D structures in the temperature range between 10 K and 300 K. The temperature at which acoustic phonons become the dominant scatterers to limit the mobilities depends on the combination of three factors, i.e. the Ge composition, the strain and the interface quality [199, 200]. Below 50 K, all parallel conduction is frozen out and only the lowest subband in the QW is occupied. Hence, the sheet density stays constant, and the hole mobility, which is mainly limited by acoustic-phonon scattering above 10 K, decreases linearly with increasing temperature.

![Figure 9.10](image)

**Figure 9.10** Temperature dependence of the hole mobility and the sheet density of the three single-step 2DHG samples in the temperature range 10 - 300 K.

At temperatures around 50 K, the sheet density starts to rise slowly as higher subbands in the QW become populated due to thermal activation of the holes. A further increase in temperature induces the development of a parallel conduction channel in the supply layer, which results in a stronger rise of the hole concentration.
and consequently a more rapid fall in the hole mobility with rising temperature. At higher temperatures, however, the negative slope of the mobility can be seen to decrease in samples BF216 and BF218, similar to the low Ge content samples of [187]. This behaviour can only be explained by the very high hole concentration in these samples and a subsequently strong parallel conduction in the Si supply layer that has a higher hole mobility than the 2DHG in the channel.

At room temperature, the mobility is primarily dependent on the acceptor concentration and the Ge composition in the channel [105, 106, 201-204]. In general, the hole mobility falls with increasing acceptor concentration and rises with increasing Ge content, although this relationship is also dependent on the chosen alloy scattering potential. Assuming a doping concentration in the low $10^{18}$ cm$^{-3}$ for all three samples, the mobility should lie around 125, 150 and 170 cm$^2$/Vs for a Ge concentration of 10%, 20% and 30%, respectively. Another crucial point of concern is the Hall scattering factor $r_H$, which is known to deviate from unity at elevated temperatures due to the nonparabolicity of the dispersion curve and the anisotropic nature (warping) of the constant energy surface of the valence band. Depending on the energy of the relaxation time and thus the scattering rates, $r_H$ decreases with increasing temperature from unity at low temperatures to about 0.4 at 300 K [205, 206]. At the same time, $r_H$ also decreases with increasing Ge content possibly due to changes in the energy dependence of the relaxation time [201]. Furthermore, a strong correlation exists between the Hall scattering factor and the hole density [205], whereby $r_H$ can be seen to increase with $n_h$. Taking all three dependences of the Hall scattering factor into account, the room-temperature mobilities of all four investigated 2DHG samples are probably underestimated by a factor of 0.4, at most.

9.4 Summary

Providing a valuable insight into the charge-carrier transport of the investigated semiconductors, the carrier mobility can be calculated from the sheet density and the zero-field resistivity of the SdH oscillations. The sheet density itself can be extracted from the fundamental field or alternatively from the gradient of the Hall resistance. In all of the present $n$- and $p$-type samples, the good agreement between the values of the sheet density as extracted from both methods confirms the absence of any parallel conduction, and the Fourier analysis indicates only the lowest subband to be occupied. Variations of the electron mobility in the $n$-type structures can only be caused by changes in the sheet density or the scattering process since the electron effective mass remains constant. Demonstrating a high structural quality, the low-density samples display high electron mobilities in the range of $1 - 1.7 \times 10^5$ cm$^2$/Vs. Much lower mobilities between 5 and $6 \times 10^4$ cm$^2$/Vs can be seen in the delta-doped structures which is due to the higher sheet density rather than a lower quality of the structure. Even lower mobilities in the second set of $n$-type samples, on the other hand, can be explained by a relatively lower structural quality. In $p$-type structures,
the dependence of the hole effective mass and the interface quality on the Ge content influence the hole mobility in opposite ways. Whereas the effective mass decreases with increasing Ge concentration, thus improving the mobility, the quality of the interface diminishes, and the corresponding increase in scattering causes a reduction in the mobility. Furthermore, a rise in the Ge content causes an increase in the sheet density as the sheet density is dependent on the confinement potential in the channel. Accordingly, the sample with the lowest Ge concentration exhibits the lowest sheet density, and despite having the highest effective mass, it displays the highest hole mobility of $7 \times 10^3 \text{ cm}^2$/Vs. A higher sheet density and corresponding lower mobility of $4.5 \times 10^3 \text{ cm}^2$/Vs can be seen in the structure with a 20% Ge content. The sample with the highest Ge content consequently displays the highest sheet density and also has the lowest mobility of only $2 \times 10^3 \text{ cm}^2$/Vs. However, an even lower mobility of just under $10^3 \text{ cm}^2$/Vs can be seen in the step-graded structure, even though the value of the hole effective mass is believed to be the lowest of the present $p$-type samples and the sheet density is lower than in the uniform sample with a 30% Ge content. Such a low mobility can only be explained by the presence of strong scattering from the additional two interfaces in the quantum well which are located within the boundaries of the wavefunction. Interestingly, the temperature dependence of the mobilities in the present $n$- and $p$-type structures is almost linear below 4.2 K, suggesting a relatively strong contribution from Coulomb scattering despite fairly high mobilities in some of the $n$-type samples.
10 Transport and Quantum Relaxation Time

Although the mobility is the principle quantity to describe the charge-carrier transport in a semiconductor, it does not provide any information about the relative importance of the various scattering potentials limiting the mobility in these structures. To improve the quality of a heterostructure and ultimately increase its mobility, it is crucial to acquire some knowledge about the dominant scattering mechanism by characterizing the system in terms of its relaxation times. Whereas the quantum relaxation time $\tau_q$ is given by the total scattering rate (eqn. (4.10)), hence incorporating both large- and small-angle scattering, the transport relaxation time $\tau_t$ is weighted by the scattering angle $\theta$ (eqn. (4.11)), thereby favouring large-angle scattering [82, 93-98]. This difference in sensitivity to small- and large-angle scattering can be used to establish whether the dominant scattering potential is long- or short-range in nature. For large-angle potentials like interface roughness, the two relaxation times are almost identical, and the ratio $\tau_t/\tau_q$ is expected to be close to unity, but if small-angle scattering associated with remote impurities dominates, $\tau_t$ is typically larger than $\tau_q$ by a factor of 10 or more [93-98]. Being defined by the solution of the Boltzmann equation in the relaxation time approximation, the transport relaxation time is directly related to the mobility via eqn. (3.4) and values for $\tau_t$ can therefore be calculated straight from the results of the preceding chapter. The quantum relaxation time, on the other hand, is only given implicitly in the exponent of eqn. (3.66), influencing the magnitude of the SdH oscillations as $\tau_q$ is determined by the collision broadening of the Landau levels [207]. This means that the effect of increased scattering upon the SdH line shape is equivalent to a rise in lattice temperature. Consequently, a value for $\tau_q$ has to be extracted experimentally from the amplitude of the SdH oscillations by evaluating the corresponding Dingle plot. Rewriting eqn. (8.1) as

$$\Delta\rho_{xx} = 4\rho_0 X(T) \exp\left(-\frac{\pi m^*}{e\tau_q} \frac{1}{B}\right)$$  \hspace{1cm} (10.1)

where $X(T)$ is the thermal damping factor $\chi/\sinh\chi$, a Dingle plot can be realized by plotting $\ln[\Delta\rho/(\rho_0 X(T))]$ against the reciprocal magnetic field as shown in Figure 10.1, and the slope gives $(-\pi m^*/e\tau_q)$ with an intercept at $1/B = 0$ of 4 [96, 98].

In certain cases, however, the Dingle plot of a real 2D system may deviate from the behaviour predicted by theory [98]. These variations can either be caused by parallel conduction or inhomogeneities in the sample, each associated with a characteristic divergence from the ideal Dingle plot. In the presence of a parallel conduction path in the supply layer, the intercept at $1/B = 0$ is smaller than the theoretical value of 4, but unless the parallel conduction is extreme, the slope of the Dingle plot still yields $\tau_q$ to a good approximation. However, if the parallel conduction path occurs in a
higher subband, $\tau_q$ includes contributions not only from intrasubband but also from intersubband scattering. Due to this difference in $\tau_q$, it is obvious that the values of $\tau_q$ cannot be compared directly when only a single subband is occupied and when two or more subbands are populated.

A more complicated behaviour of the Dingle plot can be seen in samples that have inhomogeneities. If the inhomogeneities are localized in the sample resulting in two or more regions with different densities, the Dingle plot displays two or more sets of beating oscillations. Whether a reliable value of $\tau_q$ can still be extracted from such a plot depends strongly on the amplitude of the oscillations and thus the magnitude of the density variations. Alternatively, the inhomogeneities may be manifested as a general spread of densities, leading to a destructive interference of the corresponding spread of frequencies of the SdH oscillations. The resulting dephasing decreases with increasing magnetic field, thus causing a strong curvature of the Dingle plot.

### 10.1 **Dominant Scattering in 2DEG Samples**

In order to guarantee equivalent experimental conditions for the calculation of both relaxation times, $\tau_q$ has been determined at the same values of the lattice and electron temperature as the mobility which ultimately yields $\tau_r$. In good agreement with eqn. (10.1), the Dingle plots of all $n$-type samples are well behaved at all temperatures, each displaying a straight line with an intercept of 4 within a few percent. Representative for the Dingle plots of all other samples, Figure 10.2 shows the Dingle plots of BF213 as a function of the lattice and the electron temperature. As can be seen clearly in both cases, the slope of the Dingle plot increases with $T_L$ and $T_e$, indicating a drop in $\tau_q$ with increasing temperature due to a steady rise of the phonon
population which causes further broadening of the Landau levels from increased acoustic-phonon scattering.

![Graph](image)

*Figure 10.2 Dingle plots of sample BF213 for varying lattice temperature (left) and also for varying electron temperature (right).*

The values of the transport and quantum relaxation times are shown in Figure 10.3 as a function of the lattice temperature and the electric field. While $\tau_q$ is found to be less than 1 ps in all samples except BF213 (a), $\tau_r$ of the first set of samples is larger by at least an order of magnitude. Such a big difference in relaxation times indicates small-angle scattering from remote ionized impurities to be dominant in these samples. Despite having similar quantum relaxation times, the samples of the second set display transport relaxation times that are only larger by a factor of 4 - 7. Since the ratio of $\tau_r / \tau_q$ is bigger than one, small-angle scattering is still assumed to dominate but with a much stronger contribution from large-angle scattering compared to the first set of samples. The different values of $\tau_r / \tau_q$ can be seen in Figure 10.4 which, together with the individual values of the transport and quantum relaxation time, provide valuable insight into the dominant scattering process of each sample. A summary of $\tau_r$, $\tau_q$, and $\tau_r / \tau_q$ at base temperature and low electric field is given in Table 10.1. In order to compare the samples with respect to their relaxation times, the parameters influencing $\tau_r$ and $\tau_q$ for remote impurity scattering have to be investigated. As ionized impurities are Coulomb scatterers, the force exerted on the electrons in the QW decreases with the square of the distance and thus the spacer thickness and linearly with the dopant concentration. Significantly, $\tau_r$ and $\tau_q$ are affected differently by variations of the spacer layer width. Although $\tau_q$ increases as the Coulomb potential is moved away, $\tau_r$ increases at an even faster rate because the scattering angle grows smaller at the same time, thereby reducing the contribution of the remote impurities to $\tau_r$ even further, and consequently, $\tau_r / \tau_q$ increases with increasing spacer thickness. When taking into account the screening potential of the electrons [208], it is crucial to consider not only the effect of screening on the remote impurities but also on the background impurities. A slightly stronger dependence of the mobility on the sheet density has been found for remote impurities as $\mu_s \propto n^{1.7}$ compared with $\mu_s \propto n^{1.4}$ for background impurities [95]. Furthermore, $\tau_r / \tau_q$ increases strongly with $n_s$ for scattering by remote impurities and by interface roughness as screening is more efficient in restricting changes in the momentum than reducing the total scattering rate. An additional type of screening of a different nature has to be
adopted in the treatment of small-angle scattering because real Coulomb interactions cannot be regarded as ideal two-body interactions. These multiple scattering events are correlated and cannot be treated as independent. In the case of remote ionized impurities, the scattering is produced by the fluctuations about the mean potential created by the donors. One way of solving this problem is to introduce the concept of statistical screening [209] which assumes that the probability of scattering by an impurity is weighted by the probability of scattering by a closer impurity. Hence, statistical screening reduces scattering by remote impurities, which results in a small increase in $\tau_\rho$, typically by a factor of around 1.5, but simultaneously, it leads to a much larger increase in $\tau_q$ and thus an overall decrease in $\tau_i/\tau_q$. Further modifications of the relaxation times have to be introduced if all multiple scattering events are taken into account which have the tendency to increase $\tau_i$ with decreasing $n_s$, while at the same time, they tend to decrease $\tau_q$ at an even faster rate resulting in a reduced value of $\tau_i/\tau_q$ for lower sheet densities [95, 210].

![Graphs showing transport and quantum relaxation times](image)

**Figure 10.3** Transport (top) and quantum (bottom) relaxation times of 2DEG samples as a function of the lattice temperature (left) and the electric field (right).

Despite having almost identical layer structures and very similar sheet densities, BF197 and BF213 (a) display different relaxation times and $\tau_i/\tau_q$ ratios due to slight deviations in the doping concentration, the supply layer width and the depletion by surface states. In the wider supply layer of BF197, surface segregation during growth creates a lower doping density in the vicinity of the interface to the spacer layer, which is responsible for a slight drop in the sheet density and a reduction of the scattering angle associated with ionized impurities. Yet, an increase of the surface potential over time (the time period between growth and experiment for BF197 is
much greater than for BF213 (a) may cause the corresponding depletion region to extend further into the supply layer, thus generating more ionized impurities. Contributing directly to the collision broadening of the Landau levels, the resulting higher density of ionized impurities in BF197 reduces $\tau_q$ by about 20% compared to BF213 (a). However, the greater distance of the ionized impurities to the electron gas in the QW results in a significantly reduced contribution to the momentum change of the electrons since the scattering angle is much smaller, and hence, $\tau_r$ is 50% larger in BF197. The described rise in $\tau_r$ and corresponding drop in $\tau_q$ results in a considerably increased value of $\tau_r/\tau_q$ in BF197 which is approximately 90% higher than in BF213 (a). Similar high values of $\tau_r/\tau_q$ between 10 and 30 have been found in other high-quality SiGe heterostructures [97, 167, 169, 211, 212], confirming the dominance of remote impurity scattering in 2DEG structures with wide quantum wells.

An additional type of scattering potential has to be taken into account when comparing BF213 (a) and BF213 (b) in order to explain the significant changes of the relaxation times in the aged sample. Since both samples have identical layer structures and similar sheet densities, deviations of the transport properties can potentially be caused by modifications in the supply layer, an increase in threading dislocations or by strain modulation in the channel. Again, the depletion region associated with the surface potential continuous to expand into the supply layer over time, which affects $\tau_q$ to a larger extent than $\tau_r$. In addition, partial relaxation not only generates new threading dislocations but also causes existing ones to nucleate through the previously relaxed buffer layer to create misfit dislocations at the lower interface of the QW where inhomogeneities in the array of misfit segments are known to cause strain variations in the QW [213, 214]. If the threading dislocation density increases above $10^8$ cm$^{-2}$, the related scattering potential can become strong enough to reduce the relaxation times significantly [101]. Scattering from strain distributions, on the other hand, are not likely to influence the electron transport in structures with thick buffer layers as in the present samples [208]. The observed reduction of the relaxation times in the aged sample can therefore be attributed to increased scattering from remote ionized impurities and possibly a higher density of threading dislocations which causes a substantial drop in $\tau_q$ of 40%, whereas $\tau_r$ only decreases by about half that amount. In spite of its potentially strong influence on the
relaxation times, scattering from threading dislocations is not expected to have become dominant in the aged sample because the structural relaxation process is moderate and hence the threading dislocation density is assumed to stay relatively low. The continued dominance of remote impurity scattering in BF213 (b) is further corroborated by an increase in $\tau_s / \tau_q$ of almost a third, which is consistent with a reduction of the long-range scattering angle associated with the ionized impurities in the depletion region of the surface potential.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\tau_s$ (ps)</th>
<th>$\tau_q^1$ (ps)</th>
<th>$\tau_q^2$ (ps)</th>
<th>$\tau_s / \tau_q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF197</td>
<td>17.7</td>
<td>0.90</td>
<td>0.78±0.09</td>
<td>19.7</td>
</tr>
<tr>
<td>BF213 (a)</td>
<td>11.8</td>
<td>1.13</td>
<td>1.02±0.09</td>
<td>10.4</td>
</tr>
<tr>
<td>BF213 (b)</td>
<td>9.2</td>
<td>0.68</td>
<td>0.61±0.05</td>
<td>13.5</td>
</tr>
<tr>
<td>BF313</td>
<td>5.7</td>
<td>0.51</td>
<td>0.58±0.04</td>
<td>11.2</td>
</tr>
<tr>
<td>BF314</td>
<td>6.1</td>
<td>0.59</td>
<td>0.67±0.05</td>
<td>10.3</td>
</tr>
<tr>
<td>BF741</td>
<td>3.7</td>
<td>0.51</td>
<td>−</td>
<td>7.3</td>
</tr>
<tr>
<td>BF835</td>
<td>1.7</td>
<td>0.24</td>
<td>−</td>
<td>7.1</td>
</tr>
<tr>
<td>BF851</td>
<td>1.5</td>
<td>0.29</td>
<td>−</td>
<td>5.2</td>
</tr>
<tr>
<td>BF852</td>
<td>2.3</td>
<td>0.54</td>
<td>−</td>
<td>4.3</td>
</tr>
</tbody>
</table>

1 from Dingle plot  
2 from $\ln(\Delta\rho/\rho_0) vs \ln(x/sinhx)$ plot of Figure 8.14

Table 10.1 Transport and quantum relaxation times of 2DEG samples at base temperature. The ratio $\tau_s / \tau_q$ is calculated using $\tau_q$ from the Dingle plot.

A remarkable resemblance of $\tau_s / \tau_q$ can be seen between the uniformly doped structure of BF213 (a) and the delta-doped samples BF313 and BF314, indicating a very similar dominance of remote impurity scattering in all three samples. However, the individual relaxation times in the delta-doped samples are considerably shortened due to the different concentration profile of the dopants in the supply layer. The delta-deposition of the donors in BF313 and BF314 results in a high doping concentration near the spacer layer which is not only higher in density but also closer to the QW than in BF213 (a). Displaying otherwise the same layer structure, except for small variations in the virtual substrate, the quality of the channel and the substrate is expected to be identical in all samples. Under these conditions, the extremely reduced values of $\tau_s$ and $\tau_q$ in the delta-doped samples can be explained by the high density and the close proximity of the remote impurities. Consequently, the scattering angle is larger in the delta-doped samples, which results in the remote impurities having a relatively stronger influence on $\tau_s$ and $\tau_q$ compared to BF213 (a), but a corresponding increase in $\tau_s / \tau_q$ is compensated by the amplified screening.
effect of the much higher sheet density. As the only difference between the deltadoped samples, the slightly higher doped delta-deposition in the supply layer of BF313 results in reduced values of \( \tau_i \) and \( \tau_q \) due to increased scattering from remote impurities, and despite coinciding stronger screening from a correlated higher sheet density, \( \tau_i / \tau_q \) increases marginally.

The samples of the second set are nominally reproductions of the structure used in BF213 and differ only in the grading and composition of the virtual substrate and the growth temperature. The effective contribution from remote impurity scattering is therefore expected to be very similar to that of BF213 (a), but the relatively low \( \tau_i / \tau_q \) ratios indicate the presence of another scattering potential that may dominate the electron transport in these samples. Except for remote impurity scattering, the only other scattering process that can produce \( \tau_i / \tau_q \) ratios much larger than unity is interface roughness scattering [95]. Although interface roughness is considered to produce large-angle scattering and should thus contribute equally to both relaxation times, the effect of interface roughness scattering on \( \tau_i \) can become comparable to that of small-angle scattering for very high values of the lateral correlation length \( \Lambda (k_F \Lambda >> 1) \) in which case the ratio is given by \( \tau_i / \tau_q = (k_F\Lambda)^2 / 3 \). However, it has been shown that interface roughness scattering only limits the mobility in narrow quantum wells when the channel width is less than 4 - 5 nm [165, 173, 174, 215] which is about half the width of the investigated structures. Hence, interface roughness scattering is not considered to be dominant in any of these samples but its contribution may still be significant [102]. A higher density of threading dislocations in the buffer layer is another plausible source of increased scattering whereas scattering from strain distributions is still expected to be irrelevant in these structures due to their very thick buffer layer. However, scattering from threading dislocations is not expected to be dominant either as in this case, \( \tau_i / \tau_q \) would be close to unity. It concludes that remote impurity scattering remains dominant in the samples of the second set, but a strong influence of scattering by threading dislocations and/or interface roughness scattering reduces \( \tau_i / \tau_q \) considerably. Similar combinations of small- and large-angle scattering can also be seen in other Si_{1-x}Ge_x structures which are in general of relatively low quality [164, 216-218].

As a prominent feature of most samples, it can be seen that \( \tau_q \) has a stronger temperature dependence than \( \tau_e \) resulting in an increase of the ratio \( \tau_i / \tau_q \) with temperature. This renders deformation-potential scattering comparable to small-angle scattering although acoustic phonons are not considered to be Coulomb potentials, and their collisions with electrons are only approximately elastic. To explain the different temperature dependences, it is necessary to look at the restrictions of the allowed collisions. In a degenerate gas, the Fermi energy is much larger than the acoustic-phonon energy \( (E_F >> h\omega_q) \). Consequently, scattering is concentrated around the Fermi surface, and the electron wavevector is described by the Fermi wavevector \( k_F \). In particular, phonon absorption can only take place if the electron energy fulfils the condition \( E \geq E_F - h\omega_q \), and equivalently, phonon emission is only allowed for \( E \leq E_F + h\omega_q \). From these conditions, it is clear that only phonons with long wavelengths and thus small wavevectors can interact with electrons at the Fermi level which have much larger wavevectors. Since the corresponding
momentum change of the electron is very weak, these collisions are considered to be quasi-elastic, and although acoustic-phonon scattering involves an energy exchange, it is small compared with $K_B T_e$, allowing for this type of collisions to be treated like remote impurity scattering.

10.2 **Dominant Scattering in 2DHG Samples**

Showing no signs of parallel conduction or inhomogeneities, the Dingle plots of the four $p$-type samples are well behaved at all temperatures, and compared to the $n$-type samples, they display a relatively weak $T_1$- and $T_\eta$-dependence. However, the very large spread of the Dingle plot gradients between the individual 2DHG samples indicates significant changes in the magnitude of the scattering potential for SiGe alloy channels of different Ge composition and strain. Both the temperature and the structural dependence of the Dingle plots are illustrated in Figure 10.5. It is important to mention the congruence between the Dingle plots of the original and the aged samples of BF218, which implies that $\tau_\eta$ is virtually unaffected by the ageing process despite a slight decrease of the mobility and hence $\tau_1$ over time.

![Figure 10.5](image.png)

It is apparent from Figure 10.6, which shows both relaxation times as a function of the lattice temperature and the electric field, that $\tau_1$ and $\tau_\eta$ have almost identical values in the present samples. This similarity of the relaxation times generates $\tau_1/\tau_\eta$ ratios of approximately unity, which suggests large-angle scattering to be dominant in these 2DHG structures. To determine which of the short-range scattering potentials is strongest and hence dominates the hole transport, only the most likely scattering mechanisms will be considered. In particular, theses scattering processes are due to interface roughness, interface charge, alloy disorder and strain fluctuations [107, 113]. Theoretical calculations predict $\tau_1/\tau_\eta = 1$ for interface charge scattering and $\tau_1/\tau_\eta = 0.67$ for interface roughness scattering as well as for alloy scattering [95]. A $\tau_1/\tau_\eta$ ratio smaller than unity can be achieved due to the backscattering character of the large-angle scatterers, whereby the rate at which the
net momentum of all holes is randomized will be reduced more rapidly than the electronic momentum eigenstate. Providing information on how acoustic-phonon scattering influences the two relaxation times individually, the temperature and electric field dependence of the $\tau_r/\tau_q$ ratios are illustrated in Figure 10.7. Subsequently, the values of $\tau_r$, $\tau_q$, and $\tau_r/\tau_q$ at 0.1 K and low electric field (to avoid any potential heating of the holes) are summarized in Table 10.2.

![Graph showing relaxation times vs. temperature and electric field](image)

**Figure 10.6** Transport (top) and quantum (bottom) relaxation times of 2DHG samples as a function of the lattice temperature (left) and the electric field (right).

Alloy disorder and strain fluctuations have to be taken into account when analysing the short-range scattering because they are potentially very strong scattering mechanisms. Obviously, alloy scattering depends strongly on the Ge concentration in the channel and is strongest in a Si$_{0.5}$Ge$_{0.5}$ layer. Scattering from strain fluctuations, on the other hand, is influenced by the quality of the top interface because fluctuations in the valence band edge are not only induced by misfit dislocations, as in the case of the 2DEG structures, but can also arise from interface roughness [213]. However, the screening effect induced by other holes reduces alloy scattering significantly even at moderate sheet densities [95, 113, 114, 199], and scattering from strain fluctuations is also reduced considerably by the presence of screening [114]. As a result, the low-temperature mobility of the holes should not be limited by either of these two scattering potentials in high-quality samples with a relatively low Ge content. Assuming standard values for the average height $\Delta$, the correlation length $\Lambda$ and the density of negatively charged interface impurities, interface roughness scattering has been shown to dominate the hole mobility at higher hole densities and interface impurity scattering at lower hole densities [112-114], and the two competing
mechanisms can be distinguished in each individual case by their $\tau_\parallel/\tau_\perp$ ratio. Although the exact role of carrier screening remains controversial, measured values of the screening coefficient for scattering by interface charge and interface roughness were found to be consistent with those deduced from analytic expressions [219]. Including furthermore the multiple scattering theory in the analysis of the hole mobility leads to slight deviations in the $n_\text{p}$-dependence of the theoretical mobility, but this approach is necessary to interpret the experimental mobilities satisfactorily [220]. Since the quality of the interface is strongly influenced by the growth process, variations of individual growth parameters like temperature and pressure can be used to reduce the interface roughness or the charged impurity density at the interface [115, 184, 186, 187].

Figure 10.7 Ratio $\tau_\parallel/\tau_\perp$ of all 2DHG samples as a function of the lattice temperature (left) and the electric field (right).

When comparing the relaxation times of the three samples that have a uniform Ge concentration in the channel, BF216, BF217 and BF218, the most prominent feature is the dramatic decline of $\tau_\parallel$ and $\tau_\perp$ with growing Ge content, indicating a significant rise in the effective scattering potential. At the same time, $\tau_\parallel/\tau_\perp$ can also be seen to decrease with increasing Ge concentration, which suggests either an enhancement of the dominant scattering potential or a continuously growing influence of a second scattering mechanism. Since the value of $\tau_\parallel/\tau_\perp$ falls from 1.37 to 1.0, the present structures are clearly dominated by interface charge scattering despite a fairly high sheet density in BF218. The transition of the dominant scattering mechanism from interface charge to interface roughness, however, depends strongly on the proportion between impurity density and interface roughness. For a rather smooth interface with a comparatively high density of charged impurities, this transition is shifted to much higher sheet densities. Consequently, the values of the $\tau_\parallel/\tau_\perp$ ratios suggest either a fairly high impurity density at the interface, a relatively smooth interface or, to a certain extend, a combination of both scenarios. Except for slight modifications during the growth of the channel to allow for different Ge compositions, all three samples are nominally grown under identical conditions, which suggests a constant impurity density at the interface in all samples. The interface roughness, on the other hand, is also dependent on the strain in the channel, and as a result, the interface is likely to become rougher with increasing Ge content. This would explain the observed drop in $\tau_\parallel/\tau_\perp$ as interface charge scattering stays constant while interface
roughness scattering grows stronger and thus increases its influence on the total scattering potential. Although interface charge scattering remains dominant, a certain contribution from interface roughness scattering has to be included to account for the variations in $\tau_r/\tau_t$. This interpretation of the results is also corroborated by a further drop in $\tau_r/\tau_t$ in BF218 (b) which indicates an even stronger influence of the interface roughness due to partial relaxation of the strained channel over time. In this case, interface roughness scattering has become almost as strong as interface charge scattering. A much stronger total scattering potential in the compositionally graded structure of BF638 is reflected in very low values of both relaxation times. In agreement with this analysis, the ratio of $\tau_r/\tau_t$ has fallen to a value that suggests approximately equal contributions from scattering by interface charge and interface roughness. A stronger influence of interface roughness scattering might be due to a significant reduction of the well width or the presence of two further interfaces closer to the hole gas of which one even lies clearly within the limits of the wavefunction.

<table>
<thead>
<tr>
<th></th>
<th>$\tau_t$ (ps)</th>
<th>$\tau_q^1$ (ps)</th>
<th>$\tau_q^2$ (ps)</th>
<th>$\tau_r/\tau_q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF216</td>
<td>0.74</td>
<td>0.62±0.11</td>
<td>0.73±0.08</td>
<td>1.19±0.18</td>
</tr>
<tr>
<td>BF217</td>
<td>1.37</td>
<td>1.00±0.27</td>
<td>1.07±0.15</td>
<td>1.37±0.29</td>
</tr>
<tr>
<td>BF218 (a)</td>
<td>0.31</td>
<td>0.31±0.05</td>
<td>0.35±0.03</td>
<td>1.00±0.14</td>
</tr>
<tr>
<td>BF218 (b)</td>
<td>0.26</td>
<td>0.30±0.04</td>
<td>0.35±0.03</td>
<td>0.87±0.10</td>
</tr>
<tr>
<td>BF638</td>
<td>0.09</td>
<td>0.11±0.06</td>
<td>0.15±0.02</td>
<td>0.82±0.29</td>
</tr>
</tbody>
</table>

$^1$ from Dingle plot  $^2$ from ln($\Delta\rho/\rho_\infty$) vs ln($\chi/sinh\chi$) plot of Figure 8.18

Table 10.2 Transport and quantum relaxation times of 2DHG samples at 0.1 K. The ratio $\tau_r/\tau_q$ is calculated using $\tau_q$ from the Dingle plot.

Interestingly, BF216 and BF217 both have quantum relaxation times comparable to those of the 2DEG samples, which implies similar collision broadening of the Landau levels in the different types of structures. However, the greatly reduced effect of long-range scatterers on $\tau_t$ guarantees much higher mobilities in the n-type samples compared to the p-type structures which are dominated by short-range scattering. Similar values of $\tau_r/\tau_t$ in the range between 0.7 and 1.3 have been found for more or less comparable 2DHG structures [145-147, 190, 221]. Lower $\tau_r/\tau_q$ ratios of 0.67 and even below were obtained for p-type samples with well widths below 10 nm, corroborating the importance of interface roughness scattering in structures with thin quantum wells [155]. An impressively high ratio of 2.13 was achieved in a p-type structure grown on bonded SOI material but otherwise similar to the present samples, which demonstrates a considerably reduced contribution of short-range scatterers, and remote impurity scattering is beginning to be dominant in this structure [189]. However, the highest $\tau_r/\tau_q$ ratios with value of up to 5.9 have been found in samples with a pure Ge channel despite a relatively thin quantum well.
[157]. Such a high value is certainly comparable to the values obtained in a number of 2DEG structures and correspondingly, remote impurity scattering can be assumed to be dominant. However, it has been shown recently in n-type AlGaN/GaN heterostructures that the common interpretation of the $\tau_\text{r}/\tau_\text{q}$ ratio is valid only if both relaxation times are dominated by the same scattering mechanism [222]. By calculating the transport and quantum relaxation times according to eqns (4.42) and (4.43), respectively, it was argued that $\tau_\text{r}$ and $\tau_\text{q}$ are dominated by different scattering processes, and hence, the experimentally determined values of $\tau_\text{r}/\tau_\text{q}$ alone are not sufficient to identify the nature of the dominant scattering mechanism. Under these circumstances, large ratios of $\tau_\text{r}/\tau_\text{q}$ may be obtained even when short-range scattering is dominant.

10.3 Summary

The assessment of the transport and quantum relaxation time and their ratio represents an efficient method to obtain vital information about the dominant scattering process in the present heterostructures, whereby the transport relaxation time can be derived from the low-field mobility, and the slope of the Dingle plot yields the quantum relaxation time. With $\tau_\text{r}/\tau_\text{q}$ ratios of 10 and 11, the high-quality n-type samples of the first set are dominated by remote ionized impurity scattering, but the presence of some short-range scattering from the interface and the virtual substrate prevents the ratio of $\tau_\text{r}/\tau_\text{q}$ to exceed a value of 20 as is typically achieved in GaAs/AlGaAs heterostructures. A much weaker contribution from short-range scattering can be seen in one n-type sample for which the ratio of $\tau_\text{r}/\tau_\text{q}$ almost reaches a value of 20, suggesting a much improved quality of the interface and the virtual substrate. Another important difference in the present 2DEG samples is the apparent variation of the structural integrity with regard to the aging process. The observed invariance of the sheet density, the electron mobility and the $\tau_\text{r}/\tau_\text{q}$ ratio in both delta-doped samples within the given time period suggests a temporally stable structure that does not deteriorate with time. The uniformly doped sample, on the other hand, experiences a slight drop in the sheet density and the mobility, implying a structural degradation in this sample over time. This change, however, is thought to be caused by an increase of the surface potential due to changes in the oxide rather than by the deterioration of the interface quality or the virtual substrate since the ratio of $\tau_\text{r}/\tau_\text{q}$ can be seen to increases with time. Hence, the delta-doping technique appears to provide an effective screening potential that compensates for the added ionized impurities in the extended depletion region of the surface potential. Even stronger contributions from short-range scattering by threading dislocations are responsible for considerably reduced ratios of $\tau_\text{r}/\tau_\text{q}$ in the low-quality n-type samples of the second set which exhibit values roughly between four and seven. Although all of the present p-type structures are dominated by short-range scattering, the relative contribution from interface charge and interface roughness scattering varies with changing Ge concentration. Displaying $\tau_\text{r}/\tau_\text{q}$ ratios of approximately one, the
single–step structures are dominated mainly by interface charge scattering, but the observed reduction in the value of $\tau_i/\tau_q$ with increasing Ge concentration indicates a growing contribution from interface roughness scattering. In the sample with the 30% Ge content, a time-related partial relaxation of the strained channel causes the upper interface to become rougher. The corresponding increase of the interface roughness scattering results in a drop in the hole mobility as well as the $\tau_i/\tau_q$ ratio. The step-graded structure, however, exhibits the lowest value of $\tau_i/\tau_q$, suggesting approximately equal contributions from interface charge and interface roughness scattering. The coinciding drop in $\tau_i$ as well as in $\tau_q$ compared to the uniform structures can be explained by the proximity of the upper interface due to a narrower quantum well and the presence of two additional interfaces in the boundary of the wavefunction.
11 Deformation Potential

At finite temperatures, thermal vibrations of the lattice cause intrinsic scattering in a crystal which is one of the most important processes to influence the transport of the charged carriers. Electrons and holes interact with lattice vibrations by means of absorption and emission of phonons. It is convenient to describe these interactions in terms of the deformation potential on the assumption that local deformations produced by the lattice waves are similar to those in homogeneously deformed crystals [223-225]. Phonon processes are thus related to the shift of the band edge energy associated with a homogeneous strain of magnitude equal to the local strain at a given point due to the phonon wavevector. In the nonpolar materials Si and Ge, only acoustic and optical modes contribute to the phonon processes and at temperatures below ~100 K, scattering by optical phonons can be neglected due to their relatively high energy compared to the average electron and hole energies in this temperature range. Hence, all investigations of the phonon processes in this temperature range are focused exclusively on interactions with acoustic phonons.

With regard to the acoustic modes, the dilation and uniaxial deformation potentials of the conduction band are given by $\Xi_\alpha$ and $\Xi_\alpha$, respectively [224], and for the valence band, $a$ is the hydrostatic and $b$ and $d$ the shear deformation potentials [225]. Calculated and experimental values of the deformation potential constants described in literature are summarized in Table 11.1 for Si and Ge, where $\Xi_\alpha$ refers to the $\Delta$ minima. In addition, the optical phonon energy and significant material constants used in analysing the deformation potentials are also listed.

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>references</th>
<th>Ge</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi_\alpha$ (eV)</td>
<td>8.5 - 10.5</td>
<td>[42, 111, 226-230]</td>
<td>7.2 - 10.2</td>
<td>[42, 111, 226, 229, 230]</td>
</tr>
<tr>
<td>$a$ (eV)</td>
<td>2.1</td>
<td>[111, 226, 231]</td>
<td>2.0</td>
<td>[111, 231]</td>
</tr>
<tr>
<td>$-b$ (eV)</td>
<td>2.1 - 2.4</td>
<td>[42, 111, 226-230]</td>
<td>2.1 - 3.1</td>
<td>[42, 111, 226-232]</td>
</tr>
<tr>
<td>$-d$ (eV)</td>
<td>3.7 - 5.3</td>
<td>[42, 111, 226-228]</td>
<td>3.7 - 7.0</td>
<td>[42, 111, 226, 227, 232]</td>
</tr>
<tr>
<td>$\hbar\omega_0$ (meV)</td>
<td>63</td>
<td>[35]</td>
<td>37</td>
<td>[35]</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>11.9</td>
<td>[35]</td>
<td>16</td>
<td>[35]</td>
</tr>
<tr>
<td>$\rho$ (kg/m$^3$)</td>
<td>2328</td>
<td>[35]</td>
<td>5326.7</td>
<td>[35]</td>
</tr>
<tr>
<td>$v_s$ (m/s)</td>
<td>9130</td>
<td>[233]</td>
<td>5310</td>
<td>[233]</td>
</tr>
</tbody>
</table>

Table 11.1 Deformation potential constants, optical phonon energy $\hbar\omega_0$, dielectric constant $\varepsilon$, mass density $\rho$ and longitudinal sound velocity $v_s$ for Si and Ge.
Since $n$-type structures usually have a tetragonally strained Si layer, it is only necessary to provide values for $\Xi^d$ which, like $b$, corresponds to tetragonal distortions. Equivalently, $\Xi^L$ and $d$ correspond to trigonal distortions and can thus be ignored in the present samples. Both hydrostatic deformation potentials $\Xi^L$ and $a$ are difficult to calculate or measure because they refer to energy changes in the bands on an absolute scale [40, 42, 111]. Hence, the hydrostatic term of the deformation potential is generally described by $(\Xi^L + \frac{1}{3}\Xi^d - a)$ which refers to relative changes of the conduction band edge with respect to the valence band edge and can therefore be extracted from the calculations. However, values for the dilation deformation potential of the valence band, $a$, are listed in Table 11.1 since it can be evaluated by comparing the theoretical and experimental hole mobilities in unstrained bulk Si and Ge at room temperature. Unfortunately, two possible values for $a$ can be extracted by this approach which are $2.1/-9.7$ eV for Si and $2.0/-7.56$ eV for Ge [111]. Although both values have been used in recent calculations, also with alternative values of $-10.2$ eV for Si and $-12.4$ eV for Ge [107], it has been argued that the negative values of $a$ severely underestimate the hole mobility at 300 K since they imply very strong acoustic-phonon scattering. Due to the experimental uncertainty of the mobilities at room temperature, it is not possible to determine the parameter $a$ with a precision better than about 0.5 eV [111].

In order to justify the assumption that optical phonon scattering can be neglected at temperatures below approximately 100 K, it is useful to examine the typical energies of the acoustic phonons and the associated carriers. The average energies of the acoustic phonons that are involved in the hole scattering process in Si are 2 meV at 150 K and 3.3 meV at 300 K which have to be compared to the average hole energies of 16.5 meV and 43.1 meV at the respective temperatures. Accordingly, the average acoustic-phonon energies in Ge are 0.7 meV at 100 K and 1.4 meV at 300 K, and the corresponding average hole energies are 10 meV and 41.5 meV, respectively [234].

The deformation potential has been introduced initially to describe phonon processes in unstrained bulk crystals which can be used to evaluate the corresponding relaxation times within the transport theory and which is indispensable to establish the band structure of a semiconductor. In presence of uniaxial strain, as is the case in the quantum well of 2DEG and 2DHG structures, the individual deformation potential constants can be expected to change according to the amount of stress exerted on the material [226, 229]. Additionally, the effect of alloying in the active layer of 2DHG structures is also presumed to alter the deformation potential compared to that in bulk Si and Ge. Despite the deformation potential also exhibiting a certain temperature dependence [227, 235], it is justified to ignore any such change of the deformation potential since the temperature coefficient $d\Xi^d/dT = 0.04$ meV/K is negligibly small compared to the general spread of the calculated and measured values in Table 11.1.

Since all of the present samples are uniaxially strained and the stress is thought to vary slightly in the $n$-type structures but quite considerably in the $p$-type samples, it is desirable to investigate the corresponding deformation potential in each individual sample. In general, the appropriate set of deformation potential parameters is obtained from experiments which measure shifts and distortions of the bands upon
application of selected strains to the crystal [227, 228, 236-239]. Alternatively, the deformation potential can be calculated numerically, and in certain approaches, it has been found necessary to regard these interaction parameters as adjustable in order to obtain a quantitative agreement with the experimentally determined transport data. Within the framework of the relaxation-time approximation, separate analytical solutions for the deformation potential can be found in the high-temperature as well as in the low-temperature limit. In the equipartition (EP) regime, where $\hbar \omega_k < k_B T$, the phonon absorption and emission rates are equivalent and the reciprocal mobility $\mu^1$ exhibits a simple linear dependence on temperature since the equilibrium Bose-Einstein occupation factor can be approximated to $k_B T / \hbar \omega_k$ which is linear in temperature. Measuring the mobilities over a wide temperature range in the EP regime and defining the momentum relaxation time as a function of the deformation potential provides an analytical result for the deformation potential. In the Bloch-Grüneisen (BG) regime, where $\hbar \omega_k \geq k_B T$, the constraints of degeneracy and energy conservation lead to a drastic reduction in the momentum relaxation rate at the Fermi energy due to phase-space restrictions for scattering via acoustic phonons which gives rise to a stronger temperature dependence of the scattering rate. Apart from investigating the reciprocal mobilities in the BG regime (only in high-mobility samples), the deformation potential can also be established by studying the energy-loss rate (ELR) as a function of the electron temperature since the energy relaxation of the hot carriers is accomplished by interactions with acoustic phonons.

The transition temperature at which the BG regime changes over into the EP regime ranges between 10 K in the low-density $n$-type samples and 13 K in the high-density $n$-type samples. In the 2DHG structures where the alloy composition in the channel has to be considered as well, the transition temperatures are 9 K, 11 K, 15 K and 16 K for samples BF217, BF216, BF638 and BF218, respectively. These are calculated values since the measurements do not cover the crucial temperature range between ~4.2 K and ~10 K to establish the transition temperature experimentally.

### 11.1 EQUIPARTITION REGIME

Apart from acoustic-phonon processes, all other scattering mechanisms, e.g. by remote ionized impurities or interface roughness, are approximately independent of temperature between 20 and 60 K, and hence, any temperature dependence of the mobility can only arise from acoustic-phonon scattering in this temperature range. In the equipartition regime, the carrier distribution is degenerate in most 2D heterostructures but, unlike in the Bloch-Grüneisen regime, the effect of Pauli exclusion is small and therefore does not restrict the allowed phonon processes. When describing the scattering rate in a degenerate system, the occupation of the final state has to be taken into account, and it can be assumed that a Fermi-Dirac distribution is maintained at all times. Additionally, the Fermi energy is presumed to be much larger than the energy of the acoustic phonons.
\[\hbar \omega_q \ll E_F\] (11.1)

so that scattering is concentrated around the Fermi surface and may be considered quasielastic. The acoustic-phonon dispersion relation is given by

\[\omega_q = v_s q\] (11.2)

where \(q\) is the phonon wavevector and \(v_s\) is the longitudinal sound velocity, assuming that only longitudinal phonon modes interact with the carriers in 2D systems. The Fermi energy can be calculated according to

\[E_F = \frac{2\pi \hbar^2 n_s}{g_s g_0 m^*}.\] (11.3)

In the EP regime, the relation between the acoustic-phonon energy and the thermal energy is given by

\[\hbar \omega_q < k_B T\] (11.4)

which allows the following approximation for the Bose-Einstein distribution of the acoustic phonons

\[n(\omega_q) = \frac{1}{\exp(\hbar \omega_q / k_B T) - 1} - \frac{k_B T}{\hbar \omega_q},\] (11.5)

clearly demonstrating a linear \(T\)-dependence of the phonon population. A further simplification can be applied in the temperature range below \(\sim 50\) K where only the lowest subband is populated, thereby restricting any phonon process to intrasubband scattering. Under these conditions, the acoustic-phonon scattering rate is given by the momentum relaxation time as \([107, 240-244]\)

\[\frac{1}{\tau_{ac}} = \frac{3\Sigma^2 m^* k_B T}{2\hbar^3 c_L L},\] (11.6)

where \(L\) is the width of the confinement potential and \(c_L = \rho v_s^2\) is the elastic constant associated with acoustic vibrations with \(\rho\) being the mass density and \(v_s\) the longitudinal sound velocity. Since any temperature dependence of the mobility is caused by acoustic-phonon scattering, the mobility can be written as \([245, 246]\)

\[\mu^{-1}(T) = \mu_0^{-1} + \mu_{ac}^{-1}(T) = \mu_0^{-1} + \alpha T,\] (11.7)

where \(\mu_0\) represents the mobility limited by all temperature independent processes (the \(T\)-dependence of Coulomb scattering is negligible at temperatures above \(\sim 20\) K \([161]\)) and \(\alpha T\) is proportional to the scattering rate by acoustic phonons. Since the acoustic-phonon density is proportional to temperature, a linear increase in the resistivity with temperature is indicative of increasing acoustic-phonon scattering.
The slope of this curve represents the strength of the carrier-phonon interaction which is related to the deformation potential for the corresponding energy band via

\[
\frac{1}{\mu_{sc}} = \alpha T = \frac{3 \Xi^2 m^* \kappa_B T}{2 e \hbar^2 c_L L}. \tag{11.8}
\]

Assuming the triangular potential well in which the carriers are confined near the heterointerface to be smaller than the width of the quantum well, the effective well width \( L \) can be calculated as

\[
L = \left( \frac{\hbar^2 \varepsilon \varepsilon_0}{4 m^* \varepsilon^2 n_s} \right)^{\frac{1}{3}} \tag{11.9}
\]

where \( \varepsilon \) is the dielectric constant and \( n_s \) is the sheet density. So far, all parameters are known except for the deformation potential which can subsequently be calculated by transforming eqn. (11.8) into

\[
\Xi = \left( \frac{2 e \hbar^2 c_L}{3 m^* \kappa_B} \right)^{\frac{1}{2}} \tag{11.10}
\]

This equation for the deformation potential is generally valid for both \( n \)-type and \( p \)-type structures in the EP regime. However, since \( \Xi \) refers to an effective acoustic deformation potential, its exact composition is ambiguous in both the conduction and the valence band. Hence, a closer examination of its general character in the corresponding energy band is presented in the subsequent sections.

### 11.1.1 Reciprocal Electron Mobility

All measurements of the mobility in the equipartition regime were performed in the closed cycle cryostat during the warm-up cycle. Despite a slight hysteresis in the Hall voltage between the cool-down and warm-up cycle, the corresponding individual values of the mobility differ only by a few percent, and the gradient of the mobility is affected even less since the effect of the hysteresis on the mobility can be interpreted as a simple offset that alters the mobility at each temperature equally. As the 2DEG samples of the first batch were measured shortly before the second set of measurements in the VTI and dilution fridge, the results of sample BF213 refer to its aged structure. Unfortunately, BF197 could not be measured in the closed cycle cryostat because of bad physical contacts that could not be bonded to the TO8 header used in this system. The reciprocal mobilities of both sets of 2DEG samples are shown in Figure 11.1 between 10 K and 60 K. In all samples, the increase of the gradient at temperatures above ~40 K indicates the onset of parallel conduction due to the population of the second subband. A changeover into the Bloch-Grüneisen regime can be seen in the high-density samples at the lower end of the temperature
range at which the reciprocal mobilities start to decrease more rapidly. In the low-density samples, the transition occurs at temperatures too low to affect the mobilities in the given temperature range.

![Graph showing temperature dependence of the reciprocal electron mobilities in the equipartition regime.](image)

*Figure 11.1* Temperature dependence of the reciprocal electron mobilities in the equipartition regime. At temperatures above 40 K - 50 K, the onset of parallel conduction causes the gradient of the reciprocal mobilities to increase.

The linear range of the reciprocal mobilities lies roughly between 15 K and 40 K which is large enough to yield a gradient of high accuracy with an error margin of less than 1%. The values of the material constants necessary to calculate the width of the potential well as well as the deformation potential are considered to be identical to those in bulk Si as listed in Table 11.1. Subsequently, the gradient $\alpha$, the potential well width $L$ and the resulting deformation potential $\Xi$ are summarized in Table 11.2. In all samples, the gradient of the reciprocal mobility is smaller in the EP regime than in the BG regime where the gradient correlates predominantly to the temperature dependence of Coulomb scattering. In agreement with previous reports [246], the values of $\alpha$ in the EP regime tend to increase monotonically with the electron concentration. However, the gradient also increases with $\Xi$ which, apart from $n_e$, is the only other parameters in eqn. (11.6) that varies between samples. Hence, it is important to establish whether sheet density and deformation potential are correlated. Although the sheet density determines the width of the potential well, its ultimate impact on the deformation potential is counterbalanced by its opposite influence on the gradient of the reciprocal mobilities. Consistent with theory, the deformation potential is therefore mathematically independent of the sheet density despite its direct occurrence in eqn. (11.9) and its implicit involvement in eqn. (11.10). Since there are also no alloying effects in the strained Si channel of 2DEG structures, differences in $\Xi$ can thus only arise from existing strain variations. All three samples of the first batch exhibit very similar deformation potentials between 3.4 and 3.8 eV, suggesting almost identical strain in the active layer in each of these sample. Apart from BF741 and BF852, the samples of the second batch also display comparable yet slightly higher values of $\Xi$, ranging from 4.5 to 4.9 eV, which again indicates virtually equal strain in those samples. Another interesting aspect is the relative short width of the triangular potential well compared to the thickness of the strained Si channel which at 10 nm is about five times larger. Assuming hypothetically the rectangular potential of the quantum well to be the confinement for the electrons, the values of the deformation potential would roughly more than double.
<table>
<thead>
<tr>
<th>L (nm)</th>
<th>$\alpha$ ($10^7$ V s/cm$^2$K)</th>
<th>$\Xi$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF213</td>
<td>2.21</td>
<td>3.4</td>
</tr>
<tr>
<td>BF313</td>
<td>2.17</td>
<td>3.8</td>
</tr>
<tr>
<td>BF314</td>
<td>1.81</td>
<td>3.6</td>
</tr>
<tr>
<td>BF741</td>
<td>1.98</td>
<td>4.1</td>
</tr>
<tr>
<td>BF743</td>
<td>2.13</td>
<td>4.7</td>
</tr>
<tr>
<td>BF834</td>
<td>2.19</td>
<td>4.8</td>
</tr>
<tr>
<td>BF835</td>
<td>1.93</td>
<td>4.9</td>
</tr>
<tr>
<td>BF851</td>
<td>1.92</td>
<td>4.5</td>
</tr>
<tr>
<td>BF852</td>
<td>1.80</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Table 11.2 Effective deformation potential $\Xi$ of the 2DEG samples. Also listed are the width $L$ of the triangular potential well and the slope $\alpha$ of the $T$-dependent reciprocal mobilities.

Contributing to the problem of interpreting the effective deformation potential $\Xi$, the electron-phonon coupling in Si is complicated due to the anisotropy of the electronic energy spectrum in the conduction band, which requires the description of separate deformation potentials for the longitudinal and the transverse phonon modes as defined in chapter 4 by eqns (4.35) and (4.36), respectively. A significant simplification can be made regarding acoustic-phonon scattering in degenerate 2D structures where only the longitudinal phonons have to be considered and the transverse mode can be neglected. Although the effective deformation potential in eqn. (11.6) thus refers only to the longitudinal mode, $\Xi$ is still a combination of both $\Xi_\sigma$ and $\Xi_d$ whereby the contribution of $\Xi_d$ is dependent on the scattering angle. At small angles ($\leq 15^\circ$), $\Xi$ is simply the sum of $\Xi_\sigma$ and $\Xi_d$, but the contribution of $\Xi_d$ diminishes slightly for increasing angles up to $\sim 30^\circ$ and more strongly for even greater angles. This approach is commonly used in literature for silicon inversion layers [247-251], but the effective deformation potential is also interpreted to exclusively denote $\Xi_d$ for SiGe 2DEG systems [102, 164, 252]. In the case of bulk Si, the hydrostatic deformation potential is generally assumed to have values between $-6 \text{ eV}$ and $-8 \text{ eV}$ [247-249, 253], but in SiGe, $\Xi_d$ can fall to a value of $1.1 \text{ eV}$ and even below [42, 111] which is small compared to $\Xi_\sigma$. Therefore it is justified to assume that the effective deformation potential is equivalent to $\Xi_\sigma$ in SiGe at low temperatures where only small-angle scattering is allowed. An effective deformation potential of $10 \text{ eV}$ was deduced from measurements of the thermoelectric power in an $n$-type Si MOD structure [254], whereas energy-loss rate experiments yielded values of $12 \text{ eV}$ in $n$-type Si inversion layers [247, 248] and $9 \text{ eV}$ in SiGe 2DEG structures [164, 252], also found when comparing experimental and theoretical mobilities in SiGe at low temperatures [102].
When comparing the values of $\Xi$ in Table 11.2 to the range of calculated and experimental values of $\Xi$, for bulk Si, it is obvious that the presence of compressive uniaxial strain causes a reduction in the deformation potential of the conduction band. Furthermore, it can be generally assumed that greater strain is likely to reduce the deformation potential further than weaker strain. Despite having nominally almost identical virtual substrates, the lower values of $\Xi$ found in the samples of the first batch thus suggest a greater strain in the active layer than in the samples of the second batch. This observation is also corroborated by the results of the previous chapters where it was deduced that the second set of samples have virtual substrates of a lower quality due to different growth temperatures and grading techniques. The existence of an increased number of threading dislocations in the buffer layer and misfit dislocations near the interface to the quantum well are thought to be the cause for the Si channel not to be fully strained in the samples of the second batch. As the only sample of the second set to display a deformation potential close to those of the first set, BF741 seems to have the highest quality virtual substrate in this batch. The Si channel of sample BF852, although nominally identical to BF851, appears to be the most relaxed since its value of $\Xi$ is by far the highest. It is remarkable that the values of $\Xi$ in all of the present samples are much lower than has been found previously in comparable $n$-type structures [102, 164, 252] which hardly differ from the value for bulk Si despite being under considerable strain.

### 11.1.2 Reciprocal Hole Mobility

Since the $p$-type samples have been measured at the same time and under identical conditions as the $n$-type samples, the same considerations concerning the hysteresis of the Hall voltage apply. Analogous to sample BF213, the results of sample BF218 signify the properties of its aged structure. The very low hole mobility of sample BF638 and the relatively high noise level in the closed cycle cryostat did not allow for a signal-to-noise ratio big enough to provide sufficiently accurate readings of the Hall voltage. Due to the resulting strong spread of the data points, the results for sample BF638 are therefore omitted in this section. In Figure 11.2, the reciprocal hole mobilities are presented in the temperature range between 10 K and 60 K. Parallel conduction can be seen to set in at $\sim$45 K, which is at a slightly higher temperature than in the $n$-type samples. In samples BF216 and BF217, the data points of the reciprocal mobilities do not deviate at lower temperatures, which confirms that the transition into the BG regime occurs at temperatures too low to influence the mobilities significantly above 10 K. However, sample BF218 has a higher transition temperature, and the mobility is visibly affected below $\sim$14 K. The reciprocal mobilities display a linear dependence on temperature roughly between 15 K and 45 K, again providing a slope of very high accuracy that lies within 1%. The dielectric constant, the mass density and the longitudinal sound velocity are deduced for each individual structure by linear interpolation between the values for bulk Si and bulk Ge as all three parameters are dependent on the Ge content in the strained channel.
Figure 11.2 Temperature dependence of the reciprocal hole mobilities in the equipartition regime. The onset of parallel conduction causes the gradient to increase at temperatures above ~45 K.

The width of the triangular potential well $L$, the slope of the reciprocal mobilities $\alpha$ and the resulting effective deformation potential $\Xi$ are summarized in Table 11.3. As in the case of the $n$-type samples, a weaker temperature dependence of the mobilities in the EP regime results in lower values of $\alpha$ compared to those in the BG regime. Apart from its dependence on the existing strain in the quantum well, the effective deformation potential in $p$-type samples is also determined by the alloy composition since it influences the value of the material constants. However, the alloy composition and the strain are directly correlated as the stress in the channel increases with increasing Ge concentration. Furthermore, $\Xi$ is also governed by the hole effective mass which yet again is dependent on the alloy composition and decreases with increasing Ge concentration. From eqns (11.9) and (11.10), the proportionality factor between the effective deformation potential and the hole effective mass can be derived to be $\Xi \propto m^*^{-1.17}$, which indicates an almost linear reciprocal dependence. Due to its importance for the calculation of the appropriate values of $\epsilon$, $\rho$ and $v_\rho$, the Ge concentration in the QW is also listed in Table 11.3 as is the effective hole mass.

Relating the effective deformation potential $\Xi$ to the individual constants $a$, $b$ and $d$ of the deformation potential theory is a complex problem involving the anisotropy and the degeneracy of the valence bands. Assuming an isotropic phonon spectrum characterized by the spherically averaged longitudinal and transverse elastic coefficients

$$C_l = \frac{1}{5}(3C_{11} + 2C_{12} + 4C_{44})$$  \hspace{1cm} (11.11)$$

$$C_t = \frac{1}{5}(C_{11} - C_{12} + 3C_{44}),$$  \hspace{1cm} (11.12)$$
an appropriate theoretical effective deformation potential $\Xi_{\text{eff}}$ can be defined for computational purposes which has the form [232]

$$\Xi_{\text{eff}}^2 = a^2 + \beta \left( b^2 + \frac{1}{2} d^2 \right)$$  \hspace{1cm} (11.13)

and where $\beta = C_1/C_s$. Values for the elastic coefficients can be found for example in [255]. In this context, the parameter $\Xi_{\text{eff}}$ can be interpreted to represent a simplification of the theoretical model for pure acoustic scattering. Since $\Xi_{\text{eff}}$ still refers to the theoretical deformation potential of a simplified valence band, eqn. (11.13) has to be modified slightly to provide an adequate description of the experimental effective deformation potential $\Xi$. A simple relation between the experimental and theoretical effective deformation potential is given by the following expression [255-257]

$$\Xi^2 = \frac{\beta + 2}{6\beta} \Xi_{\text{eff}}^2.$$  \hspace{1cm} (11.14)

Using average values for the deformation potential constants $a$, $b$ and $d$ as given in Table 11.1, $\Xi$ can be calculated according to this approach to have a value of 3.7 eV in Si and 4.3 eV in Ge. However, there are still some uncertainties concerning this method since various experimental techniques indicate much higher values of $\Xi$ in Si. By comparing theoretical and experimental hole mobilities in bulk Si and bulk Ge, for example, the effective deformation potential was found to be 7.9 eV and 3.7 eV, respectively [256]. Measurements of the thermoelectric power in a $p$-Si MOD structure suggested an even higher effective deformation potential of 12 eV [254], whereas the energy-loss rate in a $p$-Si inversion layer implied $\Xi$ to be 10 eV [258]. The effective deformation potential in Ge, on the other hand, seems to be reproduced fairly accurately by the experiments. Any discrepancies between theoretical and experimental values can be explained by the presence of severe nonparabolicity in the valence band which changes the temperature dependence of the effective mobility and therefore renders eqn. (11.13) inaccurate which requires a uniform $T$-dependence [255]. Materials with extremely small spin-orbit splitting, like Si, are most likely to be affected by severe nonparabolicity which would account for the substantial disparity between the theoretical and experimental values of $\Xi$. The very large spin-orbit splitting in Ge, however, guarantees the validity and accuracy of this approach, and hence, strong agreement between measured and calculated values of $\Xi$ is achieved.

Disregarding the influence of strain in the present samples, the value of $\Xi$ can be calculated by linear interpolation and is therefore expected to lie between the bulk values of Si and Ge. If the experimental values of 7.9 eV (bulk Si) and 3.7 eV (bulk Ge) are presumed to be correct, the interpolated values of $\Xi$ are approximately 7.5, 7.0 and 6.6 eV for a SiGe heterostructure of 10, 20 and 30%, respectively. These values increase considerably when the higher experimental values of $\Xi$ for bulk Si, i.e. 10 and 12 eV, are adopted. Consequently, the comparatively low values of $\Xi$ in the present samples may possibly be explained by the presence of strain in the active layer. In agreement with this postulate, it can also be suggested that higher stress
which is associated with a higher Ge concentration causes a stronger reduction of $\Xi$. However, such a trend is not visible in Table 11.3, as the experimental data is spaced even slightly closer than the interpolated values. Although the exact relation between strain and deformation potential has not been established, experimental errors are the most likely explanation for this discrepancy with the expected behaviour.

![Table 11.3 Effective deformation potential $\Xi$ of the 2DHG samples, the width $L$ of the triangular potential well and the slope $\alpha$. For convenience, the Ge content and the hole effective mass are also listed.](image)

With a value of 5.5 eV, the closest resemblance of the deformation potential to the present results has been found in a Si$_{0.8}$Ge$_{0.2}$ sample by measuring the thermoelectric power [259]. Similar experiments performed on a Si$_{0.8}$Ge$_{0.2}$ structure, however, already suggested a much lower value of 4.5 eV [260], thereby indicating a much stronger dependence on strain than found in the present samples. Measurements of the energy-loss rate of comparable $p$-type structures, all with a 20% Ge composition in the alloy, partly confirmed the value of 4.5 eV [261] but also suggested an even lower deformation potential of only 3 eV [262, 263].

### 11.2 Bloch-Grüneisen Regime

At very low temperatures, equipartition fails, and the acoustic-phonon energies become comparable to the thermal energy, which gives rise to a more complicated temperature dependence of the relaxation rates via the statistical occupation factors [264]. Accordingly, the temperature range of the Bloch-Grüneisen regime is typically characterized by the condition

$$\hbar \omega / k_B T,$$

whereby Pauli exclusion increasingly restricts the allowed processes because the final electron state is occupied. Scattering is confined to the Fermi surface, and the process is increasingly dominated by spontaneous emission because scattering by absorption becomes negligible as the phonon population diminishes strongly and

$$n(\omega_q) = \frac{1}{\exp(\hbar \omega_q / k_B T) - 1} \ll 1.$$
Scattering electrons right across the Fermi surface requires acoustic phonons with in-plane wavevector \( q \approx 2k_F \) which may be substantial enough for the corresponding phonon energy to be greater than \( k_B T \). Pauli exclusion tends to inhibit the emission of such phonons which cease to be appreciably excited and are no longer contributing to the scattering process. Due to the energy considerations, only phonons with small wavevectors can contribute to the relaxation rate in the vicinity of the Fermi energy. As the phonon population diminishes and exclusion increasingly restricts the number of allowed processes, the scattering rate drops off rapidly as the system enters the Bloch-Grüneisen regime. The transition into the BG regime occurs roughly at a temperature \( T_{BG} \) that fulfills the condition

\[ k_B T_{BG} \approx 2\hbar v \frac{1}{k_F} \]  

where the Fermi wavevector \( k_F \) can be obtained directly from the expression

\[ k_F = \left( \frac{4\pi n_e}{G_s G_v} \right)^{1/2}. \]

A crucial point concerning the energy relaxation of the electrons is the fact that acoustic-phonon scattering is considered to be quasielastic not only in equipartition but also in the Bloch-Grüneisen regime. Whereas the momentum of the electrons and the phonons involved in this process remains unchanged, their energy may vary slightly. Consequently, hot electrons can lose energy exclusively by interactions with acoustic phonons as this is the only process that allows a limited energy exchange. Another important aspect is the possibility to characterize the Fermi-Dirac distribution of the electrons in terms of the electron temperature rather than energy. Investigating the relationship between the energy-loss rate of hot electrons and the corresponding electron temperature thus provides vital information about the temperature dependence of acoustic-phonon scattering. Since both the energy-loss rate and the electron temperature can be determined experimentally, it is possible to calculate the deformation potential by establishing a theoretical model that describes the energy-loss rate in terms of the electron temperature.

In order to study the energy loss from hot electrons to the lattice by emission of acoustic phonons, it is necessary to establish the correlation between the energy-loss rate \(-(dE/dt)\) per electron and the input power per electron \( P_e \) which is given by

\[ P_e = e\mu E^2 \]  

where \( E \) is the applied electric field. In steady state, the energy gain supplied by the electric field is exactly balanced by the energy loss to the lattice. An alternative and slightly more practical way to calculate the input power per electron from the experimental parameters is simply given by the product of the current density \( J_x \) along the Hall bar and the electric field \( E_x \) divided by the sheet density \( n_s \)

\[ P_e = J_x E_x / n_s. \]
Both methods of calculating the input power are equivalent and should produce identical results. The electron temperature can be obtained by analysing the amplitudes of the SdH oscillations at constant lattice temperature and different current densities as described in chapter 9.

Any theoretical description of the energy-loss rate has to include an expression that accounts for the momentum conversation in the direction perpendicular to the 2D layer. A common method to resolve this problem employs a numerical model that assumes the variational ansatz for the wavefunction in the lowest subband [253]. Following this approach, the energy-loss rate has been calculated for heterostructures of the III-V material GaAs/Al$_x$Ga$_{1-x}$As [266, 267], for Si inversion layers [248, 258, 268] and also for Si/Si$_{1-x}$Ge$_x$ heterostructures [164, 261]. Alternatively, the wavefunction normal to the interface can also be calculated by using a self-consistent interface potential [269, 270]. In both techniques, certain approximations for the momentum conversation are made which introduce an additional source of error. The resulting inaccuracies can only be avoided if no approximations are made and all of the final integrals are calculated numerically [271, 272]. In general, the power loss of a degenerate electron gas due to scattering by acoustic phonons can be calculated from a set of energy balance equations. The net average energy-loss rate can be evaluated from the energy gained by the 3D bulk phonons from the hot electrons

\[
\left\langle \frac{dE}{dt} \right\rangle = \int_0^\infty \frac{dE}{dt} f(E) dE \int_0^\infty f(E) dE \tag{11.21}
\]

where $E$ is the electron kinetic energy and $f(E)$ is the Fermi-Dirac distribution for an electron temperature $T_e$ that is higher than the lattice temperature $T_L$ and which has the form

\[
f(E) = \frac{1}{\exp[(E-E_F)/k_BT_e] + 1} \tag{11.22}
\]

The net energy-loss rate of an electron with energy $E$ can be written in the form

\[
\frac{dE}{dt} = \int_0^{2\pi} [E_+ (\theta) - E_- (\theta)] d\theta \tag{11.23}
\]

where the differential energy change due to scattering of the 2D electrons with wave-vector $k$ can be expressed for phonon absorption as

\[
E_+ (\theta) = \int h\omega_q \left| l(q_z) \right|^2 P(q_{xy}, q_z) N_q \left[1 - f(E + h\omega_q) \right] dq_z \tag{11.24}
\]

and equivalently for phonon emission as

\[
E_- (\theta) = \int h\omega_q \left| l(q_z) \right|^2 P(q_{xy}, q_z) (N_q + 1) \left[1 - f(E - h\omega_q) \right] \Theta(E - h\omega_q) dq_z. \tag{11.25}
\]
Notably, $\hbar \omega_q$ is placed inside the integral in eqns (11.24) and (11.25) because the differential scattering cross sections depend on the phonon energy. In eqns (11.23) - (11.25), $\theta$ is the scattering angle, $P(q_{xy}, q_z)$ is the scattering probability factor, $N_q = n(\omega_q)$ is the phonon occupation number and $\Theta(E)$ is the Heaviside function (or unit step function). The phonon components parallel and normal to the plane are denoted by $q_{xy}$ and $q_z$, respectively. The form factor $|I(q_z)|^2$ is the squared overlap integral describing the coupling of the electron wavefunction to the normal component of the scattered phonon and thus accounts for momentum conservation in the direction perpendicular to the plane. Assuming only the lowest subband to be occupied, the overlap integral relates to the narrowly localized wavefunction $\phi$ of the ground state normal to the interface via [273]

$$I(q_z) = \int \phi_0^*(z) \exp(iq_z z) dz. \quad (11.26)$$

In order to solve the overlap integral accurately, a self-consistent wavefunction has to be chosen. However, if the potential well is approximated to be triangular, a good estimate for the energy of the lowest subband can be given by using the variational wavefunction [99, 274]

$$\phi_0(z) = \left( \frac{b}{2} \right)^2 z \exp(-\frac{1}{2} b z) \quad (11.27)$$

where $b$ is the variational parameter inversely correlated to the extension of the wavefunction, thus giving an estimate of the maximum wavevector at which the electron-phonon coupling falls off. When substituting eqn. (11.27) into eqn. (11.26), the form factor reduces to the much simpler form

$$|I(q_z)|^2 = b^6/(b^2 + q_z^2)^3 \quad (11.28)$$

and a correct choice of the variational parameter $b$ minimizes the total energy of the system. The scattering probability factor can be calculated individually for each scattering mode, and for acoustic-phonon scattering via deformation-potential coupling, it is defined as [266]

$$P(q_{xy}, q_z) = \frac{m^* \Xi^2 \hbar \omega_q}{8\pi^2 \hbar^2 c_s S^2(q_{xy}, T_e)} \quad (11.29)$$

where static screening of the deformation potential is accounted for by the screening factor $S$ which depends on both the in-plane component of the acoustic phonons and the electron temperature [275-277].

Despite these simplifications, the remaining integrals still have to be calculated numerically, which is time intensive and requires specialized programmes. A slightly different approach employs not only physical but also mathematical approximations to deduce an analytic expression for the energy-loss rate. For parabolic, spherical bands in the BG regime, the energy-loss rate to phonons with small energies by unscreened deformation-potential coupling can be expressed analytically [242, 278].
Significantly, the analytic term is characterized by a simple power law for the temperature dependence in the BG regime where $P_e$ is proportional to $T_e' - T_L'$ and the exponent $\gamma$ can vary depending on conditions such as the mode of coupling and the degree of screening. Generally, unscreened deformation-potential scattering is described by $\gamma = 5$, as is the case in eqn. (11.30), whereas the exponent is reduced to $\gamma = 7$ for unscreened piezoelectric scattering. In the presence of static screening, the temperature dependence increases, and the respective exponents are raised by two. Hence, screened deformation-potential scattering can be identified by a power law of $\gamma = 7$, and screened piezoelectric scattering displays an exponent of $\gamma = 5$. A slightly different analytical expression with the same $T$-dependence has also been used successfully to calculate the energy-loss rate in weakly screened Si/Si$_{1-x}$Ge$_x$ heterostructures [262, 263]. It should be noted that in all existing theories, the description of the phonons is approximated by their bulk modes.

Since acoustic-phonon processes involve a change in the direction of the momentum but not in its magnitude, it is sensible to define a momentum relaxation time $\tau_m$ similar to the transport relaxation time $\tau$, as defined in eqn. (4.11). Although both relaxation times describe the rate at which the net momentum of all electrons is randomized, $\tau_m$ relates purely to acoustic-phonon scattering, whereas $\tau$ includes all scattering processes, i.e. acoustic-phonon and Coulomb scattering. In heterostructures with very high electron mobilities, momentum and transport relaxation times are usually identical in the BG regime. However, the relatively strong contribution of Coulomb scattering in the present samples is expected to cause significant deviations between the values of $\tau_m$ and $\tau$. Accordingly, $\tau_m$ should be somewhat larger than $\tau$ in both $n$-type and $p$-type samples, which justifies a continued separate notation of both relaxation times. In the unscreened case, the momentum relaxation time is given by [242, 278]

$$\frac{1}{\tau_m} = \frac{15\Xi^2 m^*}{2\pi^2 \hbar^2 v_s^6 k_f^3} \left[ \left( \frac{k_B T_e}{k_B T_L} \right)^6 + \left( \frac{k_B T_L}{k_B T_e} \right)^6 \right] \frac{k_B T_e}{k_B T_L}$$

(11.31)

which also displays the typical $T^5$-dependence. Since acoustic-phonon scattering is the only process to allow a limited energy exchange in the BG regime, it is possible to define an energy relaxation time $\tau_e$ for this process which can also be interpreted to describe the efficiency of the acoustic phonons to cool the hot electrons. The energy relaxation time is defined by [248]

$$\tau_e = \frac{\pi^2 k_f^2}{6E_F} (T_e^2 - T_L^2) P_e^{-1}$$

(11.32)

where the cooling efficiency decreases with increasing $\tau_e$. As the number of phonons that are emitted by the electrons increases with $T_e$, the probability rises for some of these phonons to be re-absorbed by the electrons which is likely to reduce $\tau_e$. 

246
11.2.1 ENERGY-LOSS RATE OF HOT ELECTRONS

The electron heating experiments of the first set of n-type samples were performed in the 15 T dilution fridge at a lattice temperature of 0.05 K with the exception of BF213 (a) which could only be cooled down to a base temperature of 0.1 K and BF197 which could not be cooled below 1.7 K due to problems with the dilution fridge at the time. A base temperature of 0.4 K was the minimum that could be reached in the 4 T Heliox system for the electric field experiments of the second batch of 2DEG samples. In all experiments, the magnitude of the electron temperature is limited by the resolution of the SdH oscillations as the amplitude of the oscillations decreases strongly with increasing electric field. Accordingly, the smallest viable resolution is defined by a minimal SNR that still guarantees an acceptable error level.

The input power per electron $P_e$, shown in Figure 11.3 as a function of $T_e-T_l$, has been calculated not only from eqn. (11.19) but also from eqn. (11.20), and the results of both methods display only minor deviations. A distinct feature visible in the two graphs is a general increase of the energy loss per electron with rising lattice temperature. Due to this dependence, it is only possible to compare the values of the energy-loss rate directly between samples displaying the same lattice temperature.

Figure 11.3 Energy-loss rate per electron $P_e$ vs $T_e-T_l$ for both sets of n-type samples. The lattice temperature is 0.05 K for the first batch - except for BF197 ($T_l = 1.7$ K) and BF213 (a) ($T_l = 0.1$ K) - and 0.4 K for the second batch. The analytic term of eqn. (11.30) is fitted to the data points with a fixed exponent of $\gamma = 5$ to determine the deformation potential $\Xi$. 

247
However, such a direct comparison is not necessary for the fitting procedure which is used on every sample separately to determine the corresponding value of the deformation potential $\Xi$. As demonstrated in Figure 11.3, the analytic expression of eqn. (11.30) for the energy-loss rate is fitted to the data points of each individual sample which yields a unique value for the deformation potential as the only adjustable parameter in this expression. Despite using a fixed exponent of $\gamma = 5$ for the power law of the analytic term, this approach delivers an adequate accuracy of the fitted curves with an error margin generally below $\sim 10\%$. Yet, in order to establish the exact nature of the electron-phonon interactions, the fitting procedure has to be modified slightly to determine the correct power law of the energy-loss rate in each sample. Defining not only the deformation potential $\Xi$ but also the exponent $\gamma$ as adjustable parameters in the fitting function, and thus changing the fixed $T^5$-dependence of eqn. (11.30) to a variable power law of the type $(T'_e - T'_L)$, yields similar values for $\Xi$ but may produce a slightly different power law. Such a modified approach for the fitting procedure is shown in Figure 11.4 where the diverging values of $\gamma$ are represented by different curvatures of the fitted curves.

![Graphs showing exponential fit of the energy-loss rate $P_e$](image)

**Figure 11.4** Exponential fit of the energy-loss rate $P_e$ where both the deformation potential $\Xi$ and the exponent $\gamma$ are fitting parameters to establish the correct $T^2$-dependence of the energy-loss rate for each sample.

The results for the exponent $\gamma$ and the deformation potential $\Xi$ are summarized in Table 11.4 together with the values of $\Xi$ as deduced from the gradient of the reciprocal mobilities in the equipartition regime. Any variation of the exponent from
its anticipated value of $\gamma = 5$ signifies a deviation of the phonon processes from the initially presumed case of unscreened deformation-potential scattering. The processes most likely to affect the nature of the electron-phonon interactions are screening effects or piezoelectric contributions. In general, it is problematic to include the effects of screening theoretically since the screening efficiency is reduced strongly for such short-range coupling. Furthermore, most theoretical models assume static screening, whereas screening of acoustic phonons should be treated as dynamic at high carrier densities and low temperatures [279]. However, dynamic screening requires the application of certain corrections for the degeneracy and is thus difficult to account for accurately. The inclusion of static screening consequently tends to overestimate the deformation potential. Apart from claiming that screening might be reduced by other scattering mechanisms in any case, the conventional procedures used to account for screening of a deformation potential have been questioned in general due to the difference between a deformation potential arising from a strain and a potential distribution resulting from an external electric field [249].

<table>
<thead>
<tr>
<th></th>
<th>$T_L$ (K)</th>
<th>exponent $\gamma$</th>
<th>$\Xi_{ELR}$ (eV)</th>
<th>$\Xi_{\sigma^{-1}}$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF197</td>
<td>1.7</td>
<td>4.7</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>BF213 (a)</td>
<td>0.1</td>
<td>4.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>BF213 (b)</td>
<td>0.05</td>
<td>4.9</td>
<td>3.2</td>
<td>3.4</td>
</tr>
<tr>
<td>BF313</td>
<td>0.05</td>
<td>4.8</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>BF314</td>
<td>0.05</td>
<td>4.8</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>BF741</td>
<td>0.4</td>
<td>4.8</td>
<td>4.0</td>
<td>4.1</td>
</tr>
<tr>
<td>BF851</td>
<td>0.4</td>
<td>4.8</td>
<td>3.8</td>
<td>4.5</td>
</tr>
<tr>
<td>BF852</td>
<td>0.4</td>
<td>4.7</td>
<td>5.9</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Table 11.4 Fitted values of the exponent $\gamma$ and the deformation potential $\Xi$ for n-type samples from the energy-loss rate of eqn. (11.30). For comparison, the calculated values of $\Xi$ according to eqn. (11.10) from the gradient of the reciprocal mobilities are also listed.

The fact that $\gamma < 5$ in all samples suggests a small contribution of electron-phonon interactions via piezoelectric coupling to the energy relaxation process. Concerning SiGe systems, similar observations have so far only been reported for a p-type structure [262] where the origin of the piezoelectric coupling has been attributed to electronic polarization of the slightly ionic layers (the internal displacements of the electronic charges relative to the cores) and charge transfer effects (strain-induced change in ionicity). Despite the fact that piezoelectric scattering is expected to become more significant as the lattice temperature decreases, no corresponding trend regarding the value of $\gamma$ is noticeable in the present samples for which $T_L$ varies between 1.7 K and 0.05 K. An equivalent dependence of $\gamma$ on strain is also not
obvious as the samples of the second batch have the lowest strain and the related exponents should thus deviate the least from the theoretical value of five. Within the first set of samples, however, time-related relaxation in samples BF213 (b), BF313 and BF314 reduces the strain and subsequently the contribution of piezoelectric scattering. Variations of the screening potential may also influence the value of $\gamma$ which renders any further analysis of the exponent ambiguous. A thorough investigation into the role of piezoelectric scattering in the present 2DEG samples could not be carried out within this framework due to time limitations.

For the samples of the first batch and also for sample BF741 of the second batch, the values of $\Xi$ as deduced from the fit of the energy-loss rate and those calculated from the gradient of the reciprocal mobilities are in excellent agreement, which clearly demonstrates the validity of both methods in evaluating the deformation potential. However, the analysis of the energy-loss rate in the BG regime is expected to allow a more accurate assessment of $\Xi$ than the measurement of the $T$-dependence of the mobility in the EP regime since the electron heating process is a purely phonon-related phenomenon and is therefore not affected by Coulomb scattering. Only samples BF851 and BF852 display a certain discrepancy when comparing the respective values of $\Xi$, but even then, the variance remains below 15% in both cases.

The assumption made earlier that the deformation potential decreases with increasing strain is further corroborated by the slight increase in $\Xi$ for BF213 over time due to a partial relaxation of the virtual substrate and thus a small reduction in strain. Samples BF197 and BF213 (a) display almost identical values for the deformation potential despite having virtual substrates that are the most dissimilar within the first set of samples. As both structures were measured relatively shortly after they were grown, this suggests that the variations in $\Xi$ in the aged samples of BF213 (b), BF313 and BF314 should be attributed to different degrees of relaxation, and that directly after growth, BF313 and BF314 had an initial deformation potential similar to that of BF197 and BF213 (a). In other words, the quality of the virtual substrate seems to be identical in all samples of the first set directly after growth, but the subsequent relaxation develops at different rates in each sample. The virtual substrates in the samples of the second batch, however, appear to be of lower quality to begin with since the corresponding measurements were performed after a shorter time period than those for the aged structures of the first batch. Alternatively, it is also possible that the relaxation rate of the buffer layer in the samples of the second set is much greater over time, and hence, the strain is reduced a lot further in a shorter time period.

So far, only one other experiment has been reported in which the energy-loss rate of a Si$_{0.7}$Ge$_{0.3}$ 2DEG structure has been analyzed [164, 252]. As in this case, the theoretical description deals only with deformation-potential coupling, but it also includes a static screening factor to account for the effect of free-electron screening of the scattering potential [99, 266, 275-277, 280]. The experimental data of $P$, coincides satisfactorily with the calculated energy-loss data when employing the screening factor and a deformation potential of 9 eV. This relatively high value of $\Xi$ is surprising as it represents a deformation potential within the margins of bulk Si and therefore suggests no change in $\Xi$ for a lattice under considerable strain. A possible
explanation could be a minimal strain in the Si channel of their first two samples due to a very thin buffer layer of low quality, which is corroborated by very low mobilities and a relatively small $\tau_e/\tau_q$ ratio. However, a third sample with a much thicker buffer layer and accordingly with a higher mobility and larger $\tau_e/\tau_q$ ratio displays the same energy loss as the first two samples. Assuming a virtual substrate of higher quality in the third sample and thus a larger strain, this would indeed suggest the deformation potential to be independent of strain, which is in contrast to the analysis of the energy-loss rate in the present samples.

Similarly controversial values of the deformation potential have been reported in GaAs/Al$_x$Ga$_{1-x}$As heterostructures [281, 282], confirming the problem of analysing the energy-loss rate which, among other things, might partially be due to the simplifying approximations made in the theoretic models concerning the momentum conversation and the phonon distribution [278]. Whereas in some experiments, $\Sigma$ has been found not to vary from its accepted bulk value of 7 - 8 eV [267, 271, 272, 283], the deformation potential has also been shown to be much larger than that in other experiments with values ranging from 10 to 16 eV [246, 266, 269, 270]. As a possible explanation for the strong discrepancy between the values of the deformation potential, it has been considered that the phonon modes may be disturbed by the proximity of the surface of the crystal, which would render the energy-loss rate sensitive to the distance between the electron gas and the surface. It has also been observed that repeated thermal cycling of a sample between room temperature and liquid-helium temperatures results in an increase in the energy-loss rate, although the exact origin of this phenomenon is not quite clear yet. The most plausible explanation for the disparity, however, might be the decision whether or not to include screening in the theoretical model to calculate the energy-loss rate. When comparing both the analysis and the results of the experiments, it is evident that $\Sigma$ does not change from its bulk value when unscreened potentials are employed whereas the application of screened potentials causes the value of $\Sigma$ to increase. The choice whether to use a screened or an unscreened potential might not always be apparent from the $T$-dependence of the experimental data as it has been found in some cases to remain unchanged, although the absolute value of the energy-loss rate is reduced by screening [269]. In addition, the energy-loss rate has also been shown to be dependent on the well width [272, 284].

Displaying a $T$-dependence reciprocal to that of the energy-loss rate, the momentum relaxation time $\tau_m$ as calculated from eqn. (11.31) is plotted in Figure 11.5 against $T_e - T_L$ on a double logarithmic scale. As in the case of the energy-loss rate, $\tau_m$ is not merely a function of $T_e$ when displayed in this way but inevitably is also dependent on the lattice temperature which determines the lowest possible value of the electron temperature. Furthermore, when the electrons are heated above $T_L$, the corresponding change in the phonon scattering rate is smaller the higher the lattice temperature. The variations of $\tau_m$ in the given temperature range are therefore much smaller at $T_L = 1.7$ K (less than two decades) than at $T_L = 0.05$ K (around seven decades). Mathematically, these deviations of $\tau_m$ from the ideal straight line with a gradient of five are caused by the inclusion of the $T_L^5$-term in the analytic expression, which results in the observed flattening of the curves with increasing lattice
temperature. The curves of \( \tau_m \) for different samples that have the same value of \( T_l \) are thus approximately congruent due to identical electron effective masses and material constants in all samples. The only parameters that might vary to a certain extent are the deformation potential and the Fermi wavevector. Accordingly, the calculated curves of \( \tau_m \) can be seen to differ slightly for the samples of the second batch despite having identical lattice temperatures. Samples BF313 and BF314 of the first batch, on the other hand, display exactly the same momentum relaxation time due to very similar values of \( \xi \) and \( k_F \), whereas \( \tau_m \) of BF213 (b) is marginally smaller.

![Graph showing momentum relaxation time \( \tau_m \) as a function of \( T_C - T_L \)](image)

**Figure 11.5** Momentum relaxation time \( \tau_m \) as calculated from eqn. (11.31) as a function of \( T_C - T_L \). Due to identical effective masses and material constants, \( \tau_m \) is mainly governed by the lattice temperature at low values of \( T_C - T_L \) but \( \tau_m \) is very similar in all samples at higher values of \( T_C - T_L \). In the BG regime, \( \tau_m \) is generally much larger than \( \tau_r \).

As the most prominent feature of Figure 11.5, \( \tau_m \) can be seen to be extremely large for low values of \( T_l \) and \( T_C - T_L \), but as \( \tau_m \) decreases with increasing \( T_C \), it becomes comparable to the transport relaxation time \( \tau_r \) for electron temperatures around the transition temperature between the BG and the EP regime. The very large values of \( \tau_m \) at low \( T_C \) can be explained by the Pauli exclusion which restricts the allowed phonon processes more and more as the electron temperature continuous to decrease. With fewer electron-phonon interactions occurring, it takes longer to randomize the net momentum of the electrons. As a result of the very long momentum relaxation times in the BG regime, the mobility limited purely by acoustic-phonon scattering \( \mu_{ac} = \tau_m e/m^* \) is considerably larger than the measured
mobility $\mu = \tau_c c/m^*$ which is limited predominantly by Coulomb scattering. At $T_e = 1$ K, $\mu_{ac}$ is larger than $\mu$ on average by roughly six orders of magnitude, and $\mu_{ac}$ is also larger than $\mu_{\text{vac}}$ (as defined in chapter 9) by about five orders of magnitude. These differences between the mobilities increase to approximately eleven and nine orders of magnitude, respectively, when the electron temperature decreases to 0.1 K. The discrepancies between $\mu_{ac}$ and $\mu_{\text{vac}}$ confirm the assumption made earlier that for structures with medium to high mobilities, the momentum relaxation time due to phonon processes are best determined from the energy-loss rate and not by analysing the electron mobility which might include additional temperature-dependent contributions from Coulomb scattering.

Calculated via eqn. (11.32), the energy relaxation time $\tau_e$ is shown in Figure 11.6 on a double logarithmic scale as a function of $T_e-T_l$. Apart from this dependence on the electron temperature, $\tau_e$ is also governed by the energy-loss rate and the Fermi energy which might differ from sample to sample. In $n$-type samples, $E_f$ is only a function of $n_e$ which varies at most by a factor of two in both sets of samples. However, $P_e$ is dependent on the deformation potential, the sheet density, the well width and the lattice temperature, rendering $\tau_e$ ultimately to be a complex function of altogether five parameters, which makes it difficult to compare $\tau_e$ between different samples with regard to possible variations of the material constants.

![Figure 11.6](image.png)

*Figure 11.6 Energy relaxation time $\tau_e$ as calculated from eqn. (11.32) vs $T_e-T_l$. Despite variations in $P_e$ and $E_f$, $\tau_e$ is predominantly determined by $T_l$ at low values of $T_e-T_l$. At higher electron temperatures, however, $\tau_e$ is similar in all samples. In the BG regime, $\tau_e$ is larger than $\tau_i$ but smaller than $\tau_m$ in each sample.*
Having identical lattice temperatures and very similar deformation potentials, the curves of $\tau_e$ for samples BF213 (b), BF313 and BF314 are absolutely congruent despite small variations in $P_\tau$ and $E_\tau$. In the samples of the second batch, which also have matching yet slightly higher values of $T_L$ and different deformation potentials, the curves of $\tau_e$ diverge considerably, although the deviations in $P_\tau$ and $E_\tau$ are comparable to those in the three samples of the first set. As in the case of the momentum relaxation time, the energy relaxation time increases with decreasing lattice temperature at low values of $T_L-T_P$. This scenario can be explained by bearing in mind that the principle rates of energy relaxation are associated with the absorption and emission of acoustic phonons. Quasielastic electron-phonon scattering is therefore the only mechanism that removes energy from the hot electrons. Electron-electron scattering, on the other hand, merely redistributes the energy and establishes an average electron temperature in the process. With decreasing lattice temperature, the restrictions on the phonon processes due to Pauli exclusion allow for fewer electron-phonon interactions, which thus implies a reduced energy exchange. This means that acoustic phonons become less and less efficient at cooling the hot electrons the lower the lattice temperature. However, the $T_L$-dependence of $\tau_e$ decreases with increasing $T_P$ as $T_L^5$ becomes insignificant compared to $T_P^5$, and the energy relaxation times are similar in all samples at higher electron temperatures. At low electron temperatures, $\tau_e$ is generally larger than $\tau_i$ but smaller than $\tau_m$, which indicates that the process of energy relaxation is faster than the momentum relaxation associated with acoustic-phonon scattering but significantly slower than the process of momentum relaxation by Coulomb scattering. For the three samples at $T_L = 0.05$ K, $\tau_e$ is larger than $\tau_i$ by as much as eight orders of magnitude for $T_r = T_L$, but this disparity declines with increasing lattice temperature. At $T_L = 0.1$ K, $\tau_e$ and $\tau_i$ differ by a maximum of seven orders of magnitude, which drops even further to a value of six orders of magnitude for the three samples at $T_L = 0.4$ K. The lowest discrepancy between $\tau_e$ and $\tau_i$ with only three orders of magnitude can be seen at $T_L = 1.7$ K in sample BF197. The values of $\tau_e$ in this sample are also comparable to those reported for a similar sample at a lattice temperature of 1.55 K [164, 252]. For the three samples of the second batch, however, the values of $\tau_e$ are about an order of magnitude larger than found for equivalent samples at the same lattice temperature of 0.4 K [175].

11.2.2 ENERGY-LOSS RATE OF HOT HOLES

The electric field experiments of all $p$-type samples were carried out exclusively in the 15 T dilution fridge at a base temperature of 0.03 K, apart from BF218 (a) for which a slightly higher base temperature of 0.06 K was recorded. Concerning the highest possible hole temperature that still allows an adequate analysis of the energy-loss rate in each sample, the same considerations are applicable as in the corresponding case of the maximum electron temperatures in the $n$-type samples. In Figure 11.7, the input power per hole $P_b$ is plotted as a function of $T_r-T_L$ on a logarithmic scale (left) and also on a double logarithmic scale (right). Again, the input power has been calculated from eqns (11.19) and (11.20) to verify the accuracy of $P_b$ for which only small deviations of the individual values can be detected.

254
between the two methods. Subsequently, the deformation potential of each sample can be extracted by fitting the analytic expression of eqn. (11.30) to the experimental energy-loss data where \( \Xi \) is the only adjustable parameter.

As expected, the values of the input power in Figure 11.7 lie within the same range as those of the first set of \( n \)-type samples due to very similar base temperatures. With regard to the nature of the phonon processes, however, the most interesting aspect of Figure 11.7 is the excellent match between the experimental data and the fitted curves without having to adjust the exponent \( \gamma \) in the \( T \)-dependence of the analytic term. This remarkable agreement of the temperature dependence with theory can be demonstrated visually by plotting \( P_h \) against \( T_h^5 - T_L^5 \), which yields a straight line with a gradient of unity. Alas, the individual values of \( P_h \) cannot be resolved properly in a linear plot because they extend over a range of several decades in each sample. Alternatively, it is also possible to plot the logarithm of \( P_h \) against the logarithm of \( T_h - T_L \) to illustrate the nearly perfect \( T^5 \)-dependence of the energy-loss rate. At higher hole temperatures, the values of \( P_h \) lie on a straight line with a gradient of five, but at lower values of \( T_h \) the experimental data deviates from this straight line due to its proportionality to \( T_h^5 - T_L^5 \) rather than \( (T_h - T_L)^5 \).

Figure 11.7 Energy-loss rate per hole \( P_h \) as a function of \( T_h - T_L \). The lattice temperature is 0.03 K in all \( p \)-type samples with the exception of BF218 (a) for which \( T_L = 0.06 \) K. To determine the deformation potential \( \Xi \), the data points are fitted by eqn. (11.30) on the left and shown on a double logarithmic scale on the right to reveal the slight deviation of the data points from the straight line of a simple \( T^5 \) curve for hole temperatures close to \( T_L \).
As in the case of the $n$-type samples, the exact $T$-dependence of the energy-loss rate has been established by defining the exponent $\gamma$ as the second adjustable parameter in the fitting function. The resulting values for the exponent and both deformation potentials, deduced from the energy-loss rate as well as the reciprocal mobilities, are listed in Table 11.5. The fact that the experimentally determined values of the exponent are in excellent agreement with the anticipated theoretical value of $\gamma = 5$ indicates the deformation potential to be unscreened and implies furthermore that there are no contributions from piezoelectric scattering. Theoretically, it could be argued that small piezoelectric contributions together with the existence of moderate screening might, by coincidence, exactly neutralize their diametrically opposed effects on the exponent. Yet, the consistence of $\gamma$ in the present samples renders the conditions of this argument improbable as the sheet density and thus the degree of screening varies from sample to sample, and corresponding changes in the piezoelectric contribution that compensate precisely for these variations are highly unlikely. However, it has been suggested that the value of $\gamma = 5$ for the $T$-dependence of the energy-loss rate via unscreened deformation potentials does not change for the case when the deformation potential is weakly screened [262, 263]. While the notion of a piezoelectric contribution to the acoustic-phonon scattering is not supported by the analysis of the experimental data for the present samples, a notable contribution from piezoelectric coupling has been identified in other Si/Si$_{0.8}$Ge$_{0.2}$ heterostructures by power-loss experiments at a much higher lattice temperature of 1.75 K where the piezoelectric contribution should be less significant [262]. In concurrence with the present results, however, no contribution from piezoelectric coupling could be found in similar Si/Si$_{0.8}$Ge$_{0.2}$ samples from measurements of the energy relaxation of hot holes at a lattice temperature of around 0.35 K [261, 263].

<table>
<thead>
<tr>
<th></th>
<th>$T_L$ (K)</th>
<th>exponent $\gamma$</th>
<th>$\Xi_{ELR}$ (eV)</th>
<th>$\Xi_{e^\perp}$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF216</td>
<td>0.03</td>
<td>5.0</td>
<td>5.2±0.7</td>
<td>5.4±0.7</td>
</tr>
<tr>
<td>BF217</td>
<td>0.03</td>
<td>4.9</td>
<td>5.7±1.2</td>
<td>5.6±1.2</td>
</tr>
<tr>
<td>BF218 (a)</td>
<td>0.06</td>
<td>5.0</td>
<td>1.3±0.1</td>
<td>–</td>
</tr>
<tr>
<td>BF218 (b)</td>
<td>0.03</td>
<td>5.0</td>
<td>4.7±0.5</td>
<td>5.0±0.5</td>
</tr>
<tr>
<td>BF638</td>
<td>0.03</td>
<td>5.0</td>
<td>1.1±0.5</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 11.5  Fitted values of the exponent $\gamma$ and the deformation potential $\Xi$ for $p$-type samples from the energy-loss rate in eqn. (11.30) and, for comparison, the calculated values of $\Xi$ from the gradient of the reciprocal mobilities in eqn. (11.10).

As a major aspect of Table 11.5, it is obvious that the values of $\Xi$ derived from the energy-loss rate in the BG regime are in excellent agreement with those obtained from the reciprocal mobility in the EP regime, which not only verifies the accuracy of the previous results but also confirms the suitability of evaluating the deformation potential in the EP regime. As the only difference between both methods, the values
of $\Xi_{ELR}$ can be seen to lie further apart, suggesting a slightly stronger variation of the deformation potential with strain. Accounting for the Ge composition and strain, this behaviour of the deformation potential satisfies the conditions of the proposed postulate better than the values of $\Xi$ that were deduced in the EP regime since the experimental data is spaced slightly wider than the interpolated values. Interestingly, the values of $\Xi$ for samples BF218 (a) and BF638, which were not measured in the EP regime, are lower than those for the other samples by at least a factor of three. In analogy to the variations of $\Xi$ in the $n$-type structures, this huge discrepancy might be explained by the age of and thus the degree of relaxation in the individual samples. Having been measured relatively shortly after they were grown, samples BF218 (a) and BF638 are expected to experience the highest strain of all samples due to their respective Ge compositions. In contrast to this, samples BF216, BF217 and BF218 (b) were measured approximately 18 months after their growth, allowing for the relaxation process to reduce the strain and thus increase the deformation potential noticeably. Furthermore, the relaxation process is likely to reduce the strain more efficiently the higher the Ge content in the sample, leading to a similar level of stress in the aged samples, which would explain the almost identical spacing of the experimental data and the interpolated values. However, it is unclear whether the strain relaxation alone is sufficient to explain such a large change in $\Xi$ over time as the relaxation in the heterostructures would have to be substantial in order to attribute the differences in $\Xi$ exclusively to the reduction of the strain. The notion of a decreasing effective mass with increasing strain might contribute to the problem of fully explaining the extremely low values of $\Xi$ because it reduces the scattering probability and thus the energy-loss rate even further. Assuming comparable magnitudes of strain as in the $n$-type samples of BF197 and 213 (a) due to similar Ge concentrations and age, the values of $\Xi$ should be comparable in all four samples (BF197, BF213 (a), BF218 (a) and BF638). Yet, the deformation potential in both $p$-type samples is still considerably lower by at least 1.5 eV. Accounting for the variations in the material constants and different effective masses, this discrepancy of $\Xi$ between the $n$-type and $p$-type samples might indicate a different dependence of the deformation potential on strain in structures under tensile strain (2DFG) compared to compressively strained structures (2DHG). The lowest deformation potential reported so far for a $p$-type structure has a value of 3 eV [262, 263] which is comparable to the values of $\Xi$ in the first set of $n$-type samples. The slightly higher values of $\Xi$ in the present aged $p$-type samples, on the other hand, lie within the range of reported values between 4.5 eV [260, 261] and 5.5 eV [259].

Finally, it is possible to analyze the momentum and energy relaxation times which are shown in Figure 11.8 as a function of $T_H - T_L$ on a double logarithmic scale. There are two major differences that are obvious when comparing the progression of $\tau_m$ in the $p$-type samples to the progression of $\tau_m$ in the $n$-type samples. The most noticeable distinction is the rather large spread of $\tau_m$ within the different 2DHG structures which covers more than two decades and does not diminish at higher values of the hole temperature as it is not caused by discrepancies in the lattice temperature but by different values of the hole effective mass and the material constants and also because of stronger deviations in the deformation potential than in the 2DEG samples. The other interesting feature is the very low phonon scattering...
rate in some of the \( p \)-type structures. Displaying the highest Ge concentrations, the lowest hole effective masses and the smallest deformation potentials of the present \( p \)-type samples, BF218 (a) and BF638 exhibit values of \( \tau_m \) that are up to two orders of magnitude higher than those found in the \( n \)-type samples at a comparable lattice temperature of \( T_L = 0.05 \) K. However, the momentum relaxation times in the three aged \( p \)-type samples are of the same order as those in the \( n \)-type samples. Despite much lower experimental mobilities in the 2DHG structures, the discrepancy between \( \mu_m \) and \( \mu \) in BF216, BF217 and BF218 (b) is approximately equal to that found in the \( n \)-type samples due to a larger hole effective mass. In BF638, on the other hand, the long momentum relaxation time and the small hole effective mass causes this divergence to increase by three orders of magnitude.

![Graphs showing momentum and energy relaxation times](image)

**Figure 11.8** Momentum relaxation time \( \tau_m \) vs \( T_h - T_L \) (left) and energy relaxation time \( \tau_e \) vs \( T_h - T_L \) (right) as calculated from eqns (11.31) and (11.32), respectively. Due to different values for the effective mass and material constants, both relaxation times vary considerably for the different samples despite a constant lattice temperature.

A similar spread as for \( \tau_m \) can be seen for the various curves of the energy relaxation time for exactly the same reasons, i.e. different values of \( m^* \), \( \Xi \) and the material constants in each sample. The strong influence of the deformation potential on \( \tau_e \) is clearly visible as the curves of the three aged structures are relatively closely spaced due to fairly similar values of \( \Xi \), whereas the very low deformation potentials of BF218 (a) and BF638 cause \( \tau_e \) in those samples to be longer by at least an order of magnitude. Despite a marginally lower lattice temperature of \( T_L = 0.03 \) K, the energy
relaxation times of the three aged p-type structures are a little shorter than those in the n-type samples which are at $T_1 = 0.05$ K due to a slightly elevated energy-loss rate, mainly caused by higher values of the hole effective mass and the deformation potential. When comparing both relaxation times $\tau_m$ and $\tau_v$ it can be concluded that the quasielastic scattering process associated with acoustic phonons is much more efficient at relaxing the energy of the holes than randomising their net momentum since $\tau_m \gg \tau_v$ in all samples. For $T_h = T_L$, however, the difference between $\tau_i$ and $\tau_v$ is even larger and ranges from eight orders of magnitude for BF217, which is identical to the corresponding disparity in the n-type samples at $T_L = 0.05$ K, to just under eleven orders of magnitude for BF638.

From the calculation of the deformation potential and the momentum and energy relaxation times, it can be concluded that a decreasing deformation potential in conjunction with a correlated reduction in the hole effective mass causes a significant decline in the phonon scattering rate. In contrast to the present type of 2DEG structures which all have similar virtual substrates, it is therefore generally possible to optimize a 2DHG structure with respect to its phonon scattering rate, irrespective of the Coulomb scattering rate. However, any conclusions from the current analysis of the results in the BG regime for the p-type samples (and also to some extent for the n-type samples) have to be drawn cautiously and with certain reservations since the deformation potential and both relaxation times were calculated using the analytic expressions of eqns (11.30) - (11.32) which are strictly valid only for parabolic spherical bands.

### 11.3 Summary

Phonon processes are best described by the deformation potential as it is related to the shift of the band edge energy associated with the local strain due to lattice vibrations. The deformation potential can be calculated either from the gradient of the temperature-dependent reciprocal mobilities in the equipartition regime or from the energy-loss rate in the Bloch-Grüneisen regime. An excellent agreement between the values of the deformation potential from both methods can be seen in the present n- and p-type samples, even though the energy-loss rate is linked exclusively to acoustic phonon scattering while the reciprocal mobility also incorporates contributions from Coulomb scattering. In the equipartition regime, however, Coulomb scattering is largely independent of temperature and hence does not affect the slope of the temperature-dependent reciprocal mobilities. Compared with their corresponding relaxed bulk values, the deformation potentials of the strained n-type as well as p-type structures are reduced considerably due to the presence of uniaxial stress. Since the value of the deformation potential is dependent on the strain in the quantum well, the quality of the virtual substrate in the 2DEG samples has a direct influence on how much the value of $\Xi$ is reduced. Accordingly, the deformation potential in the high-quality n-type samples of the first set displays values in the
range of 3.4 - 3.8 eV while the deformation potential in the 2DEG samples of the second set is slightly larger, reaching values as high as 5.5 eV. This discrepancy can be explained by the weaker stress in the samples of the second set due to the lower quality of their virtual substrates. Low values of the deformation potential are beneficial to the electronic transport at room temperature where acoustic and optical phonon scattering dominate as they indicate less interaction effects between the carriers and the lattice vibrations. In accordance with the results of the previous chapters, the relaxation of the channel in one of the \( n \)-type samples due to time-dependent structural degradation is thought to be marginal since the value of its deformation potential increases only slightly over time. However, the ageing process seems to have a stronger effect on \( E \) in the 2DHG structure which might suggest a different dependence of the deformation potential on strain in structures under tensile and compressive strain. Exhibiting values of around 1 eV directly after growth, the deformation potential increases significantly to 4.7 - 5.7 eV over time due to a partial relaxation of the channel. Consistent with the variation of the strain in dependence of the Ge composition, the value of \( E \) falls with increasing Ge content. While the initial drop of the deformation potential from its corresponding bulk value is comparable to that in the 2DEG structures, its subsequent increase over time results in a reduced hole mobility at room temperature. Producing only small fitting errors, the energy relaxation of the hot electrons and hot holes are best fitted by the analytical expression for unscreened deformation-potential coupling, suggesting no or only weak screening and no contributions from piezoelectric coupling in the present samples.
12 Conclusions

Magnetotransport measurements have been performed at high magnetic fields up to 15 T and low temperatures down to 30 mK on a variety of high-quality $n$- and $p$-type Si/ Si$_{1-x}$Ge$_x$ heterostructures grown by GSMBE in order to study the transport properties of the resulting two-dimensional electron and hole gas systems. Apart from providing valuable information about the electronic transport, i.e. the carrier mobility and the sheet density, which is vital for device applications as it allows the structures to be characterized electrically, the analysis of the low-temperature magnetotransport phenomena also delivers detailed information about other transport properties, such as the effective mass and the dominant scattering processes, which directly influence the electronic transport in these structures. One of the most important factors to determine the transport properties in general is the quality of the grown structure, especially the quality of the interfaces of the quantum well and the quality of the virtual substrate in the case of $n$-type heterostructures. By determining the transport properties experimentally, the characteristic of individual layer segments of the structure can be established separately, and, together with a structural characterization from techniques such as TEM, XRD or SIMS, it provides a crucial feedback to the growth process in order to help improve the quality of the structures grown by the GSMBE technique. However, the other dominant factor that governs the transport properties is the structural layer design which can be modified to tailor the band structure according to specific requirements. Hence, a band structure model was employed to aid the process of optimizing the individual transport properties, regardless of the quality of the structures. With respect to their suitability in the fabrication of $n$- and $p$-channel MODFET devices, the structural integrity of the Si/SiGe heterostructures has also been investigated over a time period of 18 months to establish whether the quality of the sample and the strain in the quantum well deteriorate with time and thus change the characteristic of the transport properties.

Indicating a 2DEG of high quality in the quantum well of the $n$-type structures, the well-pronounced SdH oscillations in the longitudinal resistivity exhibit dissipationless minima at filling factors corresponding to the main Landau level states as well as the spin-split and valley-split states. Coinciding with the minima of the SdH oscillations, the quantized Hall plateaux in the transverse resistivity are clearly resolved and display the expected value of $h/\nu e^2$. The SdH oscillations of the $p$-type samples, however, are dominated by minima associated with odd filling factors, which is a well-established property of $p$-type SiGe systems in which the effective Landé factor is sufficiently large to cause spin splitting to be comparable to the cyclotron energy, thus forcing a spin degeneracy of up and down spin states of successive Landau levels at even filling factor. One of the $p$-type samples experiences a phase transition to the Hall insulating phase at filling factors $\nu \leq 2$. In another $p$-type sample, the Hall plateaux can be seen to occur in antiphase to the SdH oscillations at lower magnetic fields. Excellent agreement between the sheet densities
as deduced from the fundamental field of the SdH oscillations and form the slope of the Hall resistivity clearly demonstrates the absence of any parallel conduction in the present n- and p-type samples at temperatures below 4.2 K.

From the temperature dependence of the SdH oscillations, the electron and hole effective masses were extracted for the n- and p-type samples, respectively. As expected, the electron effective mass was calculated to be 0.19±0.02 \( m_e \) in all 2DEG structures, irrespective of the structural layer design and largely independent of the magnetic field. The hole effective mass, on the other hand, is affected by uniaxial strain, by the Ge composition in the quantum well and also by the sheet density, which is mainly due to the nonparabolicity of the valence bands. Furthermore, the nonparabolicity effects cause the hole effective mass to increase with magnetic field. Accordingly, the hole effective mass at zero field had to be extrapolated for each sample and the corresponding values have been ascertained to vary between 0.33±0.09 \( m_e \) for a Ge concentration of 10% and 0.27±0.04 \( m_e \) for a 30% Ge content in the channel. However, the lowest effective mass of the present p-type samples could be observed in the step-graded structure, for which \( m^*(B=0) \) has a value of 0.17±0.09 \( m_e \). This drop in \( m^* \) is thought to be caused by an increased splitting of the HH1 and LH1 subbands, effectively reducing the nonparabolicity effects in the HH1 band.

Representing the electronic transport properties, the carrier mobility and sheet density are the most important parameters to characterize an electronic system concerning possible device applications with lateral transport. Since the electron effective mass is identical in all n-type structures, possible variations in the mobility can only arise from changes in the sheet density or the scattering processes. High electron mobilities in the range of \( 1 - 1.7 \times 10^5 \) cm\(^2\)/Vs have been found in the low-density samples, suggesting the structure of those samples to be of high quality. Considerably lower mobilities between 5 and \( 6 \times 10^4 \) cm\(^2\)/Vs were observed in the high-density delta-doped structures, whereby the drop in mobility can be ascribed to the higher sheet density rather than to a lower quality of the structures. However, a very low structural quality is responsible for even lower mobilities in the second set of n-type samples, for which the sheet density varies greatly. In p-type structures, the dependence of the hole effective mass on the Ge composition has to be taken into account as it affects the mobility in the opposite way than the dependence of the interface quality on the Ge content. The highest hole mobility of \( 7 \times 10^3 \) cm\(^2\)/Vs was found in the structure with the lowest Ge concentration and also with the lowest sheet density. It should be noted that the Ge content and the sheet density are closely related as the confinement potential decreases with falling Ge composition. Accordingly, the structure with the highest Ge content displayed the highest sheet density, and although the effective mass decreases, the sample has also the lowest mobility of only \( 2 \times 10^3 \) cm\(^2\)/Vs. Despite having the lowest hole effective mass of all p-type samples, the step-graded structure displays the lowest hole mobility of just under \( 10^3 \) cm\(^2\)/Vs because its wavefunction extends over both interfaces within the quantum well. All n- and p-type structures exhibit an almost linear temperature dependence below 4.2 K, characteristic for structures with a relatively strong contribution from Coulomb scattering.
Information about the dominant scattering process can be obtained by assessing the ratio of transport to quantum relaxation time. Whereas the transport relaxation time can be derived directly from the low-field mobility, the quantum relaxation time has to be extracted from the magnetic field dependence of the SdH oscillations by evaluating the corresponding Dingle plot. The high-quality n-type samples of the first set display $\tau_r/\tau_q$ ratios of 10 and 11, which indicates scattering by remote ionized impurities to be dominant in these structures. In GaAs/AlGaAs heterostructures, the ratio of $\tau_r/\tau_q$ is usually larger than 20, suggesting still a relatively strong contribution from short-range scatterers in the present n-type samples from the interface and the virtual substrate. In one structure, however, the quality of the interface and the virtual substrate appears to be superior compared to the rest of the samples because the ratio of $\tau_r/\tau_q$ reaches a much higher value of almost 20. Concerning the structural integrity of the samples over time, the $\tau_r/\tau_q$ ratio of both delta-doped samples remains constant within the given time period, suggesting the quality of these structures not to deteriorate with time. This is further corroborated by the invariance of the sheet density and the electron mobility in both samples as measured at the two different dates. The uniformly doped sample, on the other hand, experiences a slight drop in the sheet density and the mobility, implying a structural degradation in this sample over time. However, since the ratio of $\tau_r/\tau_q$ increases at the same time, this is thought to be caused by an increase of the surface potential due to changes in the oxide rather than by the deterioration of the interface quality or the virtual substrate. Hence, the delta-doping technique appears to provide an effective screening potential that compensates for the added ionized impurities in the extended depletion region of the surface potential. The low-quality n-type samples of the second set display $\tau_r/\tau_q$ ratios roughly between 4 and 7, indicating even stronger contributions from scattering by threading dislocations.

Exhibiting $\tau_r/\tau_q$ ratios of approximately one, the single-step p-type structures are considered to be dominated by interface charge scattering. However, the contribution from interface roughness scattering is believed to increase with growing Ge concentration, which is reflected by a corresponding decrease in the ratio of $\tau_r/\tau_q$. A further increase of scattering by interface roughness and a correlated drop in the hole mobility can be seen in the aged sample with a 30% Ge content, which is attributed to a partial relaxation of the strained channel, thereby increasing the roughness of the upper interface. Experiencing the lowest value of $\tau_r/\tau_q$, the step-graded structure is dominated approximately equally by interface charge and interface roughness scattering. The coinciding significant drop in $\tau_r$ as well as in $\tau_q$ can be explained by the proximity of the upper interface due to a narrower quantum well and the presence of two additional interfaces in the boundary of the wavefunction.

In the n-type as well as in the p-type samples, the values of the deformation potential as calculated from the reciprocal mobilities in the equipartition regime agrees well with those deduced from the energy-loss rate in the Bloch-Grüneisen regime. This congruence was observed despite the fact that the energy-loss rate is exclusively associated with acoustic phonon scattering whereas the reciprocal mobility is also affected by Coulomb scattering. However, in the temperature range of interest,
Coulomb scattering is independent of temperature and hence does not influence the analysis of the gradient. All investigated samples, both n- and p-type structures, experience a strain-induced reduction of the deformation potential compared to the corresponding values in the bulk material. Depending on the strain in the channel and thus the quality of the virtual substrate, the deformation potential in the 2DEG samples decreases to values roughly between 3 and 4 eV. Describing the interaction effects between the carriers and the lattice vibrations, low values of the deformation potential are beneficial in reducing scattering by acoustic phonons which, together with optical phonon scattering, represents the dominant scattering process at room temperature. As the strain in the p-type samples varies strongly with the Ge content in the channel, the deformation potential is lowest in the structures with the highest Ge concentration. Interestingly, the ageing process appears to have a stronger effect on the deformation potential in p-type structures than in n-type structures, for which its value increases only marginally over time. Directly after growth, the deformation potential in the 2DHG samples is reduced to about 1 eV, whereby the drop from the bulk value to that of the strained structure is comparable to the corresponding drop in the n-type structures. After a partial relaxation of the strained channel over time, however, the deformation potential increases to 4.7 - 5.7 eV, depending on the Ge composition. Although these values are still below the interpolated bulk values, the associated increase in deformation-potential scattering has adverse effects on the hole mobility at higher temperatures. The energy relaxation of the hot electrons as well as the hot holes can be described theoretically to a good approximation by the analytical expression for unscreened deformation-potential coupling.

**Future Work**

In conclusion, the n-type structures that were grown initially have been found to be of high quality, demonstrating the suitability of the GSMBE growth technique to realize virtual substrates with a low density of threading dislocations and to produce high-quality interfaces. Although subsequent attempts to grow virtual substrates with a reduced total thickness failed to reproduce structures of the same quality, it should be possible to increase the quality of the thicker virtual substrate even further to limit short-range scattering from threading dislocations. At the same time, the quality of the interface has to be improved if realistic $\tau_e/\tau_i$ ratios between 20 and 30 are to be achieved, ultimately increasing the low-temperature mobility. However, for a corresponding improvement of the electron mobilities at room temperature, it is vital to minimized the parasitic parallel conduction in the supply layer, whereby further developments of the delta-doping technique promise to provide better control over the doping profile, thus helping to eliminate all parallel conduction while the required sheet density in the channel is maintained.

As for the hole mobility in the p-type structures, an improvement of the interface quality by reducing the Ge concentration has to be weighted against the correlated increase in the hole effective mass. In general, the benefit of a lower effective mass due to a higher Ge content is counterbalanced by an increase in the sheet density and
the scattering rate to produce structures with lower hole mobilities. Apart from improving the GSMBE growth process to restrict the amount of interface charges and/or reducing the roughness of the interface for quantum wells with a higher Ge concentration, compositional grading in the channel offers the possibility of realizing a low hole effective mass (high Ge content) and a better quality of the upper interface (low Ge content) in the same structure. Having already grown simple step-graded quantum well structures that also pull the wavefunction away from the still inferior upper interface, further research should be concentrated on producing quantum wells in which the Ge composition is graded linearly to avoid detrimental interfaces within the channel. As a vital tool to model the band structure and engineer the wavefunction, the application of the self-consistent Poisson-Schrödinger solver significantly facilitates the process of designing appropriate structures that experience reduced scattering, yet display a low effective mass. Similar to its effect in the \( n \)-type samples, delta-doping of the supply layer might help to suppress parallel conducting channels at room temperature and should therefore be examined, as well.

Apart from focusing on altering the structural design to optimize the transport properties, the fabrication of gates is of utmost importance not only for device applications but also for simple Hall bar structures as it allows the modulation of the sheet density, and the position of the wavefunction can be altered without modifications in the design of the structures. Being a function of \( n_s \), the low-temperature mobility can thus be optimized by modulating the sheet density, and at room temperature, the low-mobility parallel conduction channel in the supply layer can be suppressed. The ability to move the position of the wavefunction in a uniform quantum well is greatly beneficial when establishing the quality of the upper and lower interface in \( p \)-type and also to a certain extent in \( n \)-type structures. Furthermore, the position of the hole wavefunction at which scattering from both interfaces is at a minimum can be determined, thus helping to specify the grading profile, especially whether the maximum Ge concentration should be located somewhere in the middle of the channel or in the vicinity of the lower interface. In order to maintain the high quality of the structures, the fabrication of the gates requires a limited thermal budget to prevent strained layers from relaxing and the diffusion of Ge and the dopants. Operating within these limitations, plasma oxidation is a promising technique that produces oxide layers comparable in quality to thermal oxides but at substantially lower temperatures. Alternatively, it is also possible to produce Schottky contacts, which is a simple but effective method to realize high-quality gates. Due to their versatility of use and enormous benefit regarding the analysis techniques, the fabrication of gated Hall bar structures should introduce a host of additional tools to evaluate the quality of Si/SiGe \( n \)- and \( p \)-type heterostructures and ultimately optimize their transport properties.
References

References


