PHYSICS AND APPLICATIONS OF

$Si/Si_{1-x}Ge_x$ MODULATION DOPED STRUCTURES

V. Venkataraman

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ELECTRICAL ENGINEERING

JANUARY 1994

Abstract

We have investigated transport of electrons and holes at low temperature in $Si/Si_{1-x}Ge_x$ modulation doped heterostructures. We used the properties of the two-dimensional hole gas to probe the $Si/Si_{1-x}Ge_x$ interface and fabricated symmetric p-type double heterostructures for the first time by rapid thermal chemical vapor deposition. We have also fabricated two-dimensional electron gases on graded relaxed $Si_{1-x}Ge_x$ buffer layers with maximum low temperature mobility of 45,000 $cm^2/V{\cdot}s$, limited by the background doping in our reactor. We achieved the highest carrier density 2.8×10^{12} cm^{-2} at 10 K and the lowest sheet resistivities 69 $\Omega/\square$ at 10 K and 140 $\Omega/\square$ at 77 K. Using a two-mask self-aligned process, we fabricated modulation doped field effect transistors with low gate leakage currents at room temperature and maximum transconductance of 20 mS/mm limited by parasitic contact resistance.

We have compared electron and hole transport and demonstrated alloy scattering in $Si_{1-x}Ge_x$ strained channels to be the dominant scattering mechanism. From the experimental analysis, we established the empirical alloy scattering potentials of 0.8 eV for electrons and 0.6 eV for holes for the first time in two-dimensional strained $Si_{1-x}Ge_x$ layers. We propose strong intervalley coupling through alloy scattering and develop quantitative models for photoluminescence no-phonon line intensities and valley splitting energies in high magnetic fields.

Acknowledgments

I am indebted to my adviser, Prof. Jim Sturm, for his guidance and support during my stay at Princeton. His boundless energy and indefatigable enthusiasm for experimenting in the laboratory never cease to amaze me. Without his encouragement and advice, my efforts would have surely been in vain. I also thank him for giving me the opportunity to disseminate this work through conferences and journals

I enjoyed working with the other students in my group, Peter Garone, Peter Schwartz, Erwin Prinz, Xiadong Xiao, Chee-wee Liu, Zeljka Matutinovic-Krstelj, Anthony St.Amour, Louis Lanzerotti, Arvind Reddy and Chung-chih Wu. It was truly an international team and making friends with them was a very memorable and enriching experience. I am especially grateful to Chee-wee for all the long hours he spent to maintain the reactor, grow and characterize the graded relaxed buffers, used for my experiments in the latter half of this thesis. His contribution to this work is greatly appreciated.

I also thank the other professors in the department and their students for providing me with equipment used to carry out my experiments and valuable technical assistance. I look forward to interacting with them in the future. I am grateful to my readers Professors Steve Forrest and Paul Prucnal for their comments and suggestions. They were kind enough to accommodate my pressing time constraints to read the manuscript. I would also like to thank Dr. Ya-Hong Xie of AT&T Bell Laboratories for his helpful ideas and samples to improve my work.

Finally I am grateful to my family for all their love, understanding and patience through my graduate life in the United States. To them, I dedicate this thesis.

Contents

List of Tables

List of Figures

x

Introduction

Motivation

The past few years have witnessed a frenzy of research activity in $Si/Si_{1-x}Ge_x$ strained layer materials and devices [1] . This heterojunction technology can potentially be integrated into the current VLSI environment, with large-scale impact in the growing microelectronics market. Developments in SiGe-base HBT's have been most rapid with peak cut-off frequencies reaching 100 GHz today, far exceeding the best all-silicon bipolar transistors. In contrast, progress in SiGe based field-effect devices has been limited. The CMOS field-effect transistor is the workhorse of modern VLSI industry. Among heterostructure electronic devices too, the modulation doped field-effect transistor (MODFET) based on III-V semiconductors dominates the industry. MODFETs currently hold all device records for lowest noise, power and delay. With peak f_T's exceeding 250 GHz and switching delays less than 5 ps, they are the fastest three-terminal devices today. Yet, the entire III-V heterojunction devices industry occupies only a few percent of the worldwide microelectronics market. Furthermore most of this share is in optoelectronics where silicon is a poor competitor. That proven heterojunction concepts cannot be put into more widespread use is most distressing. With the advent of SiGe/Si strained devices, there is renewed interest of integrating heterojunction devices with conventional silicon technology. Considering

the outstanding performance of III-V MODFETs, it is natural to ask if modulation doping and MODFETs are feasible using the Si/Ge system. This thesis attempts to answer this question. In this work, we systematically study modulation doping of both holes and electrons experimentally using structures grown by Rapid Thermal Chemical Vapor Deposition (RTCVD) , address some of the fundamental transport issues that are vital for future device modelling and demonstrate some experimental transistors.

1.2 Brief history

The microelectronics industry today is dominated by silicon MOS and bipolar transistors. These are homojunction devices where carrier flow is controlled spatially by the introduction of suitable dopant atoms. To overcome some of the limitations and design trade-offs involved in these devices, the possibility of using multiple semiconductors was actively pursued in the 1970's. These new heterostructure devices utilized two or more different semiconductors grown epitaxially on a common substrate. Because of the different electronic structure and band-gaps, carrier accumulation (or depletion) could be controlled independent of doping profiles. Furthermore, it was possible to control the potential and electric fields seen by electrons and holes independently. This added flexibility eased some of the constraints in traditional device design and led to rapid developments in Heterojunction Bipolar Transistors (HBT's) and Heterostructure Light Emitting Diodes (LED's). With the advent of advanced epitaxial growth techniques in the 1970's like Molecular Beam Epitaxy (MBE) and Chemical Vapor Deposition (CVD), it was possible to control composition and thickness on a monolayer scale. This resulted in a gamut of quantum effect devices based on tunneling, size quantization and other quantum mechanical effects.

Most of the heterostructure research till today has been carried out in the AlAs/GaAs

material system. These two III-V compound semiconductors are lattice matched to 0.1% and have similar chemical and structural properties. Epitaxial growth across the hetero-interface is almost perfect with negligible interface states or misfit dislocations. However the dominant semiconductor, silicon, lacked such a chemically similar lattice-matched partner. Although the III-V polar compounds GaP and AlP are reasonably lattice matched to silicon, they are chemically very different. In addition, they act as dopants in the silicon lattice. Initial attempts of growing Si/GaP heterostructures in the late 1970's were met with a host of cross-doping, surface segregation and three-dimensional growth problems and were quickly abandoned. Subsequent efforts concentrated on the more chemically similar Si/Ge system. The 4% difference in the lattice constants of Si and Ge, however, was a major stumbling block. This mismatch was accommodated by misfit dislocations at the interface and threading dislocations propagating to the surface through all active layers. The dislocations introduced efficient deep-level traps that reduced carrier lifetimes and increased leakage currents significantly. This problem was overcome in the early 1980's with the development of strained layer epitaxy, when it was realized that the lattice mismatch can be accommodated by homogeneous strain in sufficiently thin epitaxial layers without the generation of dislocations [2]. Subsequently rapid advances were made in the strained Si/Ge heterostructure technology.

Thesis outline

Chapter 1 provides the introduction and the purpose of undertaking this thesis work.

Chapter 2 elaborates on modulation doping structures, the specific issues in the $Si/Si_{1-x}Ge_x$ system, and briefly reviews the basic scattering mechanisms that determine the mobility in these structures.

In chapter 3 we study modulation doping of holes. We use the two-dimensional

hole gas (2DHG) formed in these structures at the interface as a sensitive probe of boron segregation, background impurity and interface quality in our growth system.

In chapter 4 we study n-type modulation doping on thick, relaxed SiGe buffer layers. Such a design is necessary to provide the strained silicon channel and sufficient conduction band offset for the modulation doping. We investigate limits on low-temperature mobilities and deduce the background doping concentration in our growth system.

In chapter 5 we demonstrate the feasibility of transistors based on modulation doping in this material system. For this purpose, we fabricated devices using a simple completely self-aligned two-step process.

Chapter 6 presents a fundamental limitation of transport in SiGe alloys, namely alloy scattering. We experimentally derive the value of this scattering potential and try to correlate this mechanism with optical studies like photoluminescence. We also propose a novel explanation of the "valley-splitting" issue, observed recently in multi-valley conduction band systems including silicon.

Finally in chapter 7 we present a synopsis of this thesis and insights it provides into our understanding of transport mechanisms in the strained SiGe material. We speculate on directions along which future work might proceed to enhance this knowledge.

Si/Si$_{1-x}$Ge$_x$ modulation doped heterostructures

2.1 Introduction

Modulation doped structures were first conceived by Dingle *et. al.* in 1978 [3]. In such structures, carriers are confined in a narrow two-dimensional potential well spatially separated from the dopants. This suppresses coulombic scattering and leads to very high low-field mobilities at low temperatures. Practical requirements of a high purity heterostructure with negligible interface defects are best satisfied in the AlAs/GaAs lattice-matched compound semiconductor system where most of the studies have been carried out. With continuing refinements in epitaxial growth techniques, very impressive mobility results exceeding 10 million cm^2/V·s have been achieved in this system. Modulation doping has also been demonstrated in other compound semiconductor systems like Al$_{0.47}$In$_{0.53}$As/Ga$_{0.48}$In$_{0.52}$As lattice matched to InP and strained AlGaAs/InGaAs on GaAs substrates. In sections 2.2 and 2.3 the equilibrium band structure and carrier transport properties of a generic modulation doped heterostructure are reviewed while Sec. 2.4 describes some specific issues related to the Si/Ge system.

2.2 Equilibrium band structure

In its simplest form, a typical modulation doped electron structure consists of a heterojunction between two semiconductors with a conduction band offset and which is intentionally doped n-type on the higher energy side. The advantage of such a configuration can be seen from a band diagram as shown in Fig. 2.1. The bulk fermi level on the doped side of the junction lies closer to the conduction band than the undoped side. Under thermal equilibrium, the fermi levels align through electron transfer across the interface. The free electrons and their positively charged parent donors form a dipole whose strong electric field confines the mobile carriers close to the interface. Analogous hole structures can be formed with a sufficient valence band offset. The situation is very similar to the common Si metal-oxide-semiconductor (MOS) system in which the inversion charge induced by a gate electric field is confined at the oxide-semiconductor interface. This system has been extensively characterized and a comprehensive review has been given by Ando et al. [4]. Unlike the Si–SiO$_2$ interface, the interface in a good semiconductor heterostructure is almost defect-free. Furthermore, by introducing an undoped "spacer" region between the dopant and the interface, ionized impurity scattering is reduced, resulting in high carrier mobilities at low temperatures.

In the rest of this section we review how to find the two-dimensional (2-D) electron density in such a system as a function of structural parameters. Because of the narrow potential well at the interface, the motion along the growth direction (z-axis) is quantized into two-dimensional subbands [5]. In the effective mass approximation, the electronic states are characterized by a subband index i and two-dimensional wave vector $\mathbf{k} = (k_x, k_y)$ along the interface. The envelope wavefunction and energy are given by:

$$\Psi_{i,\mathbf{k}}(\mathbf{r}, z) = \chi_i(z) \, \exp(i\mathbf{k} \cdot \mathbf{r}) \qquad (2.1)$$
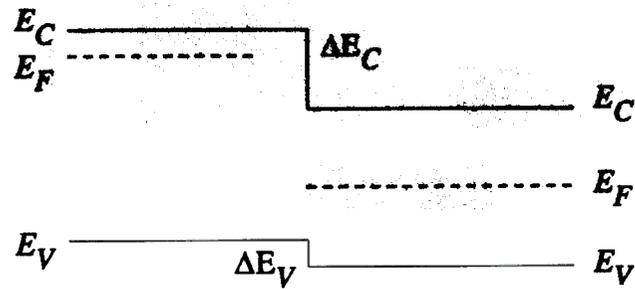
(b)
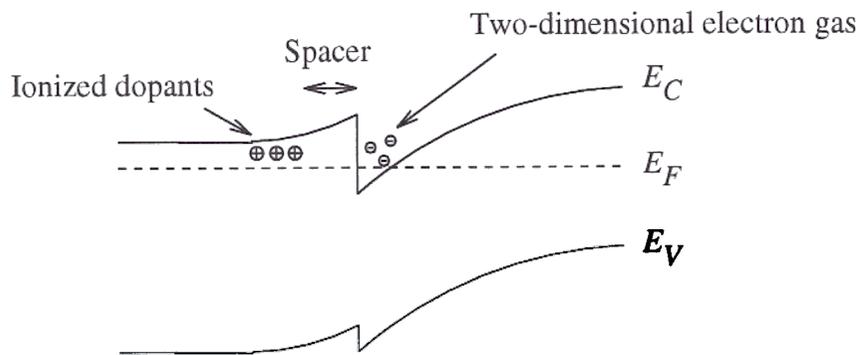
Figure 2.1: Schematic band diagrams before charge transfer (a) and after equilibrium (b) of a typical modulation doped structure. E_C, E_V and E_F are the conduction band, valence band and fermi level respectively.

$$E_i(k) = E_i + \hbar^2 k^2 / 2m_{xy}$$

where $\hbar$ is the reduced Planck constant, m_{xy} is the transverse effective mass (assumed to be isotropic) in the plane and $\mathbf{r}(x, y)$ is the *in-plane* position vector. The subband edges E_i are given by the solution of the Schrödinger equation

$$\frac{-\hbar^2}{2m_z} \frac{d^2 \chi_i(z)}{dz^2} + V(z) = E_i \chi_i(z)$$

where m_z is the longitudinal effective mass along the growth direction and $V(z)$ is the confining potential. Since this potential is determined by the charge density distribution (through Poisson equation) and therefore $\chi_i(z)$, we have a coupled system of differential equations that must be solved self-consistently. A standard way to avoid this is to use the variational approximation. In the following discussion we make the following assumptions:

1. Low temperature limit holds ($T \rightarrow 0$).

2. Only the lowest subband is occupied.

3. The wavefunction is strongly confined to the interface with negligible penetration into the potential barrier.

4. The system is isotropic in the plane of the 2-D gas.

The validity of these assumptions in interpreting our experimental data is discussed in Sec. 2.4. We use the Fang-Howard-Stern [6] one parameter variational function:

$$\chi_0(z) = \sqrt{\frac{b^3}{2}} z \exp\left(\frac{bz}{2} \right)$$

This wavefunction is used to determine the potential $V(z)$ as a function of the variational parameter b, the integrated two-dimensional electron density n and the unintentional background depletion charge density n_{dep}. Using this potential, the total

Hartree energy of the electron system is minimized to obtain (in SI units)

$$E_0 = \left(\frac{3}{2}\right)^{5/3} \left(\frac{e^2\hbar}{m_z^{1/2}\epsilon}\right)^{2/3} \frac{(n_{dep} + \frac{55}{96}n)}{(n_{dep} + \frac{11}{32}n)^{1/3}}$$

with

$$b = \left(\frac{12m_z e^2}{\epsilon\hbar^2}(n_{dep} + \frac{11}{32}n)\right)^{1/3}$$

At low temperatures, the electron system is highly degenerate and behaves like a two-dimensional fermi gas. The density of states in two-dimensions is independent of energy and given by:

$$\mathcal{D}(E) = g_s g_v \frac{m_{xy}}{2\pi\hbar^2} \tag{2.1}$$

In the above equations, ϵ is the dielectric permittivity of the semiconductor, e the electron charge, g_s the spin degeneracy and g_v the valley degeneracy. The electron density in the lowest subband at zero temperature is now given by

$$n \quad \int_{E_0}^{E_F} \mathcal{D}(E)dE = g_s g_v \frac{m_{xy}}{2\pi\hbar^2}(E_F - E_0)$$

where E_F is the fermi level. The carrier density has thus been related to the level and the subband energies.

To evaluate the quantities n, E_0 and E_F from the built-in parameters of the heterojunction, namely the doping concentration N_D, the conduction band offset ΔE_C and the undoped spacer thickness W_{sp}, thermal equilibrium conditions *viz.* constant fermi level across the structure must be invoked. Fig. 2.2 shows the conduction band in detail. From Gauss's law, the electric field $\mathcal{E}$ at the interface is given by

$$\mathcal{E} = e\frac{(n + n_{dep})}{\epsilon}$$

If we neglect the difference in dielectric constants and fixed interface charges, the electric field is continuous across the heterojunction. The potential drop in the undoped spacer Φ_{SP} and in the doped region Φ_D are given by:

$$\Phi_{SP} = \mathcal{E} W_{SP} \tag{2.10}$$
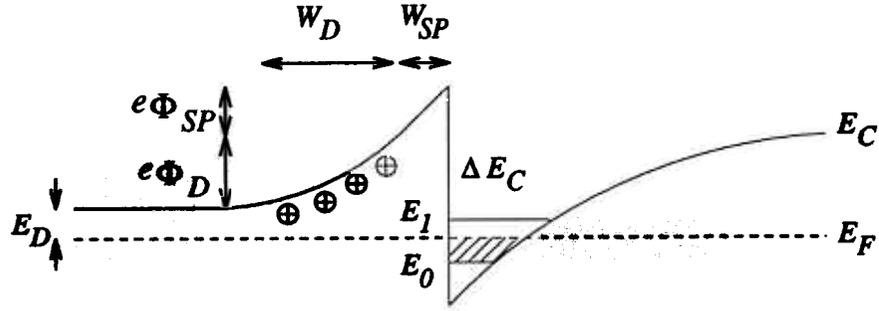
Figure 2.2: Detailed schematic of the conduction band showing two quantized sub-band levels with only lowest level occupation at zero temperature. Energies E_0, E_1 and E_F are measured with respect to the conduction band minimum at the interface. E_D is the donor level below the conduction band.

$$\Phi_D = \frac{eN_D W_D^2}{2\epsilon} = \frac{e\mathcal{E}^2}{2eN_D} \qquad (2.11)$$

The effective width of the ionized dopants W_D is given by the charge neutrality condition $N_D W_D = n + n_{dep}$. At low temperatures, the fermi level on the doped side of the junction coincides with the donor level E_D Under thermal equilibrium condition we can readily see from Fig. 2.2 that

$$E_0 + (E_F - E_0) + E_D + e\Phi_D + e\Phi_{SP} = \Delta E_C \qquad (2.12)$$

On substituting the various quantities from Eqns. 2.5, 2.8, 2.9, 2.10 and Eqn. 2.12 can be solved for the equilibrium carrier density at the interface in terms of the structural parameters N_D, W_{SP}, ΔE_C, E_D, m_z, m_{xy} and ϵ. In Chap. 3 and 4, we will use this expression to fit the experimental data.

2.3 Review of two-dimensional carrier transport

2.3.1 Scattering mechanisms

The various scattering mechanisms that determine the transport properties of a two-dimensional electron gas in a modulation doped structure have been well characterized
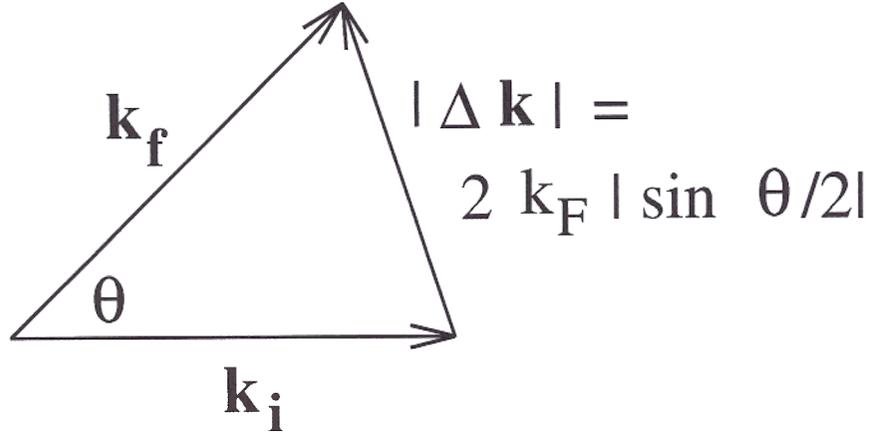
Figure 2.3: Elastic scattering geometry in two-dimensions. k_i and k_f are the initial and final wavevectors respectively which are equal to the fermi wavevector, k_F, in magnitude. The scattering angle is θ.

previously by a number of workers [6]. Broadly they can be classified as inelastic and elastic mechanisms. Inelastic scattering is mainly caused by phonons and dominates the transport at high temperatures (T $> \approx$ 40K). In this thesis, we concentrate only on the low temperature (T $\leq$ 10K) mobilities and attempt to explain the experimental data using simple theoretical models. We will therefore ignore phonon scattering and discuss the three major elastic scattering mechanisms at low temperature, namely, coulombic scattering, alloy scattering and interface roughness scattering. In the next section we display expressions for the mobility limited by these mechanisms.

2.3.2 Low field mobility

For elastic scattering, one can define a momentum relaxation time as a function of energy $\tau(E)$ (assumed to be isotropic) in the Boltzmann transport equation and obtain for the low-field mobility μ,

$$\mu \quad \frac{e <\tau>}{m_{xy}} \qquad (2.13)$$

where $<\tau>$ is the momentum relaxation time averaged over the energy distribution of the carriers. For a degenerate system at low temperature this reduces trivially to $<\tau> = \tau(E_F)$.

The elastic scattering geometry is shown in Fig. 2.3. The final state two-dimensional wavevector $\mathbf{k_f}$ is equal in magnitude to the initial state wavevector $\mathbf{k_i}$ but rotated by a scattering angle θ. The scattering is thus completely described by the magnitude of the fermi wavevector k_F (or equivalently, E_F) and the angle θ. The momentum relaxation time, $\tau(E_F)$, is the weighted average of the relaxation time $\tau_r(E_F, \theta)$ over all scattering angles:

$$\frac{1}{\tau(E_F)} \quad \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\tau_r(E_F, \theta)} (1 - \cos\theta)\, d\theta \qquad (2.14)$$

Physically this means that although all collisions determine the relaxation time, only large angle scattering degrades mobility significantly. The factor $(1 - \cos\theta)$ ensures that small angle scattering ($\theta \sim 0$) does not contribute to the transport time. We will see below how this affects the different scattering mechanisms.

Given the scattering potential V_{scatt} it is possible to calculate the relaxation time using Fermi's Golden Rule:

$$\frac{1}{\tau_r} = \frac{2\pi}{\hbar} |< \Psi_{k_i} \mid V_{scatt} \mid \Psi_{k_f} >|^2 S\mathcal{D}(E_F) \qquad (2.15)$$

where $M \equiv < \Psi_{k_i} \mid V_{scatt} \mid \Psi_{k_f} >$ is the matrix element between the initial and final states, and S is the sample area. Substituting Eqns. 2.15, 2.14, and 2.7 into

Eqn. 2.13, we finally get:

$$\mu = \frac{2\pi e \hbar^3}{m_{xy}^2 S \int_0^{2\pi} |M|^2 (1 - \cos\theta) \, d\theta} \tag{2.16}$$

Note that g_s 1 in the above expression because the scattering does not change the spin state. Also a single-valley band ($g_v = 1$) has been assumed. The effect of multi-valley degeneracies that occurs in the case of Si/Si$_{1-x}$Ge$_x$ is discussed in Sec. 2.4

Eqn. 2.16 suggests that the mobility is inversely proportional to the square of the effective mass. This is the case for short-range potentials like alloy scattering and interface roughness scattering. For the long-range coulombic scattering, the potential is screened by the two-dimensional gas. The strength of the screening is proportional to the density of states (Eqn. 2.7). Thus $M \sim m_{xy}^{-1}$ and the mobility becomes independent of the mass. This result is important when comparing low temperature mobilities of different material systems like AlGaAs/GaAs and Si/Si$_{1-x}$Ge$_x$.

Coulombic scattering

Coulombic scattering is caused by ionized impurities either as intentionally introduced dopants separated by a spacer layer (remote impurity scattering) or background impurities incorporated during the growth throughout the heterostructure (background impurity scattering). In the following the wavefunction is assumed to be tightly confined at the interface and can be approximated by a delta function. Because of the slowly-varying long-range nature of the potential, states with large differences in wavevector (large θ) are weakly coupled and the screened matrix element, M_{imp}, peaks strongly in the forward direction (small θ):

$$M_{imp}|^2 \quad \frac{N_{imp}}{} \left(\frac{2\pi e^2}{\epsilon q_0}\right)^2 \int_{z_{min}}^{z_{max}} \exp(-4k_F z \sin\frac{\theta}{2}) \, dz \tag{2.17}$$

where N_{imp} is the concentration of impurities (assumed constant) of charge e distributed from z_{min} to z_{max} and $q_0 = 2m_{xy}e^2/\epsilon\hbar^2$ is the two-dimensional screening

constant. The exponential factor insures the dominance of small-angle scattering and the mobility remains fairly high.

For a remote dopant concentration of N_D we integrate Eqn. 2.17 from z_{min} $W_{SP} + W_{2DEG}$ to $z_{max} = W_{SP} + W_D + W_{2DEG}$ where $W_{2DEG} = 3/b$ is the average distance of the 2DEG from the interface and use Eqn. 2.16 to obtain:

$$\mu_{RI} \quad \left(\frac{16e}{\pi\hbar}\right) \frac{k_F^3}{N_D \left(\frac{1}{(W_{SP}+W_{2DEG})^2} \quad \frac{1}{(W_{SP}+W_D+W_{2DEG})^2}\right)}, \quad k_F = \sqrt{\frac{4\pi n}{g_s g_v}} \quad (2.18)$$

Similarly for background impurity concentration N_B, we integrate Eqn. 2.17 from $z_{min} = -\infty$ to $z_{max} = +\infty$ and substitute in Eqn. 2.16 to get:

$$\mu_{BI} = \frac{ek_F}{2\pi\hbar N_B}, \quad k_F = \sqrt{\frac{4\pi n}{g_s g_v}} \quad (2.19)$$

Hence the mobilities $\mu_{RI} \sim n^{3/2}$ and $\mu_{BI} \sim n^{1/2}$ are strong functions of the carrier density, characteristic of coulombic scattering. In the case of a multi-valley system like Si$_{1-x}$Ge$_x$ these equations continue to hold because the slowly-varying coulomb potential cannot couple states with large differences in wavevectors and causes no intervalley scattering.

Alloy scattering

In disordered random alloys like Si$_{1-x}$Ge$_x$, the concept of band structure, Bloch waves, etc is in principle invalid because of the absence of strict translational invariance of the crystal potential V_C. Nevertheless their behaviour in a variety of transport and optical experiments suggests that their electronic structure is akin to their ordered counterparts. The simplest way to get around this difficulty is to use the "virtual crystal approximation" and write the aperiodic crystal potential which is the sum of the atomic potentials V_{Si} and V_{Ge} at random lattice sites $\mathbf{R}$

$$V_C(\mathbf{r}) = \sum_{\mathbf{R}_{Si}} V_{Si}(\mathbf{r} \quad \mathbf{R}_{Si}) + \sum_{\mathbf{R}_{Ge}} V_{Ge}(\mathbf{r} - \mathbf{R}_{Ge}) \quad (2.20)$$

as the sum of a periodic potential due to an "average" atom and a fluctuating disorder potential which can be treated as a perturbation:

$$V_C(\mathbf{r}) = \sum_{all\ \mathbf{R}} (1-x)V_{Si}(\mathbf{r}-\mathbf{R}) + xV_{Ge}(\mathbf{r}-\mathbf{R}) + V_{dis}(\mathbf{r})$$

where the disorder potential V_{dis} is a sum of random strongly localized functions effective only within a unit cell. Thus alloy scattering, in contrast to impurity scattering, is short-ranged and screening is ineffective. The delta-function like potential couples all states in K-space equally and the matrix element M_{alloy} is independent of the scattering angle and given by [6]

$$M_{alloy}\mid^2 = \frac{\Omega_0}{S}x(1-x)V_{alloy}^2 \int \chi_0^4(z)\,dz \tag{2.22}$$

where Ω_0 is the atomic voume, V_{alloy} is a parameter that characterizes the strength of the disorder potential and x is the alloy concentration. Substituting in Eqn. 2.16, we obtain:

$$\mu_{alloy} = \frac{e\hbar^3}{m_{xy}^2\Omega_0 x(1-x)V_{alloy}^2 \int \chi_0^4(z)\,dz} \tag{2.23}$$

The alloy scattering limited mobility is strongly dependent on effective mass and causes significant mobility degradation in large mass systems like Si$_{1-x}$Ge$_x$ (see Sec. 2.4) . Alloy scattering by the barrier is negligible because the wavefunction penetration is very small. However when the channel is also an alloy, the integral in Eqn. 2.23 becomes significant. Using Eqn. 2.4 for the wavefunction, we obtain

$$\mu_{alloy} = \frac{e\hbar^3}{m_{xy}^2\Omega_0 x(1-x)V_{alloy}^2} \frac{16}{3b} \tag{2.24}$$

where $b \sim n^{1/3}$ is the variational parameter given by Eqn. 2.6. Thus alloy scattering depends strongly on x and weakly on the carrier density ($\mu_{alloy} \sim n^{-1/3}$). This mechanism has been extensively investigated in Chap. 6. It should be noted that the classical formula for alloy mobilities in three dimensional systems [7] diverges at low temperatures and cannot be used for 2-D systems.

Interface roughness scattering

The maximum low temperature mobilities in the Si-SiO$_2$ system has been found to be limited by scattering at the "rough" oxide-semiconductor interface. Ando [8] has modelled this scattering potential as a sum of uncorrelated localized potentials similar to alloy scattering. The nature of this scattering is short-range and therefore leads to a mobility expression similar to Eqn. 2.23. Because the interfaces in lattice-matched semiconductor heterojunctions like AlGaAs/GaAs and Si/strained Si$_{1-x}$Ge$_x$ are much more uniform on an atomic scale, interface roughness is not expected to play a role in limiting the mobility below $\sim 5 \times 10^5$ cm^2/V·s at low carrier densities ($< 10^{12}$ cm^{-2}). Indeed, magnetotransport experiments indicate that the highest mobilities in pure silicon channels are limited by long-range coulombic scattering. We will therefore ignore this mechanism in the rest of this work.

From Eqn. 2.16 the total mobility due to all scattering mechanisms is given by:

$$\mu_{total}^{-1} = \mu_{RI}^{-1} + \mu_{BI}^{-1} + \mu_{alloy}^{-1} \qquad (2.25)$$

This expression, the so-called "Matthiessen's Rule", is valid only for degenerate low-temperature systems where all scattering takes place at the fermi energy.

2.3.3 Transport in magnetic fields

In a magnetic field, the resistivity (or conductivity) of a two-dimensional isotropic system is no longer a scalar, but a tensor:

$$\rho = \begin{matrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{xx} \end{matrix} \qquad (2.26)$$

Classically, for low magnetic fields ($\mu B \ll 1$), the transverse resistivity ρ_{xy} is given by the Hall coefficient:

$$\rho_{xy} = r_H \frac{B}{ne} \qquad (2.27)$$

(a)   (b)   (c)

B = 1   B = 1.5   B = 2
G = 2   G = 3   G = 4
ρ_{xx} low ρ_{xx} high ρ_{xx} low

$\hbar\omega_c$
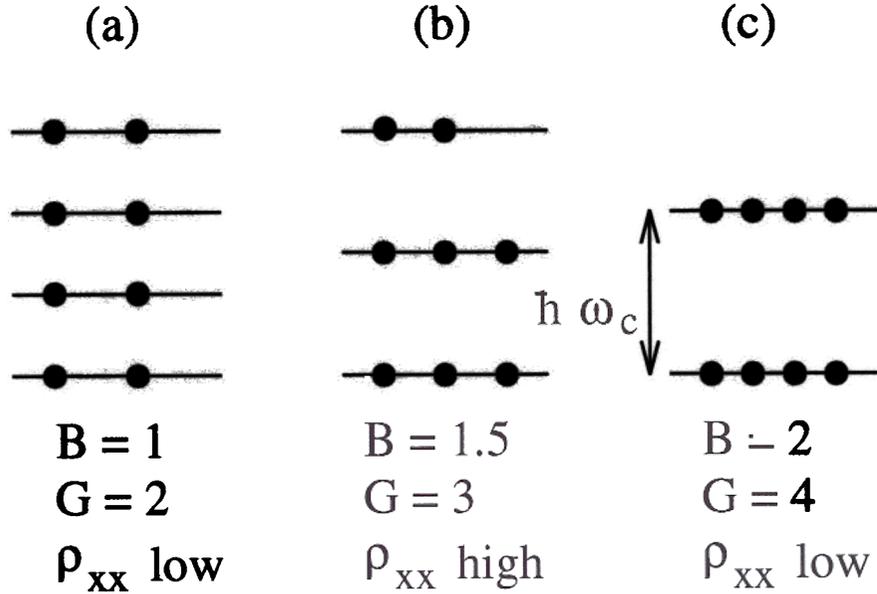
Figure 2.4: Schematic diagram showing Landau level structure for a hypothetical system of 8 electrons in a magnetic field. The units for the field (B) and the degeneracy of each level (G) is arbitrary. As the field increases and the occupancy of the highest level changes, the longitudinal resistance (ρ_{xx}) oscillates in magnitude.

where B is the magnetic field and $r_H = <\tau^2>/<\tau>^2$ is the Hall factor. Again, for degenerate low temperature systems, the Hall factor reduces trivially to unity. Hence there is no difference between Hall concentrations (or mobilities) and the actual carrier concentration (or drift mobility) [1]. The longitudinal resistivity ρ_{xx} is not affected by low magnetic fields to first order.

Under the influence of high magnetic fields ($\mu B \gg 1$), the energy subbands become further quantized into equally spaced Landau levels separated by the cyclotron energy $\hbar\omega_c$, $\omega_c = eB/m_{xy}$ (see Fig. 2.4). Each Landau level has a degeneracy per unit area given by $g_s g_v eB/2\pi\hbar$ (assuming spin degeneracy is not resolved). As the magnetic field is increased, both the occupancy of each level and the separation between levels increases. When an integer number of levels are fully occupied and $k_B T \ll \hbar\omega_c$, where

[1]This has been confirmed experimentally by comparing the Hall concentration of a single 2-D gas with the concentration obtained from Shubnikov-deHaas oscillations of the longitudinal resistance.

k_B is the Boltzmann constant, there are no available states for scattering into and the resisitivity is very small $(\rho_{xx} \rightarrow 0)$ [2]. On the other hand, when the top level is only half-filled, maximum scattering can take place and the resistivity is high. Thus the longitudinal resistance oscillates as a function of the magnetic field. The period of these Shubnikov-deHaas oscillations is easily calculated. For a given density of carriers, n, let i levels be completely filled for a magnetic field B_i. Then

$$n = i\frac{g_s g_v e B_i}{2\pi\hbar} \tag{2.28}$$

As the field is increased to B_{i-1}, the number of fully occupied levels drops by one:

$$n = (i-1)\frac{g_s g_v e B_{i-1}}{2\pi\hbar} \tag{2.29}$$

From the above equations, we obtain

$$n = \frac{g_s g_v e}{2\pi\hbar\Delta\left(\frac{1}{B}\right)}, \quad \Delta\left(\frac{1}{B}\right) = \frac{1}{B_i} - \frac{1}{B_{i-1}}$$

The Shubnikov-deHaas oscillations are therefore periodic in the reciprocal magnetic field. The period can be used to determine the carrier density from the above equation. Alternatively, if the density is independently determined from the Hall slope, the degeneracies g_s and g_v can be established.

Whenever an integer number of Landau levels are completely filled and ρ_{xx} shows a minimum, the transverse resistivity ρ_{xy} exhibits a plateau. From Eqns. 2.27 and 2.28 we can see that the value of the plateau is quantized as $h/ig_s g_v e^2$ where h $(= 2\pi\hbar)$ is the Planck constant. The Quantized Hall Effect was first observed in Si-MOSFET's in 1980 by Klaus von Klitzing *et. al.* [9] and has been, since then, extensively characterized in the AlGaAs/GaAs system. It has also been observed in other III-V systems and strained Si/Si$_{1-x}$Ge$_x$ heterostructures.

---

[2]In the presence of a strong magnetic field, the eigenstates of the ideal system are no longer Bloch states. Instead they are the well-known harmonic oscillator eigenstates with no average velocity. If a driving electric field is added perpendicular to the magnetic field, then *all* eigenstates acquire an average drift velocity given by E/B in a direction perpendicular to both the fields. Hence all states in a *full* Landau level (in the absence of disorder) carry the same current without resistance.

2.4 Modulation doping in the Si/Si$_{1-x}$Ge$_x$ strained system

The band structure of Si$_{1-x}$Ge$_x$ has been theoretically and experimentally investigated by a number of groups [10, 11, 12, 13]. In thin quantum wells, where the lattice mismatch is fully accommodated by biaxial strain, the dramatic effect of strain on the band-gap and heterojunction band alignments has been clearly observed in optical absorption, photoluminescence and transport studies [11, 14, 15, 16]. As mentioned before, in a virtual crystal approximation, the band structure of bulk disordered Si$_{1-x}$Ge$_x$ alloys is similar to bulk silicon (for $x < 0.85$) with six-fold degenerate conduction band minima at the $\Delta -$ point and a degenerate valence band consisting of the light hole, heavy hole and spin-orbit split-off bands (see Fig. 2.5). All samples used in this work were grown on $<100>$ substrates so that, by symmetry, the system is isotropic in the growth plane. The in-plane biaxial strain splits the six-fold valley degeneracy in the conduction band into two-fold and four-fold degenerate valleys with energy configuration depending on sign and magnitude of the strain as shown in Fig. 2.5. The effective masses of these ellipsoidal valleys for strained Si ($x = 0$) were determined from cylotron resonance experiments to be close to the bulk unstrained silicon values ($m_z = 0.98, m_{xy} = 0.19$) [17]. The degeneracy at k = 0 in the valence band is similarly lifted. The in-plane constant energy surfaces have been mapped out by angle-resolved magnetotunneling [18] and found to be warped. The effective mass in this case strongly depends on strain and varies from $m_z \sim m_{xy} = 0.44$ for $x = 0$ to $m_z \sim m_{xy} = 0.32$ for $x = 0.25$. The relatively large effective masses means that quantum size effects are not as important as in the more common electrons in AlGaAs/GaAs system. Also, the density of states (Eqn. 2.7) is large so that occupation of the second subband does not occur until very high carrier densities. For electrons, $\mathcal{D}(E) \sim 1.6 \times 10^{11}$ cm^{-2}meV^{-1}, so that, assuming $E_1 - E_0 \sim 20$ meV, the second level gets occupied at a density of 3.2×10^{12} cm^{-2} . Hence the assumptions
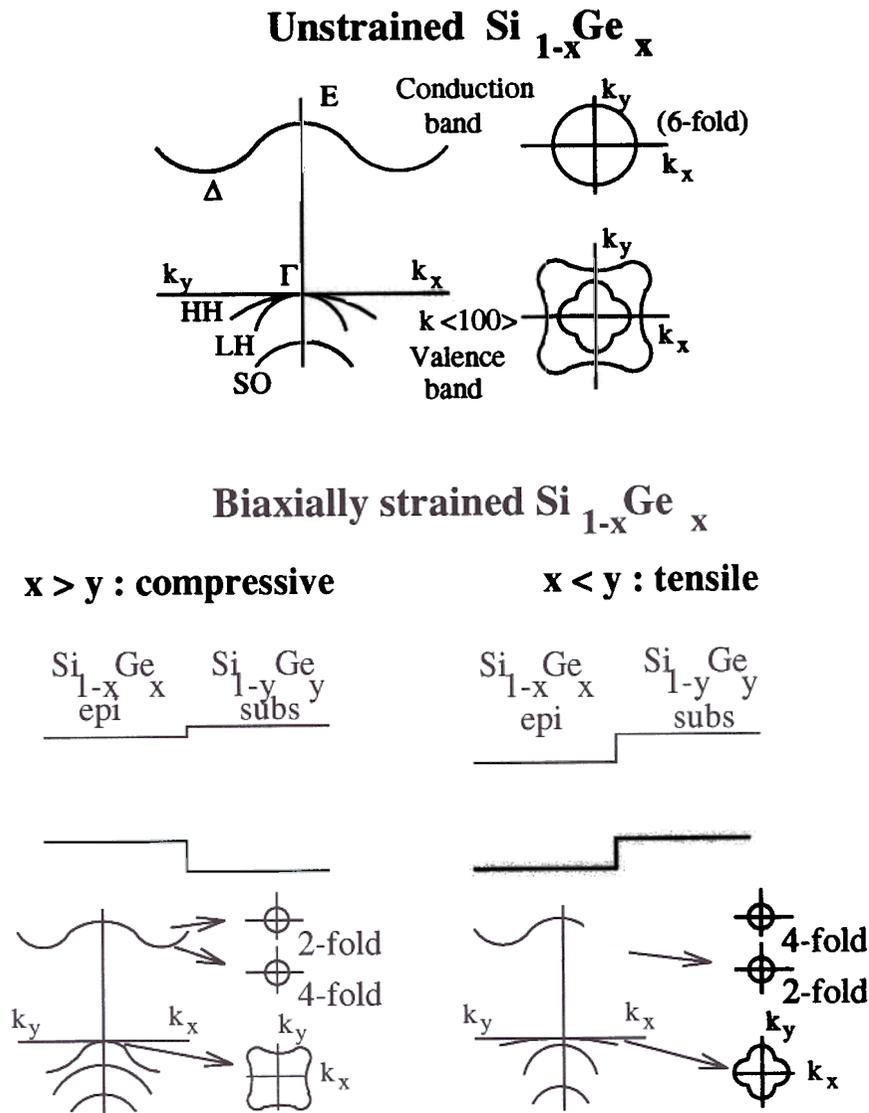
Figure 2.5: Schematic band structure of Si$_{1-x}$Ge$_x$ alloys showing the effect of strain on degeneracies and band alignments. HH, LH and SO represent the heavy hole, light hole and split off valence bands in the unstrained case. The strain splits the degeneracies and affects the band alignment. Note that a substantial conduction band offset is obtained only for tensilely strained case.

in Sec. 2.2 are well justified.

Fig. 2.5 also shows the strong influence of strain on band-alignments for the general case of a strained Si$_{1-x}$Ge$_x$ epitaxial layer on a relaxed Si$_{1-y}$Ge$_y$ substrate. These alignments have been calculated theoretically by van de Walle and Martin [10]. When $x > y$ (compressive strain) nearly all the band offset lies in the valence band, with negligible conduction band offset. Modulation doping with strained Si$_{1-x}$Ge$_x$ on Si substrates was, in fact, first used to establish band alignments in this material system [16]. The subsequent realization that $x < y$ or tensile strain is necessary to obtain sufficient conduction band offset led to the fabrication of the first two dimensional electron gases [19] Due to the lack of a Si$_{1-y}$Ge$_y$ ($y > 0$) substrate, the initial layers were grown on thick epitaxial Si$_{1-y}$Ge$_y$ layers on Si substrates with numerous dislocations. The optimization of this "relaxed" buffer to obtain the highest mobilities in this system ($> 100,000$ cm^2/V·s) has been the subject of intense investigation in the last 2–3 years and forms a major part of this thesis (Chap. 4).

The RTCVD growth system at Princeton

The epitaxial layers described in this thesis were grown by Rapid Thermal Chemical Vapor Deposition (RTCVD) which is a combination of rapid thermal processing and chemical vapor deposition [20]. A schematic of the reactor is shown in Fig. 2.6. It consists of a 17.5 cm diameter cylindrical quartz tube in which a quartz stand supporting the wafer is placed. The wafer is heated by a bank of tungsten-halogen lamps which allows for rapid temperature switching and optimized growth temperatures for individual epitaxial layers. The temperature is monitored by detecting transmission of 1.3μm and 1.5μm infrared light from semiconductor lasers through the wafer [21]. A feedback system controls the wafer temperature accurately to within a few degrees. This is important because the growth rates depend exponentially on temperature

**RTCVD Reactor with Temperature Monitor**
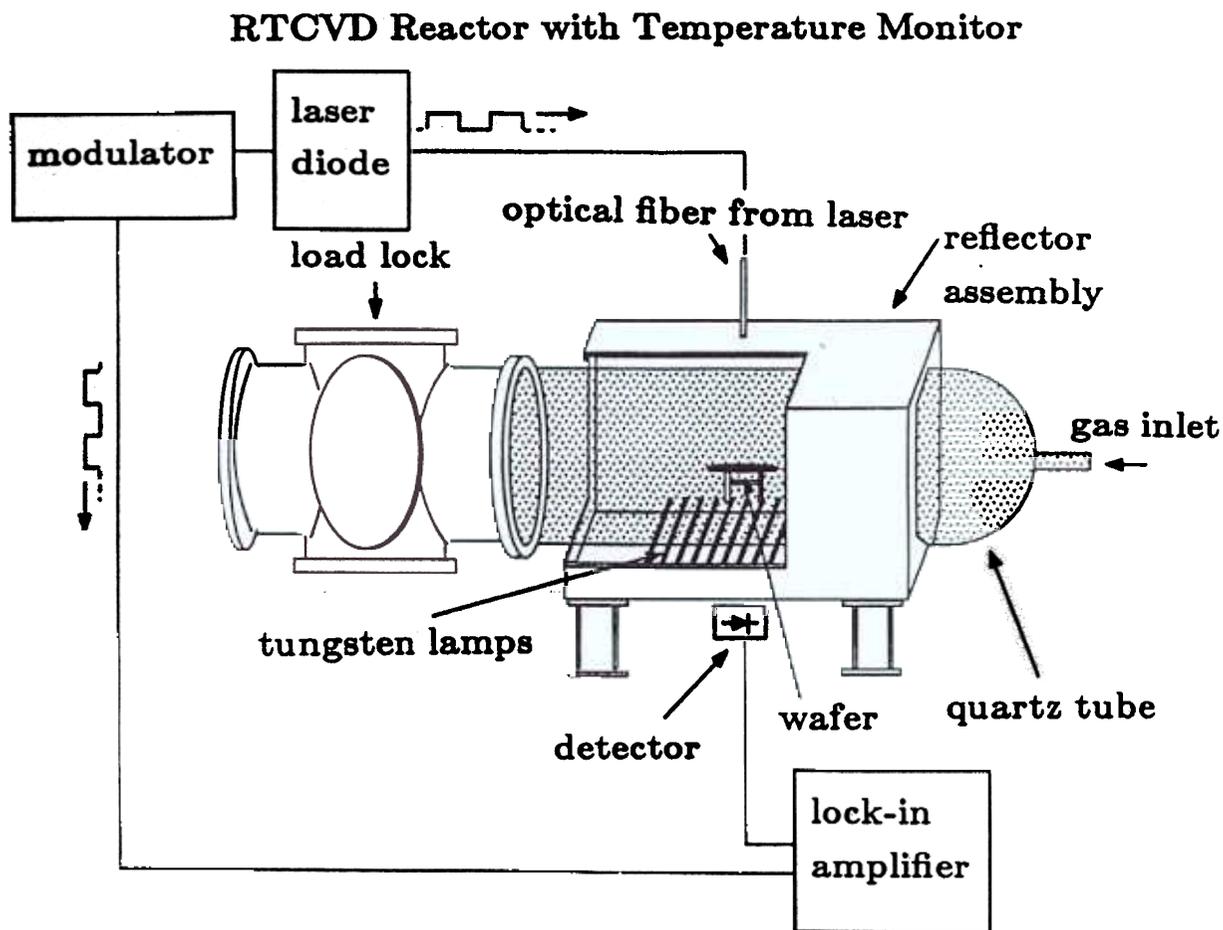
Figure 2.6: Schematic picture of the RTCVD growth system at Princeton. The pumping system consisting of a rotary vane mechanical pump with a burn-box exhaust is not shown. A recently added Auger analysis chamber to the rear of the load lock is also not shown in this diagram.

under our growth conditions.

The gases used are dichlorosilane (SiCl$_2$H$_2$) and silane (SiH$_4$) for silicon and germane (0.8% GeH$_4$ in H$_2$) for germanium. The doping is accomplished simply by switching a phosphine (70 ppm PH$_3$ in H$_2$) or diborane (10 ppm B$_2$H$_6$ in H$_2$) source on and off. The gas mixture is established in a hydrogen carrier gas prior to injecting into the reactor and exhausted by a rotary vane mechanical pump. The wafers are transferred into the reactor tube through a load lock which reduces residual oxygen contamination in the epitaxial films to less than 10^{18} cm^{-3}. This ensures high carrier lifetime in our films [22] and could also play a crucial role in reducing interface states for successful modulation doped structures.

Under a typical growth pressure of 6 torr, silicon layers are grown at 700°C (30 Å/min) and 1000°C (0.5 μm/min) using 26 sccm of dichlorosilane in 3 slpm of hydrogen. The Si$_{1-x}$Ge$_x$ layers are grown at 625°C by adding germane to this mixture, depending on the Ge content required. For instance, using 26 sccm of SiCl$_2$H$_2$ and 100 sccm of GeH$_4$ results in a Si$_{0.80}$Ge$_{0.20}$ film ($\sim$100 Å/min) while the same SiCl$_2$H$_2$ flow with 450 sccm of GeH$_4$ results in a Si$_{0.62}$Ge$_{0.38}$ film ($\sim$300 Å/min). The growth rate increases with Ge content in the film [23] and hence, for obtaining the linearly graded structures described in Chap. 4, it was necessary to change the germane flow non-linearly with time.

The RTCVD system described above has showed considerable promise in the past in the growth of Si and strained Si$_{1-x}$Ge$_x$ epitaxial layers. Heterojunction Bipolar Transistors (HBTs) fabricated with this system showed near-ideal base currents with high Early voltages [24, 25]. More recently strained Si$_{1-x}$Ge$_x$ quantum wells grown by RTCVD at Princeton demonstrated the first band-edge photoluminescence [14]. In the rest of this thesis we will learn the capabilities and limitations of this reactor to grow high mobility modulation- doped structures for field-effect transistor applications.

Two-dimensional Hole Gases on Si $<100>$ substrates

3.1 Introduction

The first experiments of modulation doping in Si/ $Si_{1-x}Ge_x$ were carried out using coherently strained epitaxial $Si_{1-x}Ge_x$ layers grown on the commonly available $<100>$ silicon substrate. These alloy layers had to be sufficiently thin in order to accommodate the lattice mismatch by strain. The "critical thickness" for the generation of misfit dislocations has been investigated theoretically and experimentally by various groups. It was found that although the original Matthews-Blakeslee equilibrium theory imposed too stringent criteria for practical needs, it was possible to overcome these limitations under non-equilibrium "metastable" conditions. The metastable critical thickness limits have been estimated by People et al. [26] and yield numbers roughly an order of magnitude higher, depending on growth temperature. It then became more practical to grow strained layers with sufficient germanium content, and a number of low temperature epitaxial growth techniques like Molecular Beam Epitaxy (MBE) and Chemical Vapor Deposition (CVD) were adapted for this purpose.

To make full use of these capabilities, it is essential to characterize dopant profiles and interface quality on a monolayer scale. While mass spectroscopy techniques like Secondary Ion Mass Spectroscopy (SIMS) or Rutherford Backscattering Spectroscopy

(RBS) do not have the necessary sensitivity or spatial resolution, imaging techniques like Cross-sectional Transmission Electron Microscopy (XTEM) are too time consuming, expensive and incapable of distinguishing dopants. However, from Chap. 2.2 we know that the equilibrium and transport properties of the two-dimensional gas are strong functions of the spacer thickness, doping concentration and interface roughness. We therefore use the 2DHG as a sensitive probe on an atomic scale.

Normal and Inverted Interfaces

A well known problem in the growth of AlGaAs/GaAs heterostructures by MBE is dopant segregation towards the growth direction. Some dopant species have a lower surface energy compared to their host atoms and continue to 'ride" on the surface after the source is turned off. This results in a sharp turn-on profile, but a long tail towards the surface. In "normal" modulation doped structures, where the dopant is placed *above* the channel, this poses no problem. However in 'inverted" structures, the dopants are placed *below* the channel and segregate into the channel region during growth. The increased coulombic scattering degrades the mobility by more than an order of magnitude compared to normal structures. Although a number of growth "tricks" like surfactants, growth interruptions and superlattice buffers have alleviated this problem recently, the best mobilities are still obtained in normal structures. Technologically, inverted structures are important for Modulation Doped Field Effect Transistor (MODFET) design because the carriers can be placed much closer to the gate. Furthermore dopant segregation affects the design of other devices like HBTs and Resonant Tunneling Diodes (RTDs). The first $Si/Si_{1-x}Ge_x$ hole gases [27], grown by MBE by the AT&T group, showed unequal transport characteristics for the normal and inverted interfaces. Subsequently the IBM group [28] presented evidence for virtually identical interfaces in samples grown by by Ultrahigh Vacuum

CVD (UHV-CVD). This is plausible since dopant segregation is a strong function of surface energetics which varies depending on the growth environment, such as hydrogen-terminated surfaces. Additional complications in a CVD environment like autodoping and slow gas transients also cause asymmetrical broadening of dopant profiles. In order to study this problem in our RTCVD system, we compared the 2DHG properties at both interfaces.

3.2.1 Sample growth and structure

The RTCVD reactor at Princeton has been described in detail in Sec. 2.5. The doping interfaces were achieved within the Si for modulation-doping experiments simply by switching the diborane source on and off. Once growth is started, the sample is not cooled until the final structure is complete. For all experiments described in this chapter, 4-in diameter n-type wafers (~ 1.0 Ωcm resistivity) were used. first few samples had contact problems and the cap layer doping was increased. Too much doping (> 30 sccm B_2H_6 corresponding to $\sim 6 \times 10^{18}$ cm^{-3}), however, led to large parallel conduction at low temperatures and degenerately doped behaviour. A summary of these initial investigations is given in Appendix A. In the next few sections, we discuss in detail three samples #514, #529 and #370.

Fig. 3.1 shows the normal (sample #514) and inverted (sample #529) structures. These *single* heterostructures consist of an undoped 400 Å Si$_{0.87}$Ge$_{0.13}$ channel separated by 150 Å from a 1.5×10^{18} cm^{-3} boron-doped layer on *one* side. The active layers are grown on a ~ 1 μm buffer and terminated by a heavily doped (3×10^{18} cm^{-3}) Si cap which prevents surface depletion of the channel. The doping and thickness of the cap layers were carefully adjusted to minimize "parallel" conduction at low temperature while compensating a maximum surface charge of 4.5×10^{12} cm^{-2} . Fig. 3.2 shows the schematic energy band diagram for these structures. The band alignment is type-I and the two-dimensional hole transport occurs in the *compressively* strained

p^+ - Si cap ~ 150 Å (relaxed)	p^+ - Si cap ~ 150 Å (relaxed)
i- Si cap ~ 150 Å (relaxed)	i- Si cap ~ 600 Å (relaxed)
p^+ - Si ~ 40 Å (relaxed)	
i- Si spacer ~ 150 Å (relaxed)	$Si_{0.87}Ge_{0.13} \sim 400$Å (strained)
$+ + + + + $ 2DHG $+ + + +$	$+ + + + + $ 2DHG $+ + + +$
$Si_{0.87}Ge_{0.13} \sim 400$Å (strained)	i- Si spacer ~ 150 Å (relaxed)
	p^+ - Si ~ 40 Å (relaxed)
i-Si buffer ~ 1 μm (relaxed)	i-Si buffer ~ 1 μm (relaxed)
(100) n – Si substrate	(100) n – Si substrate

Figure 3.1: Normal (sample #514) and inverted (sample #529) p-type modulation doped structures. The doping levels are 1.5×10^{18} cm^{-3} for the supply layer and 3×10^{18} cm^{-3} for the cap layer. The position of the 2DHG is schematically indicated.

$Si_{1-x}Ge_x$ layers.

3.2.2 Hall Measurements

After growth, the wafer was cleaved into $\sim 3 \times 3$ mm^2 square pieces for electrical measurements in the Van der Pauw configuration. In a few cases, the data was compared to that obtained from lithographically defined Hall bar geometries and found to agree to within 10%. The samples were degreased in TCE/Acetone/Methanol using an ultrasonic bath. The native oxide was removed with a 1:50 HF solution for 30–60 seconds. For contacts, ~ 1000 Å Al was thermally evaporated and annealed at 500 °C in a forming gas atmosphere. This yielded ohmic p-type contacts good to liquid He temperatures. The sample was insulated by a sapphire piece and mounted on the cold head of a closed cycle He refrigerator system using vacuum grease. Electrical connections to the sample were established through gold wires bonded to the contacts with

SINGLE HETEROJUNCTION


DOUBLE HETEROJUNCTION
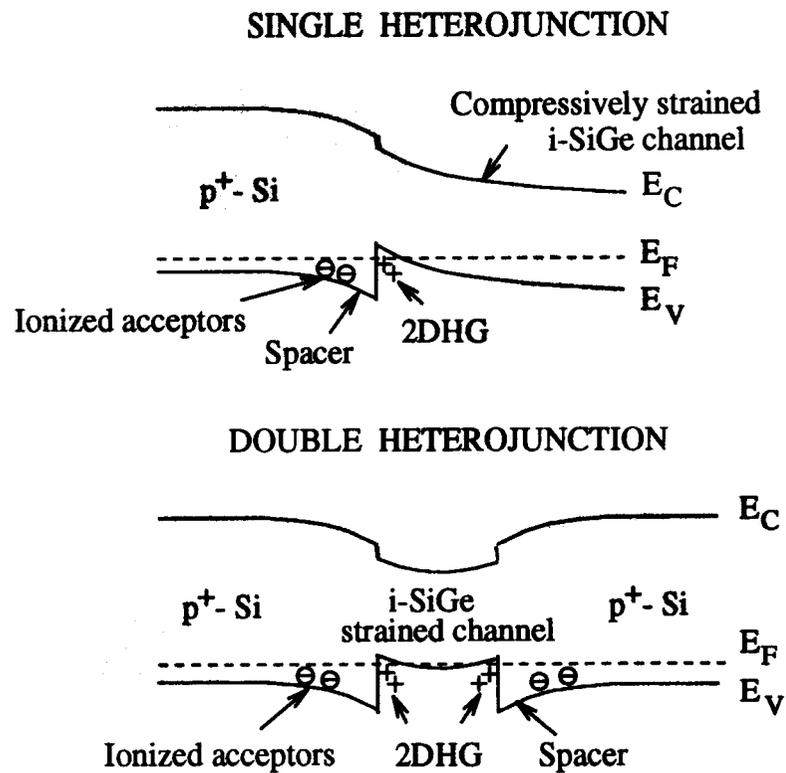
Figure 3.2: Schematic energy band diagram of single and double p-type modulation doped structures. The band alignment is type-I with negligible conduction band offset. The holes are confined in the compressively strained $Si_{1-x}Ge_x$ layers.
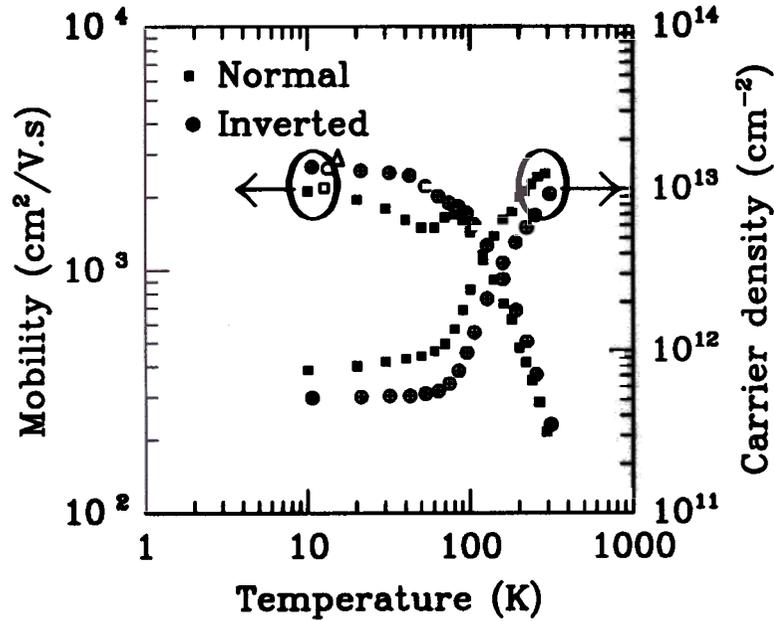
Figure 3.3: Electrical measurements for the normal and inverted structures. The open symbols represent mobility data obtained from similar samples grown by other techniques: o UHV-CVD [28], △ MBE [29], □ MBE [27].

indium. All temperature dependent resistivity and Hall measurements were carried out at an AC frequency of 17.3 Hz using sensitive lock-in phase detection technique to improve signal-to-noise ratio.

The results are displayed in Fig. 3.3. The carrier mobility increases monotonically as the temperature is lowered, reaching about 2500 cm^2/V·s at 10 K. At the same time, the carrier density decreases and saturates at about 6×10^{11} cm^{-2} , showing no signs of freeze-out. These are strong indications of the success of modulation doping in both samples. At 4.2 K, the mobility increased to 3500 cm^2/V·s and the corresponding carrier density was 4.8×10^{11} cm^{-2} . Also shown in Fig. 3.3 are published peak mobilities in samples with 20% Ge and 100 Å spacer grown by MBE [27], 15% Ge and 120 Å spacer grown by UHV-CVD [28], and 20% Ge and 100 Å spacer grown by MBE [29]. Our data is in close agreement with these values demonstrating the excellent quality of the RTCVD technique. Mobilities as high
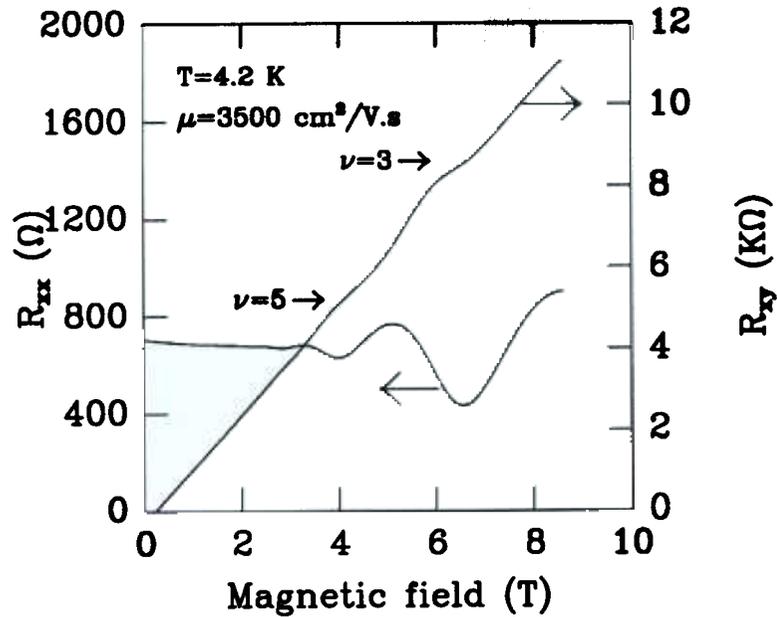
Figure 3.4: Longitudinal and Transverse resistances of the inverted 2DHG sample as a function of magnetic field at 4.2 K. The Hall resistance has a small offset at zero field due to non-symmetric sample geometry.

as 6000 cm^2/V·s have been measured in MBE grown samples at 2 K [29], but our comparisons were made around 10 K so that the UHV-CVD data could also be included.

3.2.3 Magnetotransport

The longitudinal (R_{xx}) and transverse (R_{xy}) resistances were also measured in high magnetic fields using a liquid Helium cooled superconducting magnet. The results for the inverted sample are displayed in Fig. 3.4. as a function of magnetic field at a temperature of 4.2 K. As explained in Chap. 2.3, the longitudinal resistance displays Shubnikov-deHaas oscillations which are periodic in the reciprocal field. At the same time, the transverse resistance shows a few Quantum Hall plateaus. The quantization is rather weak in this case because the condition $\mu B \gg 1$ is only weakly satisfied for this low mobility sample. Using Eqn. 2.30 we can deduce $n = 4.7 \times 10^{11}$ cm^{-2}

assuming $g_v = 1$ and $g_s = 2$. Note that although the valence band degeneracy is completely lifted by strain, the anisotropic g-factor causes the Zeeman splitting in this system to match the Landau level splitting, yielding an effective spin degeneracy of 2 [30]. This value for the density is identical to, within experimental error, the number 4.8×10^{11} cm^{-2} obtained from the transverse Hall resistance using Eqn. 2.27. This confirms the presence of a *single* two-dimensional hole gas at the inverted interface with negligible parallel conduction elsewhere. Analogous results were obtained from a magnetotransport analysis of the normal structure.

3.3   Double Heterostructures

Although nominally of similar structure, the mobility and carrier densities were not identical in the normal and inverted structures, but were 20% higher and 15% lower in the inverted structure respectively. This is contrary to what one would expect if autodoping or surface segregation caused a smaller spacer in the inverted structure. Since the samples were grown three weeks apart, it was not known if this difference was a real physical effect, or due to a small change in growth conditions. Therefore, we grew a double heterostructure incorporating both the normal and inverted interfaces at the same time. Appendix A lists some samples which initially failed to show good 2DHG behaviour and possible explanations. In the sections below, we discuss one of our better samples, #370, in detail.

3.3.1   Sample structure

The double heterostructure (sample #370) shown in Fig. 3.5 has an active Si$_{0.8}$Ge$_{0.2}$ layer with 3×10^{18} cm^{-3} doped layers on *both* sides, separated by 50 Å spacers. Sample growth conditions were identical to that in Sec. 3.2. No special precautions were taken during the growth of either interface, so that, in the absence of any asymme-

|	
p^+ - Si $\sim$ 400 Å	(relaxed)
i- Si spacer $\sim$ 50 Å	(relaxed)
+ + + + + + + 2DHG + + + + + + + $Si_{0.8}Ge_{0.2}$ channel $\sim$ 800Å (strained) + + + + + + + 2DHG + + + + + + +	
i- Si spacer $\sim$ 50 Å	(relaxed)
p^+ - Si $\sim$ 400 Å	(relaxed)
i-Si buffer $\sim$ 1 μm	(relaxed)

(100) n − Si substrate
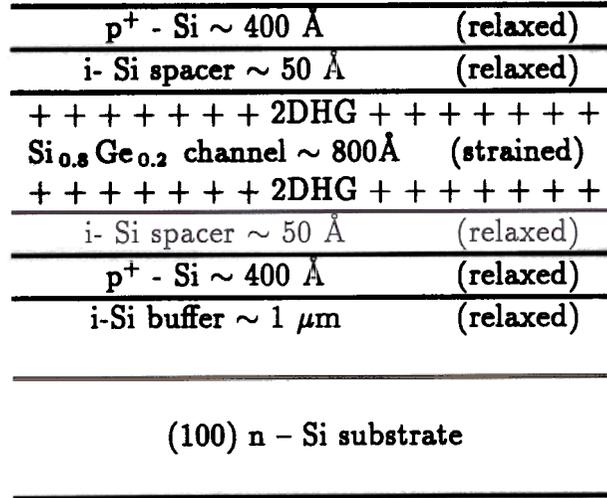
Figure 3.5: Double heterostructure sample #370 with normal and inverted interfaces at which the 2DHG is schematically indicated. The doping level is 3×10^{18} cm^{-3} and the germanium concentration in the channel is 20%.

try introduced by growth chemistry, the structure is symmetric. Fig. 3.6 shows a Secondary Ion Mass Spectroscopy (SIMS) analysis of this sample. The Ge profile in the channel is flat at 20% and abruptly turns on and off with no sign of segregation. The doping level is approximately 3×10^{18} cm^{-3} with a thickness of $\sim$ 300 Å . As expected, SIMS analysis, while confirming the doping levels and layer thicknesses, did not have the resolution to detect any asymmetry. Again we use the more sensitive 2DHG transport as a probe.

3.3.2 Hall and Magnetotransport measurements

Temperature dependent Hall measurements were undertaken on cleaved van der Pauw pieces with evaporated Al contacts, similar to the previous study. The results are displayed in Fig. 3.7. Again, modulation doping characteristics are clearly observed at low temperature, when all parallel conduction in the doped layers is completely frozen out. The higher carrier density 3.2×10^{12} cm^{-2} and lower peak mobility 1000
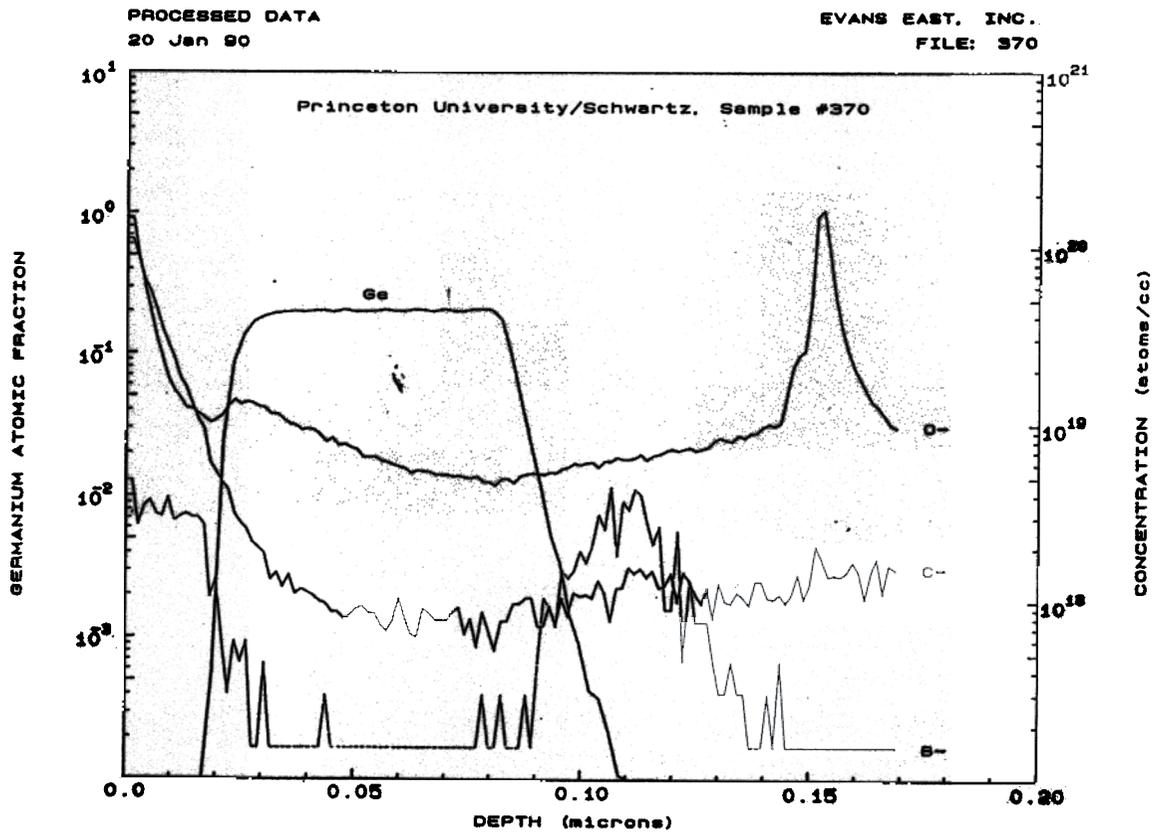
Figure 3.6: SIMS analysis of sample #370 with the symmetric double heterojunction structure. The Ge profile is flat with abrupt turn-on and turn-off profiles. The doping level is $\sim 3 \times 10^{18}$ cm^{-3} . The depth resolution is insufficient to detect the spacer and asymmetry in the doping profile.
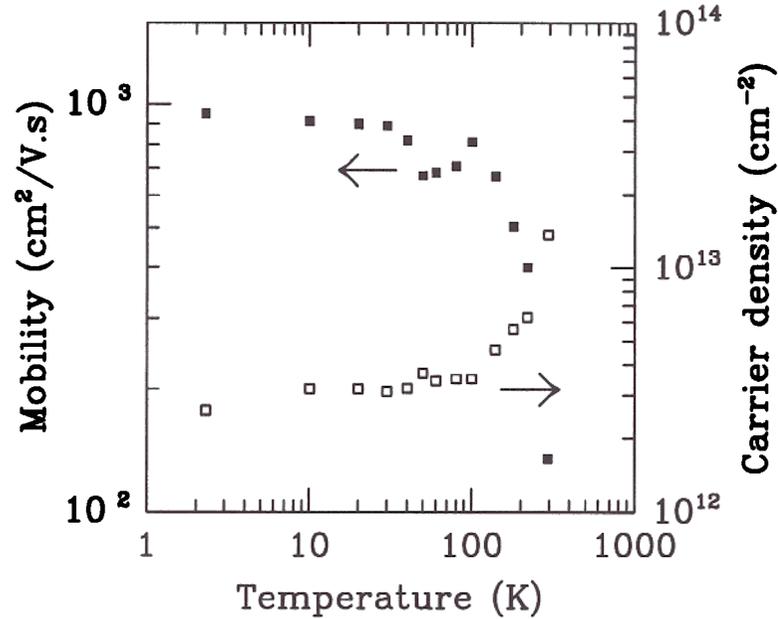
Figure 3.7: Carrier density and mobility as a function of temperature for the Si/SiGe/Si double modulation-doped heterostructure. The feature around 60 K is due to parallel conduction not yet frozen out.

$cm^2/V \cdot s$ compared to the single heterostructures is due to the higher doping, more Ge, and smaller spacer layers. This sample also exhibits magnetoresistance oscillations shown in Fig. 3.8. The presence of a two-dimensional hole gas was also verified by tilting the magnetic field away from the normal. An ideal two-dimensional system responds to the normal component of the magnetic field given by $B \cos \theta$ where θ is the angle measured from the normal. Fig. 3.9 shows the longitudinal magnetoresistance for various angles. As the sample is tilted, the oscillations become weaker and the minima move to higher fields, in order to maintain the same normal component. In the limit of a parallel field ($\theta = 90°$) the oscillations cease. A plot of the minima position as a function of angle shown in Fig. 3.10 confirms that the normal component is maintained constant by the two-dimensional nature of the carrier system.

To take a closer look at the nature of the two interfaces, a fourier analysis is performed on the periodic Shubnikov-deHaas (SdH) oscillations (in the reciprocal
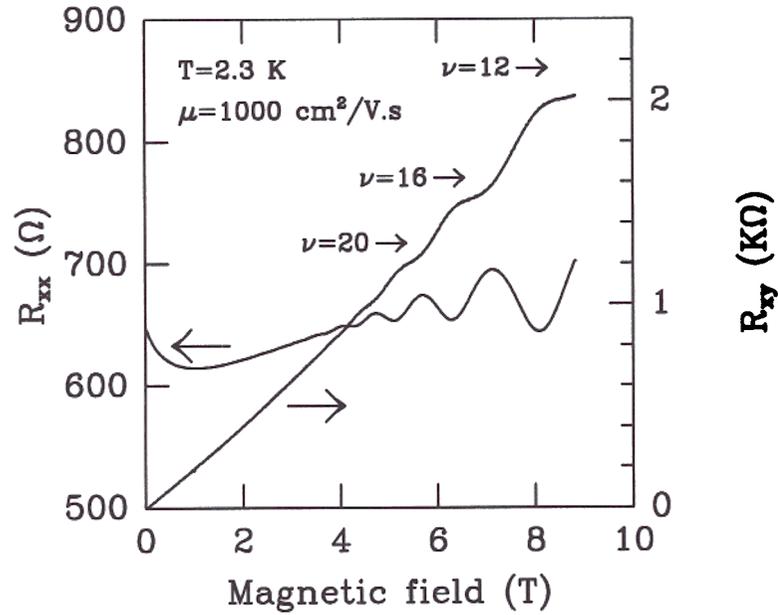
Figure 3.8: Longitudinal resistance (R_{xx}) and Hall resistance (R_{xy}) for the double heterojunction sample as a function of magnetic field. The plateau filling factors, ν, are given by $R_{xy} = h/e^2\nu$.
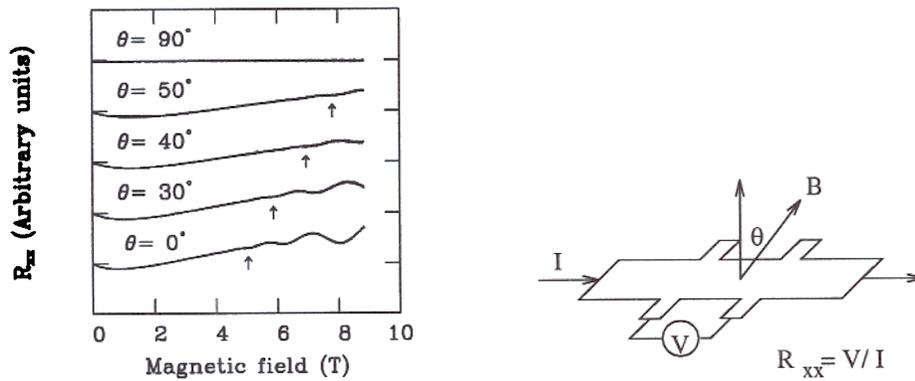
Figure 3.9: Longitudinal magnetoresistance of the double heterostructure in tilted fields. The angle θ is measured away from the normal as shown in the measurement geometry. The position of a particular minimum is tracked by the arrows as it moves to higher fields with larger angles.
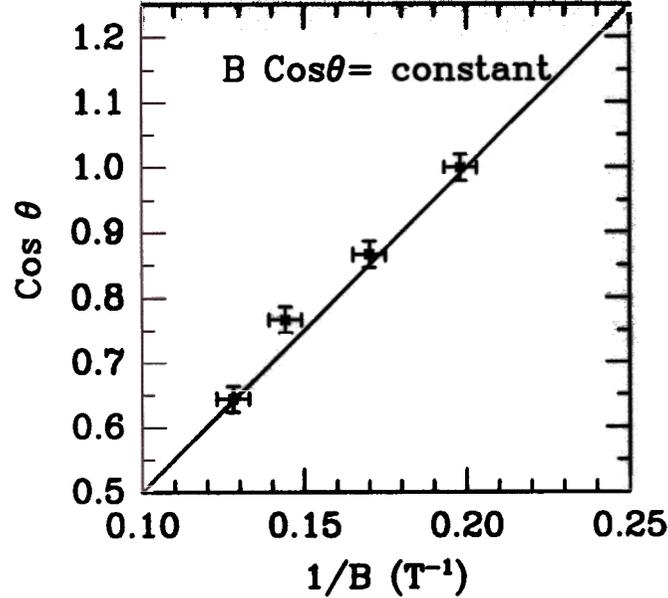
Figure 3.10: Position of the oscillation minima shown by arrows in Fig. 3.9 plotted as a function of angle of the magnetic field to show that the normal component $B \cos \theta$ is constant.

field). The spectrum shown in Fig. 3.11 displays a broad peak whose position corresponds to an SdH carrier density of 1.3×10^{12} cm^{-2} from Eqn. 2.30. On the other hand, the corresponding high-field Hall resistance R_{xy} yields a value 2.6×10^{12} cm^{-2} (Eqn. 2.27) which is *double* the SdH density. This indicates that there are *two* parallel 2DHG channels with similar densities and mobilities at the two interfaces as shown in Fig. 3.5. To rule out any other parallel conduction, *e.g.* in the doped layers, accounting for the difference, we examined the Quantum Hall effect also shown in Fig. 3.8. The quantized plateau values are normally given by $R_{xy} = h/e^2 \nu$ where ν is an even integer for a spin degeneracy $g_s = 2$ and valley degeneracy $g_v = 1$. In our case, however, we find plateaus at approximately $\nu = 12, 16, 20$ but are missing $\nu = 14$ and $\nu = 18$. Equivalently, the quantized resistance is given by $R_{xy} = h/2e^2 \nu$ which is half that of a single two-dimensional system. This unequivocally demonstrates that the double heterostructure sample has two symmetric parallel channels.
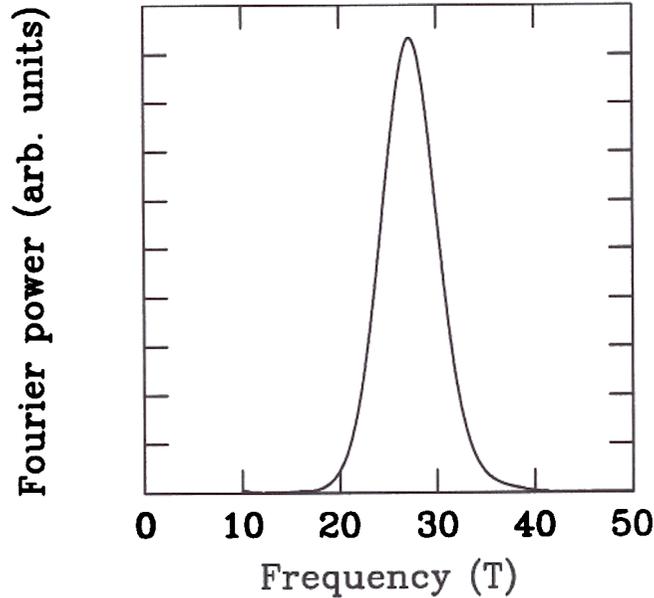
Figure 3.11: Fourier analysis of the periodic Shubnikov-deHaas reciprocal field oscillations of the double heterostructure sample. The position of the broad peak yields a carrier density of 1.3×10^{12} cm^{-2} with an error of 10%.

3.4 Analysis using theoretical models

In order to compare the densities and mobilities we obtained with the theoretical models of Sec. 2.2 and Sec. 2.3, we require relevant material parameters like band-offsets, effective masses etc. We use van de Walle's procedure [10] to calculate the position of the various conduction and valence bands in a strained $Si_{1-x}Ge_x$ alloy on a Si < 100 > substrate. Fig. 3.12 shows the results for the two-fold degenerate conduction minima (2) along the growth direction, the four-fold degenerate minima (4) in the growth plane and the heavy hole (HH), light hole (LH) and split-off (SO) valence bands. As can be seen from the figure, the lowest conduction band edge remains constant with alloying while the highest valence band edge moves almost linearly with alloy concentration. This supports the experimental inference that most of the energy difference appears as a valence band offset for this strain configuration. This valence band edge is almost spherical in nature and we use $m_z = m_{xy} = 0.44 -$
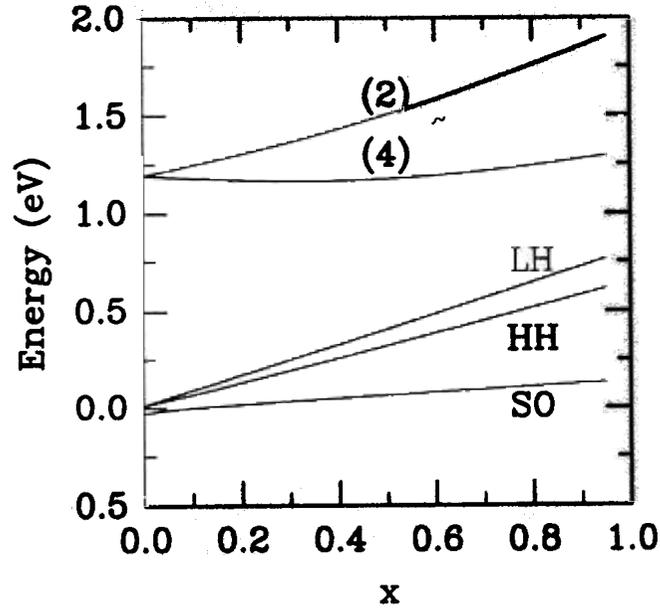
Figure 3.12: Positions of the various conduction and valence band edges in a strained Si$_{1-x}$Ge$_x$ alloy relative to a bulk Si <100> substrate. The calculation procedure is described in ref.[10].

0.42x which fits the experimentally determined cyclotron masses [31] well. Carrier density calculations for the double heterostructure sample ($x = 20\%, \Delta E_V \sim 170$ meV) are shown in Fig. 3.13 as a function of spacer thickness. The depletion charge is assumed to be small and the doping level is also neglected for a heavy doping of 3×10^{18} cm^{-3} . The density is very sensitive to the spacer thickness on the scale of angstroms and justifies the use of the 2DHG as a fine probe. For the nominal spacer of 50 Å , we obtain a theoretical estimate of 1×10^{12} cm^{-2} per interface which compares well with the experimental value of 1.3×10^{12} cm^{-2} . An error of 30% can be attributed to uncertainties in the band-offset and absolute spacer thickness. However the *relative* error in the carrier densities at the two interfaces is smaller. From the width of the peak in Fig. 3.11, we estimate a maximum difference (Δ_{dens}) of 10% in the densities at the normal and inverted interfaces. This corresponds to a *relative* difference in the top and bottom spacers (Δ_{spacer}) of 10 Å . The variation in
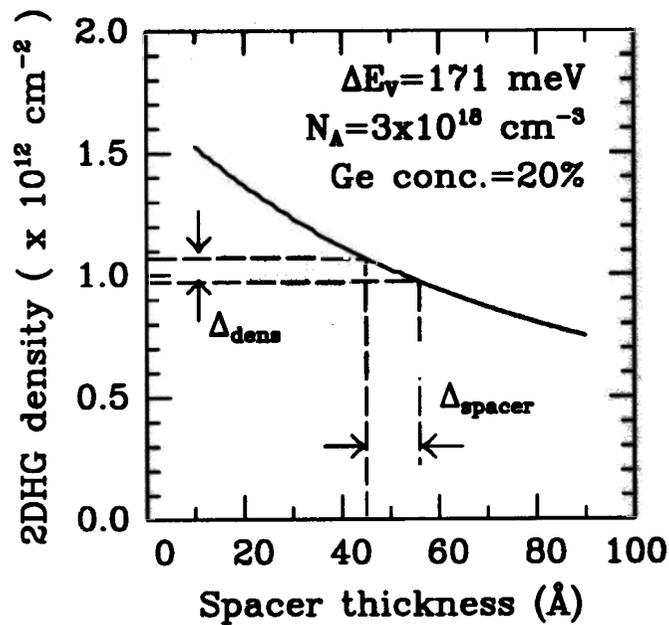
Figure 3.13: Calculation of the equilibrium 2DHG density for a valence band offset of 171 meV (corresponding to 20% Ge) using Eqn. 2.12. A 10% change in carrier density (Δ_{dens}) indicated by the dashed lines corresponds to a variation in spacer thickness (Δ_{spacer}) of 10 Å .
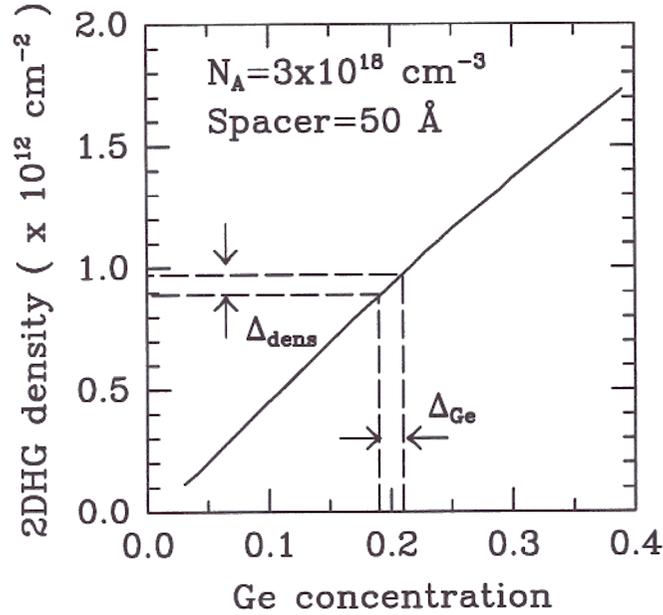
Figure 3.14: Calculation of the equilibrium 2DHG density as a function of Ge content in the channel for a spacer thickness of 50 Å . A 10% change in density (Δ_{dens}) indicated by the dashed lines corresponds to a 10% change in Ge concentration (Δ_{Ge}).

density could also be caused by a change in the Ge concentration from the bottom to the top interface. In Fig. 3.14, we show calculations of the carrier density as a function of the Ge content. A 10% density difference (Δ_{dens}) in this case would correspond to a 10% change (Δ_{Ge}) in the Ge content. For the single heterostructures, using a band-offset of 100 meV ($x = 13\%$) and a spacer of 150 Å , we calculate a density of 3.0×10^{11} cm^{-2} which is 37% below the experimental value.

However, a calculation of the low temperature mobilities using the equations described in Sec. 2.3 yields numbers which are at least an order of magnitude greater than experiment for coulombic mechanisms. For instance, a background doping[1] of 1×10^{16} cm^{-3} limits the mobility to $\sim 50,000$ cm^2/V·s . Similarly remote impurites of density 3×10^{18} cm^{-3} placed at a spacer distance of 50 Å scatter carriers with an equivalent mobility of $\sim 30,000$ cm^2/V·s   The numbers are even higher for lower

---

[1]Typical background doping levels in our samples are in the range 5×10^{15} cm^{-3} – 1×10^{16} cm^{-3} estimated from C-V measurements. See Chap. 4.4 for details.

background doping and larger spacers. Clearly, these low mobility samples are not limited by coulombic scattering.

3.5   Discussion

From the previous section, a spacer thickness change of 10 Å is inferred between the normal and inverted interfaces of the double heterostructure. This means that any boron segregation or autodoping effect during growth is of this order. Similar symmetric interface results have been obtained with the UHV-CVD growth technique at 550°C [28], but not with MBE [29, 27]. In UHV-CVD conditions, under the hydrogen desorption temperature of $\sim$ 550°C , the surface is hydrogen covered [32], which may explain the lack of segregation and presence of symmetric interfaces. Our silicon growth temperature of 700°C , when the boron interfaces were formed, is well over the hydrogen desorption temperature of $\sim$ 550°C [32]. This suggests that the high H_2 overpressure (6 Torr) or the chlorine from the dichlorosilane source in our reactor is playing an important role in the surface chemistry and resulting segregation. The gas switching times in our reactor impose a lower limit on our interface widths. The average switching time for GeH_4 from 0 sccm to 100 sccm ($\sim$ 20% Ge) at 625°C was estimated to be about 5 seconds [33]. Using an average growth rate of 60 Å/min, this corresponds to a minimum interfacial width of 5 Å .The interface width was also measured to be 3–5 Å at a growth temperature and pressure of 625°C and 6 torr respectively using X-ray reflectivity spectra [33]. These numbers agree well with the upper limit of 10Å estimated from the magnetotransport analysis in Sec. 3.4.

Although the carrier densities agree reasonably with theory, the identification of the mobility-limiting mechanism in these structures has been controversial. Previous workers suggested that the low mobility is due to background impurities [29], remote impurities [28] or interface charge [34]. As mentioned before quantitative models

incorporating these mechanisms predict much higher mobilities. Furthermore in a seven-year period of research in this area using a variety of growth techniques, the peak hole mobility for Ge concentrations between 10% and 20% has increased only marginally from 3300 $cm^2/V \cdot s$ in 1984 [27] to 7000 $cm^2/V \cdot s$ in 1991 [30]. In contrast, as described in the next chapter, peak *electron* mobilites in *pure* silicon channels has exceeded 200,000 $cm^2/V \cdot s$ in just two years. These are strong indications of a more fundamental scattering mechanism present in $Si_{1-x}Ge_x$ alloy channels. It has been suggested that alloy scattering (described in Chap. 2.3) is the dominant mechanism responsible for the low hole mobilites [35]. A quantitative estimate is however impossible at this stage because the strength of the scattering potential is unknown. The final part of this thesis explores this question in detail and quantitatively accounts for the maximum hole mobilities observed in these samples thus far.

Two-dimensional Electron Gases on relaxed SiGe buffer layers

4.1 Introduction

In the previous chapter, p-type modulation doped structures grown commensurate on Si substrates were extensively studied. As mentioned before, the insufficient conduction band offset for this strain configuration precludes the use of these structures for n-type modulation doping experiments. However, soon after the first demonstration of two-dimensional hole gases, Abstreiter *et. al.* [19] reported successful modulation doping of electrons using a special strain-symmetrized structure. Subsequent experimental and theoretical investigations confirmed that the band alignments depend crucially on the strain configuration and it was indeed possible to engineer a structure with a significant conduction band offset. As shown in Fig. 2.5 such a structure would require a larger lattice constant than that of Si, such as a relaxed SiGe alloy. A novel SiGe substrate, even if available, would negate the potential viability of these structures in the Si VLSI environment. Instead, the initial experiment [19] used a thick $Si_{0.75}Ge_{0.25}$ buffer layer directly on a silicon substrate, with no attempt at minimizing the misfit and threading dislocations generated by the lattice mismatch. The peak mobility of the two-dimensional electron gas at low temperature obtained in this structure was ~ 2500 $cm^2/V \cdot s$ and there was little progress over the next

43

five years. Renewed interest in these structures in 1991 began with record mobilities exceeding 100,000 cm^2/V·s obtained by minimizing the dislocation density in the active layers using an optimized, graded Si$_{1-x}$Ge$_x$ relaxed buffer on a conventional silicon substrate [35, 36, 37]. Fig. 4.1 shows a generic high mobility structure of this type and the associated band diagram. The band alignment is type-II and the electrons transfer from the relaxed SiGe supply layer into the *tensilely* strained Si channel. From Fig. 2.5 we know that the strain splits the six-fold degeneracy in the conduction band of the Si channel. Because the electrons selectively populate the lower energy two-fold degenerate bands with light in-plane mass and lower intervalley scattering, the room temperature mobility in these structures (~ 2000 cm^2/V·s) is higher than bulk undoped silicon [38]. This promises higher speeds for MODFETs utilizing this structure compared to conventional Si MOSFETs.

From a device point of view, both the carrier density and mobility are equally important parameters. In MODFETs, the spacer thickness is often used to trade-off density with mobility, with smaller spacers resulting in higher densities and lower mobilities. In order to increase the carrier density in the channel without sacrificing mobility, it becomes necessary to achieve maximum conduction band offset. For the strained Si/Si$_{1-x}$Ge$_x$ relaxed structures this translates to high germanium content buffers and maximum strain in the channel. All the results reported so far have been achieved with MBE or UHV-CVD growth techniques. While high mobilities have been obtained, the carrier densities have remained below 1×10^{12} cm^{-2} , with correspondingly high sheet resistivities. In this work we demonstrate n-type modulation doped structures grown by RTCVD for the first time. By increasing the Ge content in the buffer and optimizing the spacer width, we achieved a record carrier density of 2.7×10^{12} cm^{-2} in the channel. At the same time the mobilities at 77 K ($\sim 15,000$ cm^2/V·s) and 4.2 K ($\sim 32,000$ cm^2/V·s) remained fairly high. The corresponding sheet resistivities, 150 $\Omega/\square$ at 77 K and 69 $\Omega/\square$ at 4.2 K, are the lowest reported so

n$^+$-SiGe supply layer   (relaxed)

i - SiGe spacer   (relaxed)

i - Si channel   (strained)

Graded SiGe buffer (relaxed)

Si substrate

Ionized
Dopants

Spacer

2DEG

E_C

$+++$

Relaxed
buffer

Tensilely strained Si channel
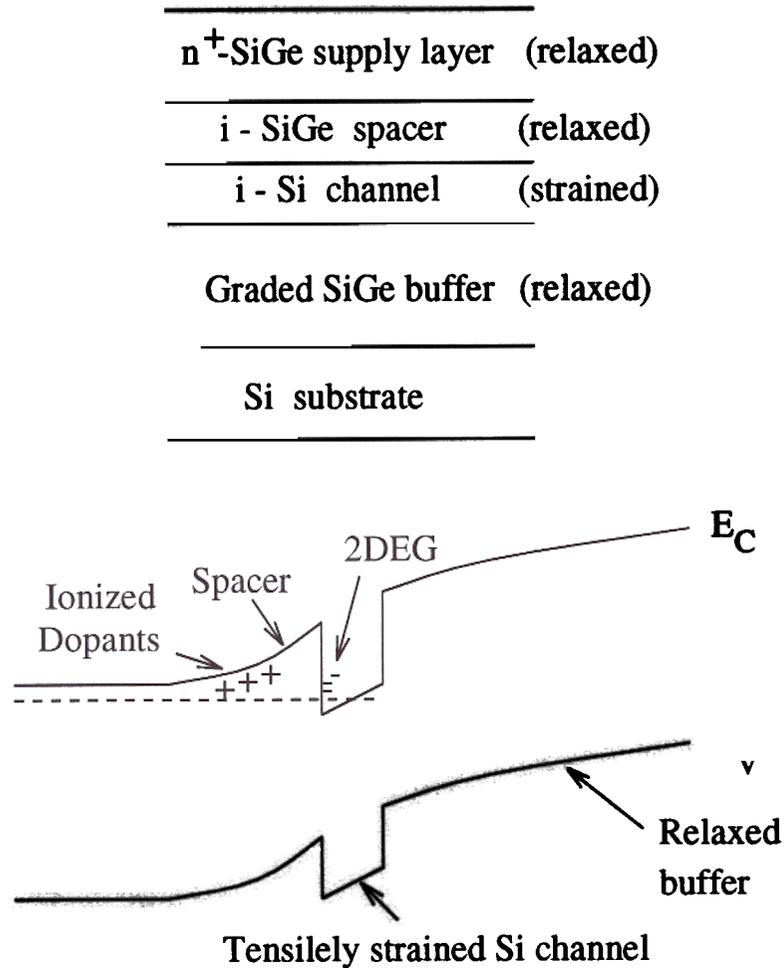
Figure 4.1: A typical high mobility n-type modulation doped structure on a relaxed graded SiGe buffer layer. The associated band diagram shows a type-II alignment with the electrons in the tensilely strained Si channel.

far at those temperatures.

4.2   Growth and characterization of relaxed buffers

In aforementioned studies [39, 36], the relaxed buffers were grown by MBE at relatively high temperatures of 750–900°C . Using UHV-CVD at 550°C , the IBM group also demonstrated successful high mobility structures [37]. Although low temperatures are generally preferable, unless very thick buffers are used, there is a possibility of incomplete relaxation. This could lead to smaller band offsets and less carrier transfer. In this work, we used RTCVD to grow the buffer at 625°C   We did not grow the buffer at higher temperatures for the following reasons:

1 The growth rate and Ge incorporation at higher temperatures was not well characterized. Accurate knowledge of these parameters is important for proper grading of the buffer.

2. A few SiGe layers grown at 800°C indicated rough morphology and reduced Ge incorporation.

3. At high temperatures there is substantial deposition on the quartz tube, thus requiring frequent venting of the system and replacement of the tube.

The Ge content was graded linearly from 0% at the substrate to 38% at the top over a distance of $\sim$ 0.5 μm followed by a uniform $\sim$ 1 μm Si$_{0.62}$ Ge$_{0.38}$ cap layer. The structure was then annealed in situ for 1 hour at 800°C to induce complete relaxation as confirmed by X-ray diffraction analysis . The development and characterization of the relaxed buffers were performed by Chee-wee Liu of our laboratory.

The dislocation density was characterized by plan-view and cross-sectional TEM and Electron Beam Induced Current (EBIC). Plan view images of a typical relaxed buffer performed by J.T. McGinn of the David Sarnoff Research Center are shown in

Fig. 4.2. The upper image shows threading dislocation segments generated by inter-action of misfit dislocations due to the lattice mismatch, with an average separation of 5 μm   This indicates a lower limit of 4×10^6 cm^{-2} for the dislocation density. We also see some areas on the wafer with densely packed "bands" of threading dis-locations as shown in the lower image. The origin of these bands is unclear at this point. Fig. 4.3 shows a cross-sectional TEM image of the same buffer. The dense network of misfit dislocations are confined close to the substrate interface. In fact, some of them penetrate into the substrate contrary to intuition. The misfit disloca-tions normally interact to produce threading segments which penetrate all the active layers and terminate at the surface. Such threading dislocations can cause excessive recombination and reduce mobility through carrier trapping. However, as seen from Fig. 4.3, the uppermost layers appear largely dislocation-free. The grading of the Ge content presumably suppresses interaction of misfits and formation of threading dis-locations. The nucleation and propagation of dislocations in these type of structures has been extensively investigated by Fitzgerald $et.$ $al.$ [39] and LeGoues $et.$ $al.$ [40].

While TEM is an excellent high-resolution imaging technique, it is ill-suited for quantitative estimations of low dislocation densities. Large area imaging methods like defect etching or Electron Beam Induced Current (EBIC) are better for this purpose. We have used both these techniques to characterize our relaxed buffers quantitatively. An EBIC image of one of our relaxed buffers is shown in Fig. 4.4. The typical threading dislocation density is determined to be $\sim 3 \times 10^7$ cm^{-2} which is slightly higher than the best results of $\sim (10^6 - 10^7$ cm$^{-2})$ obtained by other groups [35, 36]. The slightly higher number in our case may be due to the larger Ge content and steeper grading rate ($\sim 80\%/\mu$m). In contrast, the AT&T group [35] used a 30% buffer graded over 3 μm ($\sim 10\%/\mu$m), the Daimler-Benz group [36] used 30% buffers graded over 0.65–3.0 μm ($\sim 10 - 45\%/\mu$m) and the IBM group [37] used 30% step-wise graded multilayers over 0.5 μm ($\sim 60\%/\mu$m). As we shall see below, this
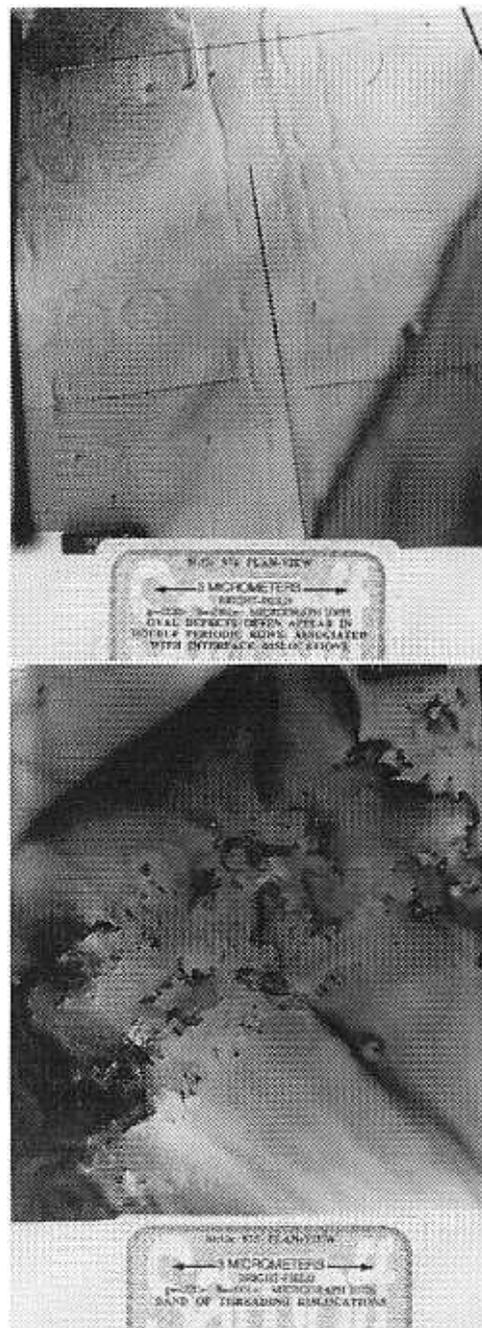
Figure 4.2: Plan-view TEM images of a typical relaxed buffer. The upper image shows threading segments with an average separation of 5 μm , equivalent to a lower limit of 4×10^6 cm^{-2} for the dislocation density. The lower image shows an area of the wafer with a "band" of densely packed threading dislocations.
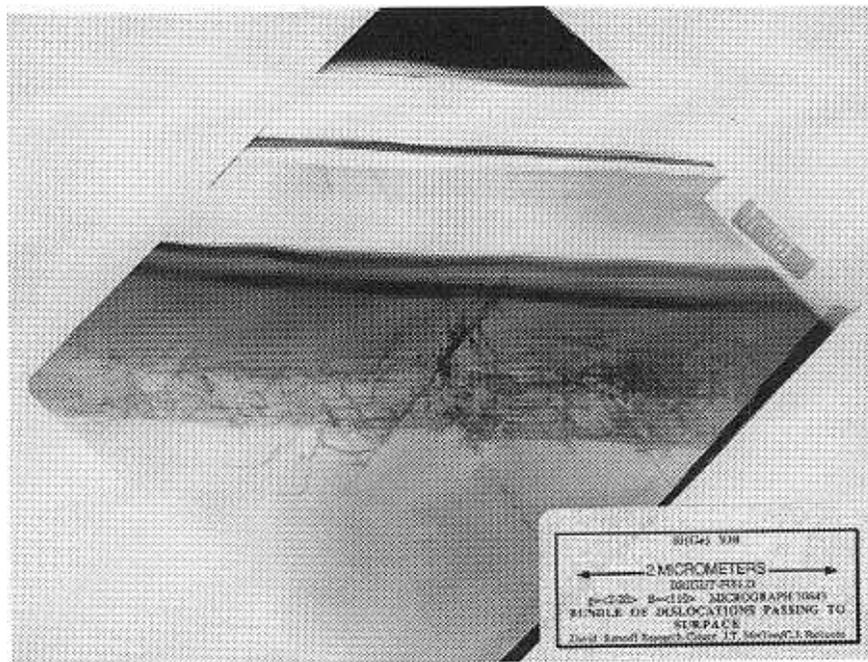
Figure 4.3: Cross-sectional TEM image of a typical relaxed buffer. Note that the dense network of misfit dislocations is confined at the substrate interface and the active layers on the top are relatively dislocation-free.
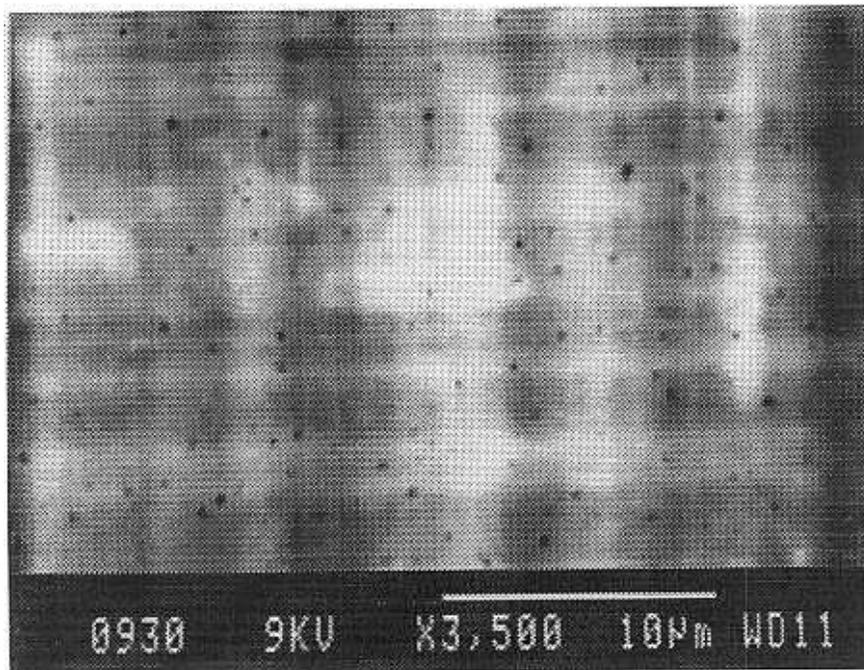
Figure 4.4: Electron Beam Induced Current (EBIC) image of relaxed buffer sample (# 930). Assuming the dark spots are threading dislocations, we infer a dislocation density of $\sim 3 \times 10^7$ cm^{-2} . This image was obtained by Chee-wee Liu.

does not adversely affect the electrical results.

4.3   Modulation doping studies

4.3.1   Sample structure

For the modulation doping experiments, the relaxed buffer was grown on a p-type Si $< 100 >$ substrate as described above, followed by the active layers consisting of an undoped strained 75 Å Si channel, an undoped $Si_{0.62}Ge_{0.38}$ relaxed spacer and a heavy $\sim 2 \times 10^{19}$ cm^{-3} phosphorous-doped 300 Å $Si_{0.62}Ge_{0.38}$ supply layer (see Fig. 4.5). The doping was estimated from spreading resistance profiling described in the next chapter. Growth conditions in our reactor were identical to those described in Sec. 3.2. The channel was kept below the critical thickness to avoid any relaxation during growth and subsequent processing. The spacer thickness was varied from 0 to 80 Å to study its effect on the carrier density and mobility. A number of initial growth runs failed to show good modulation doping behaviour due to a variety of reasons involving sample structure, contact procedures etc. A description of these samples along with probable causes of failure is given in Appendix B to aid novice experimentalists in this area. Also listed in this appendix are the growth conditions of the successful samples (#893, #907, #910 and #922) described in the next few sections. In Appendix D we list the entire computer-controlled growth sequence code for sample #910.

4.3.2   Hall measurements

As before, both Van der Pauw and Hall bar geometries were used for temperature dependent Hall measurements. A number of contact procedures were initially studied using a variety of different metals and alloys. Finally a gold-antimony (Au:Sb) alloy (0.6–24 % antimony) annealed in the range 350–400°C for 10–15 min yielded excellent

$n^+ - Si_{0.62}Ge_{0.38} \sim 300$ Å   (relaxed)

$i - Si_{0.62}Ge_{0.38}$ spacer   (relaxed)

$------2DEG--------$

$i - Si$ channel ~ 75 Å   (strained)

$i - Si_{0.62}Ge_{0.38} \sim 1 \mu m$ (relaxed)

graded SiGe buffer $0 - 38\% \sim 0.5 \mu m$

$<100> p - Si$ substrate
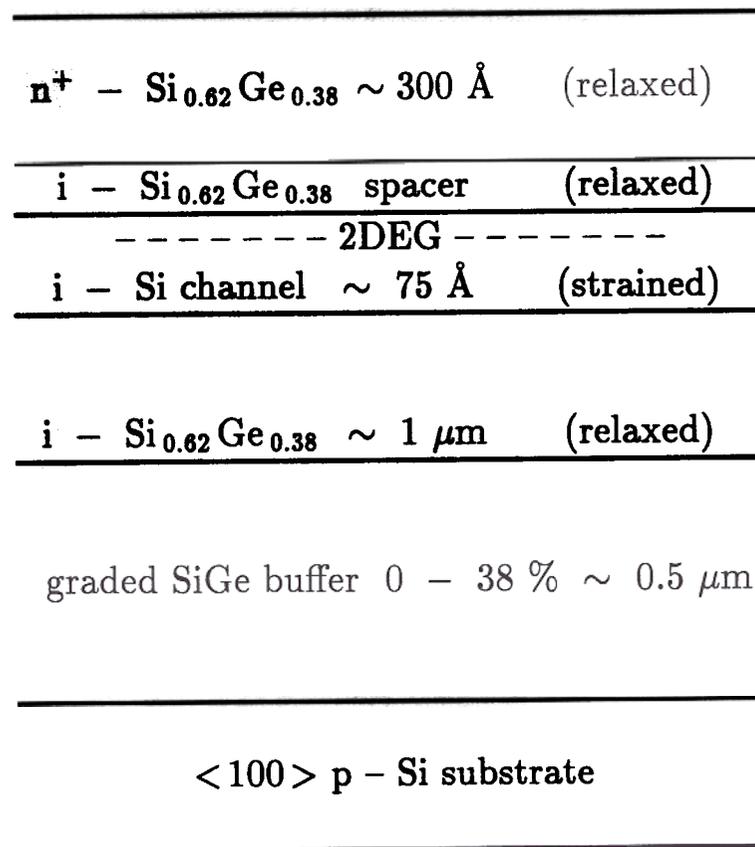
Figure 4.5: Sample structure for the modulation doping experiments. The doping level was $\sim 2 \times 10^{19}$ cm^{-3} and the spacer was varied from 0 to 80 Å .
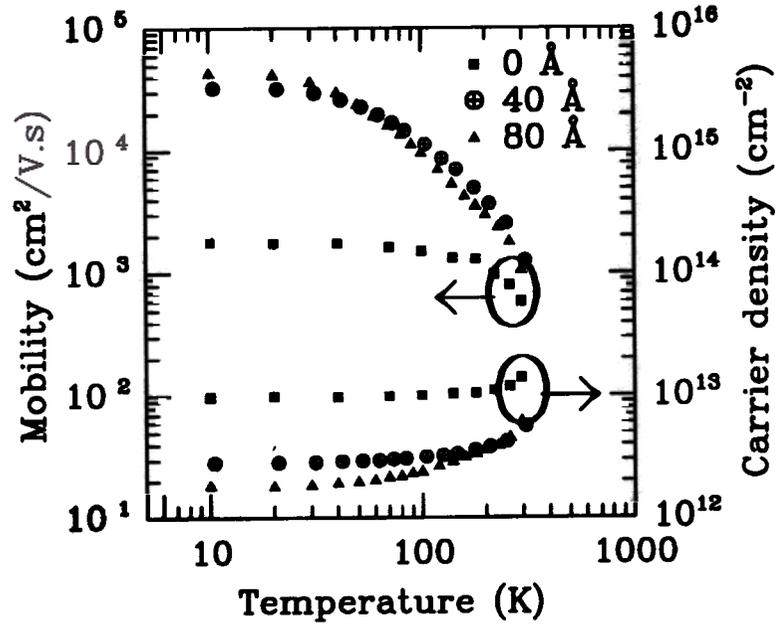
Figure 4.6: Carrier density and Hall mobility as a function of temperature for three n-type modulation doped samples with spacer thicknesses 0 (sample #893), 40 (sample #907) and 80 Å (sample #910). All samples had the structure shown in Fig. 4.5.

n-type ohmic contacts down to (at least) 0.6 K. Typically, after sample degrease and native oxide removal, 50 Å Au/ 1500 Å Au:Sb/1000 Å Au was thermally evaporated and lifted off to form the contact pads. The initial Au layer is crucial because the antimony, which evaporates first, has poor adhesion properties to the SiGe cap. This contact recipe was successful for most samples (see Appendix B for exceptions).

Samples were mounted in a closed cycle helium refrigerator with gold wires bonded to the contact pads for external electrical measurements. The results of Hall measurements are displayed in Fig. 4.6 for three samples with the same structure shown in Fig. 4.5 but different spacer widths. The mobility increases monotonically for all the samples with decreasing temperature, reaching 42,000 $cm^2/V \cdot s$ at 10 K for a spacer width of 80 Å . At the same time, the carrier density decreases and saturates at low temperatures with no evidence of freeze-out. This indicates the presence of a degenerate 2D electron gas (2DEG) confined in the strained silicon channel. At 77 K

for a 40 Å spacer, a 2DEG of density 2.9×10^{12} cm^{-2} and mobility 15,000 cm^2/V·s was obtained, which yields a sheet resistivity of 140 $\Omega/\square$, the lowest yet reported in these type of structures. The results for four samples are summarised in Table. 4.1. The clear trends of increasing low temperature mobility and decreasing carrier density with spacer thickness are expected and will be discussed later.

4.3.3   Magnetotransport

To further study the quality of the 2D electron system, we did magnetotransport experiments in a He3 superconducting magnet system. The as-grown samples showed pronounced deviation from linearity in the transverse Hall resistance (R_{xy}) at high magnetic fields. This was found to be due to parallel conduction arising from the heavily doped cap layer. If part of the cap layer is etched in a 40:20:5 HNO$_3$: H$_2$O : HF solution [41] , the parallel conduction can be removed and the Hall resistance stays linear in the high field regime. Fig. 4.7 shows the Hall resistance of the 40 Å spacer sample as a function of magnetic field for different etching times of the cap layer. After two minutes of etching, most of the parallel conduction is removed and the Quantized Hall plateaus become well defined. It should be noted that for small magnetic fields, the slope of R_{xy} remains the same (within 15%) as a function of etching time. Thus, the data of Fig. 4.6 which was taken using magnetic fields less that 0.2 T, is largely unaffected by the parallel conduction found at higher fields.

Fig. 4.8 shows both the longitudinal resistance (R_{xx}) and the transverse resistance of the 40 Å spacer sample after the cap layer etch. Well-defined Shubnikov-deHaas oscillations and quantized Hall plateaus are observed as a function of the magnetic field. Beyond 3 T where the plateaus begin to appear, the quantized values are given by $R_{xy} = h/4e^2$ indicating a total degeneracy of 4. Assuming that the Zeeman splitting is not resolved at these fields ($g_s = 2$), we obtain a valley degeneracy of two ($g_v = 2$). This result can be understood if we consider the conduction band of the

T=10 K

Sample #	Spacer Å	Density cm^{-2}	Mobility cm^2/V·s	Resistivity Ω/□	Parallel conduction ?
893	0	9.7×10^{12}	1770	363	yes
907	40	2.7×10^{12}	32350	69	no
922	75	1.7×10^{12}	38200	98	no
910	80	1.6×10^{12}	44000	83	no

T=77 K

Sample #	Spacer Å	Density cm^{-2}	Mobility cm^2/V·s	Resistivity Ω/□	Parallel conduction ?
893	0	9.9×10^{12}	1620	388	yes
907	40	3.0×10^{12}	14470	142	yes
922	75	2.6×10^{12}	8530	276	yes
910	80	2.2×10^{12}	13460	210	yes

T=300 K

Sample #	Spacer Å	Density cm^{-2}	Mobility cm^2/V·s	Resistivity Ω/□	Parallel conduction ?
893	0	1.4×10^{13}	590	754	yes
907	40	5.6×10^{12}	1250	891	yes
922	75	1.3×10^{13}	670	700	yes
910	80	6.0×10^{12}	1050	987	yes

Table 4.1: Summary of Hall data at 10 K, 77 K and 300 K for four n-type modulation doped structures with different spacer widths. The temperature dependent measurements for sample #922, whose channel was grown using silane (SiH$_4$) instead of dichlorosilane (SiCl$_2$H$_2$), are omitted in Fig. 4.6 for clarity. The last column in the table indicates whether the data included possible parallel conduction in the doped layers at that temperature.
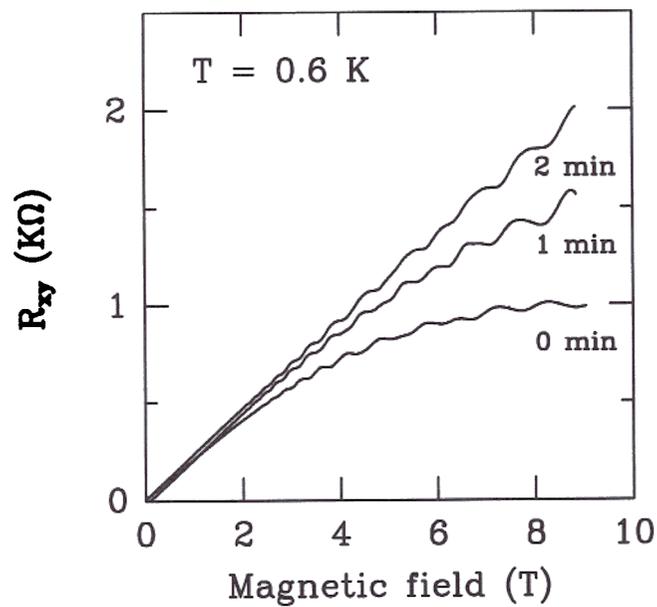
Figure 4.7: Transverse Hall resistance (R_{xy}) of the 40 Å spacer sample in high magnetic fields as a function of etching time of the cap layer. As the parallel conduction is removed by etching, the resistance becomes more linear and the plateaus become well quantized.
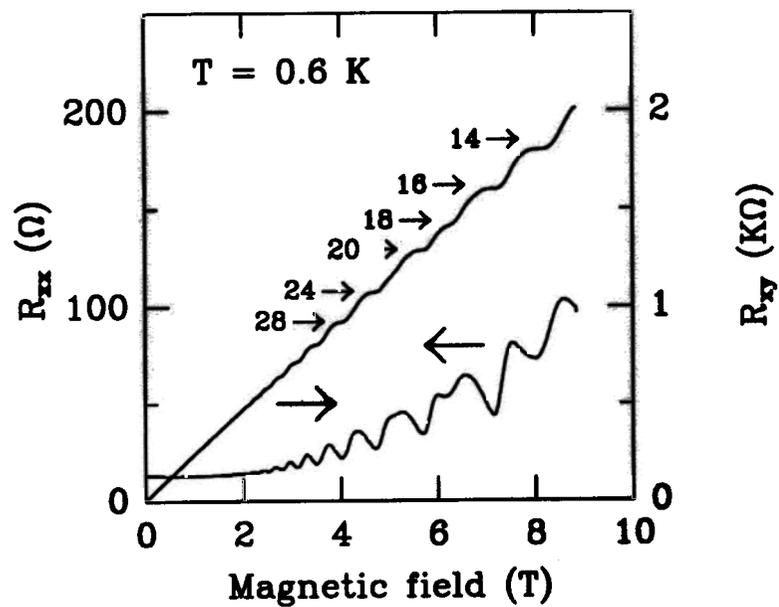
Figure 4.8: Longitudinal and transverse resistance of the 40 Å spacer modulation doped sample as a function of magnetic field at 0.6 K. The cap layer was etched to prevent parallel conduction. The plateau filling factors ν are given by $R_{xy} = h/e^2\nu$ where e is the electron charge and h is the Planck constant.
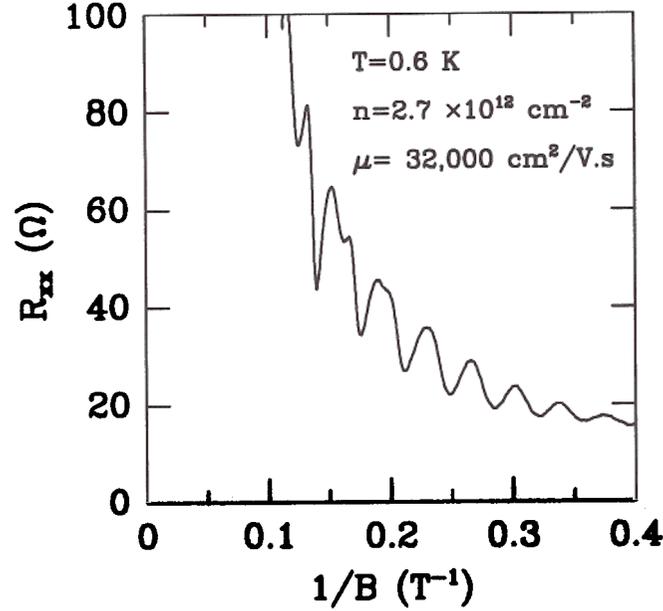
Figure 4.9: Shubnikov-deHaas oscillations in the longitudinal resistance shown in Fig. 4.8 plotted as a function of the inverse magnetic field. Note the single period corresponding to a carrier density of 2.7×10^{12} cm^{-2} at low fields. At higher magnetic fields, the frequency doubles due to spin splitting of the Landau levels.

strained silicon channel. As depicted in Fig. 2.5, the biaxial tensile strain splits the six-fold degeneracy and lowers the two valleys along the growth direction in energy. The electrons in the ground state have therefore a heavy longitudinal mass (~ 0.98 m_0) and a lighter transverse mass (~ 0.19 m_0). A transverse cyclotron mass close to silicon bulk values has been measured for the 2DEG in similar structures [17]. oscillations are periodic in the reciprocal field as displayed in Fig. 4.9, corresponding to a carrier density of 2.7×10^{12} cm^{-2} which agrees very well with the temperature dependent Hall measurements. The single oscillation period indicates that only the lowest quantized level is occupied, even at these high carrier densities. At higher magnetic fields, the oscillation frequency doubles with the spin splitting of the Landau levels being resolved. At the same time the Hall plateaus increase consecutively by even integers corresponding to the two-valley degeneracy. These observations
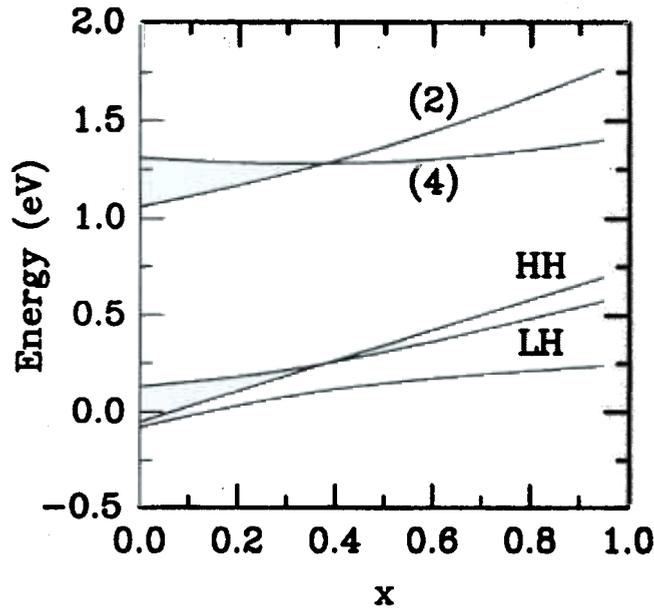
Figure 4.10: Positions of the various conduction and valence band edges in a strained $Si_{1-x}Ge_x$ alloy relative to a relaxed $Si_{0.62}Ge_{0.38}$ substrate. The calculation procedure is described in Ref. [10].

have also been reported previously by other workers [35, 36, 37]. Magnetotransport experiments on the 80 Å spacer sample yielded essentially the same results, although the filling factors were smaller due to the lower density.

4.4   Comparison with theoretical models

To estimate the band offset for the structure shown in Fig. 4.5, we repeat the calculations of Fig. 3.12 using a $Si_{0.62}Ge_{0.38}$ relaxed substrate and a strained $Si_{1-x}Ge_x$ channel. The results are displayed in Fig. 4.10. As before, the conduction bands are denoted by two-valley (2) and four-valley (4) degeneracies while the three valence bands are designated as light hole (LH), heavy hole (HH) and split-off bands (SO), using the standard notation for the unstrained case. For the pure Si channel, a type-II alignment is obtained as shown in Fig. 2.5, with a conduction band offset of $\sim$ 240
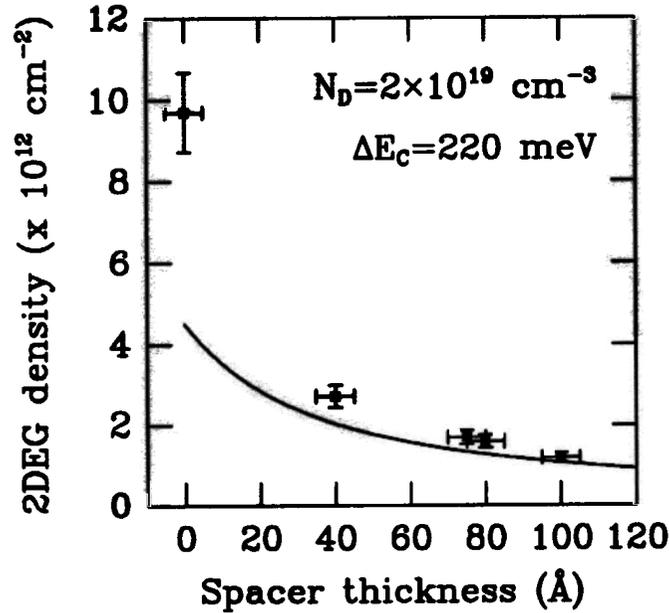
Figure 4.11: Calculated equilibrium carrier density for the n-type modulation doped structure as a function of spacer width. The solid squares are experimental data from Table. 4.1. The data point for 100 Å spacer is from sample #1344 described in Chap. 6.

meV. Using this value in Eqn. 2.12, we can estimate the equilibrium carrier density in our samples. The equations in Chap. 2.2 were strictly derived for a single-sided *infinite* heterojunction. In our case, we have a 75 Å finite quantum well as the channel. Since the extent of the wavefunction for high carrier densities ($\sim$ 30–50 Å) is less than the quantum well width, the assumption of an infinite heterojunction does not introduce any significant error. For the calculations the dopant ionization energies is neglected in the heavy doping case and the depletion charge is assumed to be much lower than the 2DEG density. The results are displayed in Fig. 4.11 along with the measured data from Table. 4.1. The agreement is fair, with the measured densities 20–30% higher due to the parallel conduction in the cap and uncertainties in the calculation parameters.

For mobility calculations at low temperature, only coulombic mechanisms need
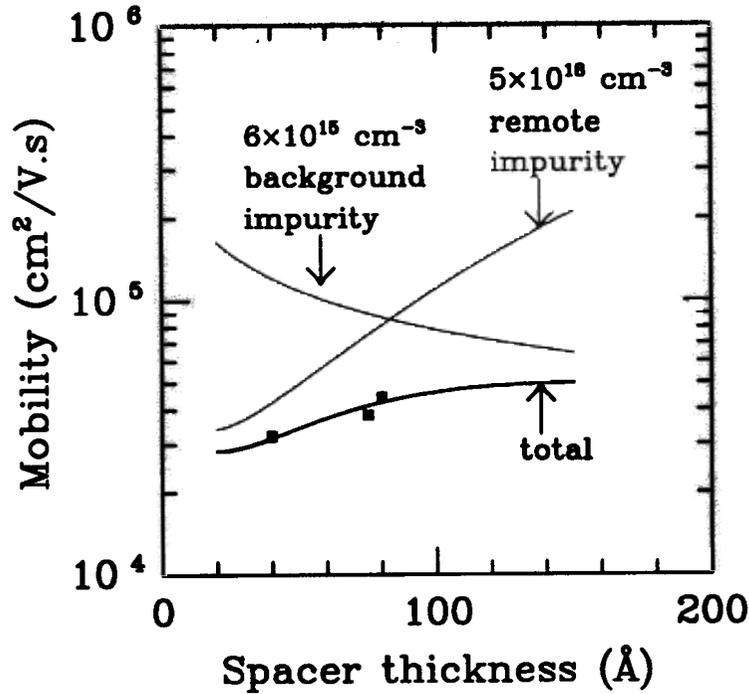
Figure 4.12: Calculated mobility due to background (μ_{BI}) and remote (μ_{RI}) impurities in the 2DEG samples. The total mobility, μ_{total} is given by $\mu_{total}^{-1} = \mu_{BI}^{-1} + \mu_{RI}^{-1}$. The solid squares are experimental data from Table. 4.1.

be considered since there is no alloy scattering in the pure silicon channel. Using the equations developed in Sec. 2.3, we calculate the mobilities due to background impurities (μ_{BI}) and remote impurities (μ_{RI}) in our structure. While μ_{RI} increases strongly with spacer thickness, μ_{BI} decreases because the decreasing carrier density leads to loss of screening of the impurities. The best fit to the experimental data is obtained for a background impurity density of $\sim 6 \times 10^{15}$ cm^{-3} and a remote impurity density of 5×10^{18} cm^{-3} (see Fig. 4.12). The background doping was determined independently using capacitance-voltage measurements on a test sample (#1096) with a 5000 Å uniform Si$_{0.62}$Ge$_{0.38}$ relaxed buffer and a doped cap structure grown at 625°C . The results are displayed in Fig. 4.13. The background doping in the SiGe layers grown at 625°C is approximately 1×10^{16} cm^{-3} and consistent (within a factor of 2) with the mobility calculations of Fig. 4.12.
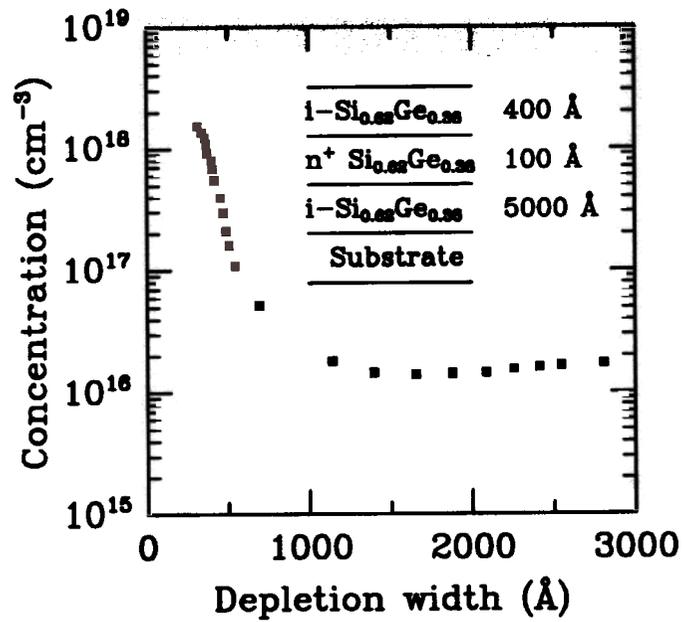
Figure 4.13: Carrier concentration as a function of depletion width from CV measurements on sample #1096. The sample structure is shown in the inset and an Aluminum Schottky barrier with area 4×10^{-4} cm^2 was used. The doped ($\sim 3 \times 10^{18}$ cm^{-3}) cap structure is necessary to prevent complete depletion of the epitaxial layer at zero bias. The background doping in the SiGe layers is $\sim 1 \times 10^{16}$ cm^{-3} .

Discussion

For the same spacer thickness, the carrier densities in our samples are consistently higher than those obtained in Refs. [35, 36]. This is mainly due to the higher germanium content, and therefore, a larger conduction band offset. We see only lowest subband occupation in our 75 Å quantum well (QW) with 2.7×10^{12} cm^{-2} carriers. In contrast, Schuberth et. al. [17] reported occupancy of the second quantized level at a density of 2×10^{12} cm^{-2} in a 75 Å well from two oscillation periods in the Shubnikov-deHaas oscillations. The samples in Ref. [17], however, had a *multiple* QW modulation-doped structure. In such structures, carrier density variations among different wells, rather than occupation of the second level, can give rise to multiple oscillation periods. Furthermore, the large effective masses coupled with a 10–20 meV ambiguity in the band offsets can lead to sizable errors in estimates of the energy levels and the Fermi-level position, which make calculations of the exact number of occupied levels difficult.

Although the mobilities we obtained are comparable to the UHV-CVD results [37], they are lower than the best MBE results [36] by a factor of 4. The calculations in the previous section show that this is consistent with a higher background doping ($\sim 10^{16}$ cm^{-3}) in our reactor. More extensive numerical calculations by Stern et. al. [42] and experiments by the IBM group [43] indicate that the maximum mobilities in samples grown by MBE and UHV-CVD with lower background doping are limited by remote impurity scattering.

For practical FETs, we need high carrier densities and therefore smaller spacers, leading to mobilities limited by remote impurities. Hence background doping is not a limiting factor. For example we have record low sheet resistivities (140 $\Omega/\square$) at 77 K. Encouraged by these numbers, we look at MODFETs utilizing this structure in the next chapter.

Chapter 5

Fabrication and characterization of n-MODFETs

5.1 Introduction

Among the first electronic devices fabricated in the $Si/Si_{1-x}Ge_x$ system was the heterojunction bipolar transistor. This technology has matured sufficiently today with demonstrations of record cut-off frequencies and integrated circuit capabilities [44, 45]. The first p-channel $Si/Si_{1-x}Ge_x$ MOSFETs used the hole confinement in a $Si_{1-x}Ge_x$ quantum well to move the carriers away from the Si-SiO_2 interface, resulting in improved mobilities [46, 47]. The performance of these devices was hampered by restricted gate voltage swings and additional alloy scattering in the $Si_{1-x}Ge_x$ well at low temperatures. With the advent of high mobility 2DEGs on relaxed buffers, the focus shifted towards n-channel MODFETs. Both enhancement and depletion mode devices were initially fabricated with high extrinsic transconductances (~ 300 mS/mm at 300 K and ~ 600 mS/mm at 77 K) but large gate leakage currents at room temperature [48, 49]. The gate leakage was attributed to poor n-type dopant turn-off in the MBE and UHV-CVD systems.

In this chapter, we describe our attempts to fabricate high-performance MOD-FETs utilizing the low sheet resistivity 2DEG structures demonstrated in the previous chapter. In the process, we studied the n-type dopant control in our RTCVD reactor in order to obtain good Schottky gates with low leakage currents. We also

64

developed a simple two-step fabrication process which completely self-aligns the gate to the source and drain contacts

Sample structure

The n-type modulation doped structure shown in Fig. 4.5 of the previous chapter had to be modified to accommodate a Schottky barrier which can control the 2DEG density. The simplest way to accomplish this task was to grow an undoped cap on top of the n^+ supply layer. The residual doping in the cap had to be low enough for minimum gate leakage while the integrated doping in the supply layer had to be reduced to ensure complete depletion by the gate. This design depends crucially on control of the dopant profile in the supply layer

We studied the phosphorous incorporation in our system by spreading resistance measurements on a sample (#1121) with a series of 3000 Å $Si_{0.62}Ge_{0.38}$ epitaxial layers grown at 625°C , doped with different amounts of phosphine (diluted 70 ppm PH_3 in H_2) and separated by 3000 Å undoped regions. Each layer was grown for 10 min and the phosphine source was simply turned on and off to achieve this structure. The results, measured by B. Halleck of the Rockwell Science Center, are shown in Fig. 5.1. The dopant incorporation increases systematically with PH_3 flow (in sccm), exceeding 2×10^{19} cm^{-3} for 100 sccm. The width of the profiles remains approximately constant and hence, to first order, we can assume that the growth rate does not depend on the phosphine flow. From simple model calculations, we estimated that a 100 Å 4×10^{18} cm^{-3} doped layer can be completely depleted by a Schottky gate using a reverse bias of less than 2 V. The corresponding phosphine flow required is therefore 5 sccm. The spreading resistance technique does not, however, have the required depth resolution to determine turn-on and turn-off profiles accurately. We used capacitance-voltage (CV) measurements on a test structure (sample #1287) for
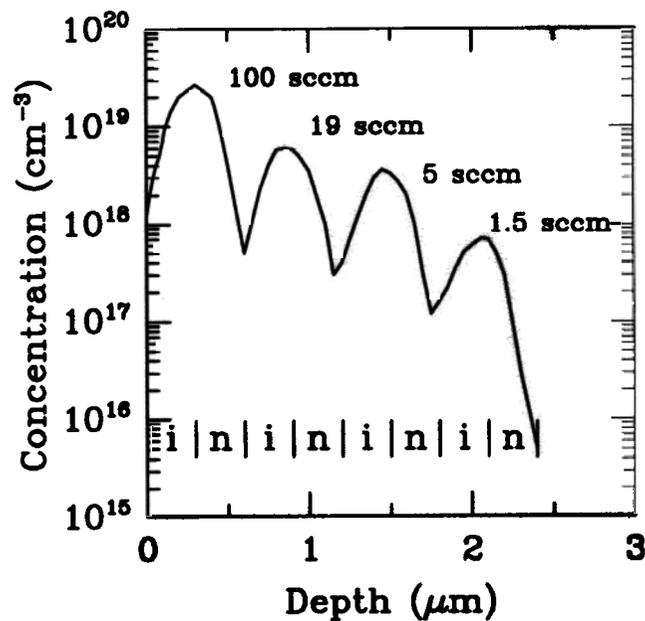
Figure 5.1: Phosphorous concentration as a function of depth for sample #1121. The structure consisted of different amounts of dopants separated by undoped regions, achieved by simply turning on and off a diluted phosphine (70 ppm in H_2) source in intervals of 10 min at 625°C . The 26 sccm DCS flow and 450 sccm of 0.8% GeH_4 in H_2 along with a 3 lpm H_2 carrier were kept constant throughout the structure. The Ge content is $\sim$ 38% and the growth rate ($\sim$ 300 Å/min) does not depend on the phosphine flow. The data was measured by spreading resistance profiling technique assuming bulk silicon mobility.
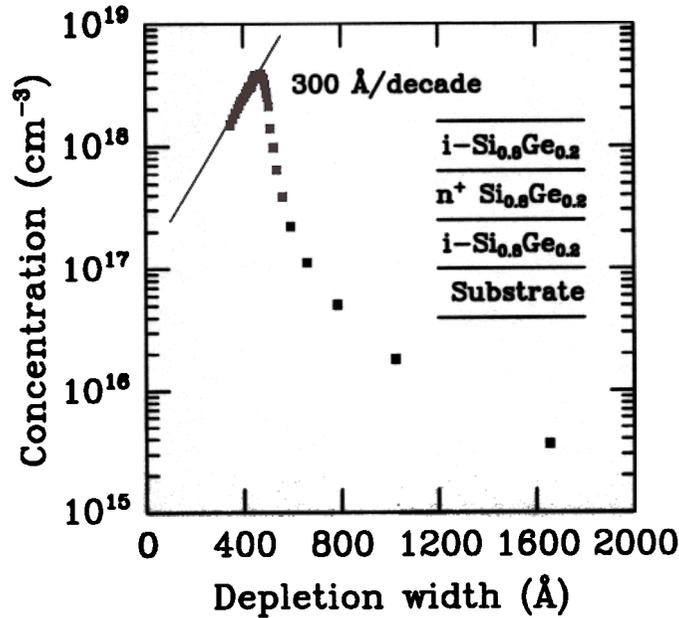
Figure 5.2: Dopant concentration as a function of depth derived from capacitance-voltage measurements on the test structure (sample #1287) shown in the inset using an Al schottky gate. The growth is described in detail in the text. The dopant turn-off rate is ~ 300 Å /decade and the peak concentration corresponding to a 4.5 sccm PH_3 flow agrees well with the spreading resistance data.

this purpose.

Sample #1287 started with 100 Å of undoped $Si_{0.8}Ge_{0.2}$ (DCS=26 sccm, $GeH_4 =$ 100 sccm) on a n-Si substrate. This was followed by 4.5 sccm of PH_3 (diluted 70ppm in H_2) switched on for 2 min for a 200 Å 3×10^{18} cm^{-3} $Si_{0.8}Ge_{0.2}$ layer. Finally the phosphine source was simply turned off for another 4 min to provide the undoped cap. All layers were grown at 625°C . layer separated from the surface by a 400 Å undoped cap. Aluminum was found to be a good Schottky gate with low leakage current at room temperature. This was the first indication that the dopant could be turned off within 400 Å to below $\sim 10^{17}$ cm^{-3} levels. Fig. 5.2 shows the actual dopant profile derived from CV measurements. The background doping is below 1×10^{16} cm^{-3} and the turn-off slope is ~ 300 Å /decade, confirming our previous estimates. It should be noted that, unlike the UHV-CVD work [49], no purging, flushing, growth interruption

or other unusual steps were taken prior to the growth of the undoped cap to ensure an adequate dopant turn-off profile. This experiment also proved that it was possible to deplete the doped layer completely without gate breakdown.

For the MODFETs, we used a combination of the 2DEG structure shown in Fig. 4.5 and the Schottky barrier cap structure described above with some modifications. The spacer was reduced to 20 Å to increase the carrier density in the channel at the expense of some mobility degradation, and the undoped cap layer was reduced to 200 Å to improve the transconductance. The final structure is depicted in Fig. 5.3. Room temperature Van der Pauw measurements indicated a carrier density of 1.8×10^{12} cm^{-2} with a mobility of ~ 1200 cm^2/V·s . The surprisingly high mobility can be attributed to the lifting of the conduction band degeneracy by the strain and selective population of the lower mass valleys. In contrast, a conventional Si n-MOSFET operating at the same carrier density (corresponding to an effective transverse field of $\sim 1.5 \times 10^5$ V/cm) will have an inversion layer electron mobility of 600 cm^2/V·s at room temperature while a 100 Å uniformly doped layer having the same two-dimensional density of carriers will have a mobility of only 200 cm^2/V·s [50]

5.3   Novel two-step processing scheme

The FETs were processed using a novel completely self-aligned two-mask scheme depicted in Fig. 5.4. The goal of the self-alignment is to reduce the number of masking steps and achieve minimum separation between source, drain and gate without sophisticated alignment tools. The gate metal, 1000 Å Al, is thermally evaporated onto the sample in the first step. Source and drain regions are then defined in positive photoresist. A chlorobenzene soak step before developing ensured an "overhang" photoresist profile for easy lift-off in the following step. The aluminum is etched away