Transport in Silicon-Germanium Heterostructures

D. Chrastina*
Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

The work presented here describes the electrical characterization of n- and p-type strained silicon-germanium systems.

Theories of quantum transport in low magnetic fields at low temperature are discussed in terms of weak-localization: the traditional theory is shown not to account for the dephasing in a 2-dimensional hole gas behaving in a metallic manner and emergent alternative theories, while promising, require refinement. The mobility as a function of sheet density is measured in a p-type pseudomorphic $Si_{0.5}Ge_{0.5}$ across the temperature range 350 mK–282 K; it is shown that calculations of the mobility based on semi-classical scattering mechanisms fail below 10 K where quantum transport effects become relevant. A room temperature Hall scattering factor has been extracted.

A new functional form has been presented to fit the resistivity as a function of temperature, below 20 K: traditional theories of screening and weak localization appear not to be applicable.

It is also demonstrated that simple protection circuitry is essential if commercial-scale devices are to be meaningfully investigated.

Mobility spectrum analysis is performed on an n-type strained-silicon device. Established analysis methods are discussed and a new method is presented based on the Bryan's Algorithm approach to maximum entropy. The breakdown of the QHE is also investigated: the critical current density compares well to that predicted by an existing theory.

Finally, devices in which both electron and hole gases can be induced are investigated. However, it is shown that the two carrier species never co-exist. Design rules are presented which may allow more successful structures to be created. Results are presented which demonstrate the success and the utility of implanted contacts which selectively reach different regions of the structure.

1. ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr. D. R. Leadley, for his wisdom, patience and guidance throughout the years I have spent performing the experimental work and data analysis presented here. I would also like to thank Prof. T. E. Whall and Prof. E. H. C. Parker for additional advice and support.

Thanks also go to the whole of the ASR group, especially to R. J. H. Morris and T. W. M. Naylor for technical support and also to Dr. R. J. P. Lander, Dr. O. A. Mironov and Dr. G. Braithwaite for support and instruction in the ways of semiconductor characterization.

I would also like to thank friends within the rest of the Department of Physics, the rest of the University of Warwick, and elsewhere: the past four years certainly would not have been so rewarding without them. I must also acknowledge the support of my parents, who have been very patient and supportive.

I would finally like to thank the Engineering and Physical Sciences Research Council for funding this work, and the University of Warwick for making its facilities available.

Except where specifically stated, all the work described in this Thesis was carried out by the author or under his direction. The devices in Chapter 5 were created by Dr. M. A. Sadegzahdeh at the University of Warwick (55/53) and Dr. H. E. Fischer at Siemens, now Infineon. The device in Chapter 6 was created by R. Ferguson at Imperial College. The devices in Chapter 7 were created by Dr. C. J. E. Emeleus at the University of Warwick.

The computer code used to calculate mobility was written by Dr. A. I. Horrell, and the computer code used to perform Bryan's Algorithm Mobility Spectrum is based on code written by Dr. J. P. Hague.

2. INTRODUCTION TO TRANSPORT IN SILICON-GERMANIUM

With low power dissipation, high integration levels, good noise immunity, high cost-effectiveness and reliability, silicon CMOS (Complementary Metal-Oxide-Semiconductor, see Chapter 7) technology occupies a dominant position in microelectronics. However, the low mobilities† of electrons and holes in silicon limits its application to relatively low frequencies, leaving III–V materials such as Gallium Arsenide (and related materials) to fulfil roles in mobile communications and the like.[1–3]

Strained layers of silicon and silicon-germanium alloy offer scope for dramatic improvements in mobility, and therefore performance. New technology may possibly be incorporated into standard silicon CMOS processing, making the transition favourable to industry.[1–3]

Room temperature mobilities in silicon MOSFETs

† Mobility is a figure of merit for semiconductor materials and is defined and explained in Chapters 3 and 4. The conductivity of a device is proportional to the product of the density of charge carrying particles (number per unit area of material) and their mobility, but the maximum frequency at which a MOSFET can be operated is related to the minimum transit time of carriers through the device and therefore to the mobility of the carriers.[3]

(Metal-Oxide-Semiconductor Field-Effect Transistor) tend to be around 300 $cm^2V^{-1}s^{-1}$ for electrons, and less than 100 $cm^2V^{-1}s^{-1}$ for holes, for sheet densities of the order of 10^{13} cm^{-2}.[1] The advantage of CMOS microelectronics is its very low power consumption (which facilitates higher packing densities) compared to bipolar or NMOS technologies: ideally, a CMOS circuit only dissipates power as it is switching state and it is therefore desirable to equalize the switching time of n and p channel devices, and minimize the switching time overall. In practice this currently means that p-channel MOSFETs in CMOS integrated circuits must be made wider than corresponding n-channel MOSFETs for their conductances to match, and there is a trade-off between channel width and packing density.

It would be advantageous to match the mobilities of electrons and holes, and to increase the mobilities of both electrons and holes overall. This will facilitate higher packing densities, higher operational frequencies or lower-power operation, depending on the requirements of the application.

In addition to this desire to contribute directly to the semiconductor industry, silicon-germanium alloy strained-layer systems can be studied from the perspective of the fundamental physics of semiconductors. The results of these studies (often at liquid-helium temperatures or employing large magnetic fields, on devices with relatively poor characteristics) can then be considered when optimizing the design of industry-level devices for room-temperature operation.

In the field of silicon-germanium research, the highest mobilities have been observed in systems with relatively low sheet carrier densities, meaning that the overall conductivity of the system is not necessarily impressive.[4–6] The relationship between mobility and carrier density is fundamental to characterizing the mechanisms which limit the mobility and will be explored throughout this thesis, especially in the realm of higher sheet densities of both electrons (Chapter 6) and holes (Chapter 5). This should lead, in conjuction with respectable mobilities, to the production of high conductivity devices with potential for high frequency operation.[7] Additionally, research is traditionally carried out on devices which may be up to the order of a millimetre in length and measured parameters may or may not scale down to integration-scale devices on the scale of a few microns in length. A device of this scale will be investigated in Chapter 5. Regarding a new approach to integrating and balancing electron and hole channels, a device which features both will be investigated in Chapter 7.

3. GENERAL THEORY OF TRANSPORT IN SILICON-GERMANIUM ALLOYS

Ignoring doping for the moment, there are two ways in which an alloy layer may be incorporated within a heterostructure. In a pseudomorphic structure (such as that described in Figure 3.1) a silicon-germanium alloy is grown directly on pure silicon. (It is normal practice not to grow the alloy directly onto the silicon substrate but to deposit a few hundred nanometres of pure silicon first.) The alloy layer is then generally capped by more pure silicon (there are issues regarding the oxidization of a silicon-germanium alloy[8]). The alloy layer will match its lattice constant to that of the pure silicon, provided it is not too thick, the germanium concentration is not too high, or the growth temperature is not too high. For this reason, whilst pseudomorphic structures are generally simple to grow, they are generally limited to low germanium concentrations and thin alloy layers.

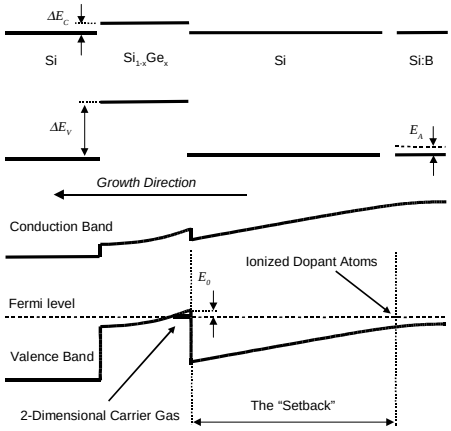

FIG. 3.1: A schematic band profile of an inverted, pseudomorphic structure which shows how a 2-dimensional carrier gas (in this case, of holes) is formed. When brought together, dopant atoms ionize and charge accumulates at the nearest heterointerface.

Alternatively, to allow for greater flexibility regarding thicker alloys with higher germanium concentrations or pure silicon under tensile strain (for electron channels, see Chapter 6) a virtual substrate may be grown. A buffer layer of alloy is grown and allowed to relax, and then the active channels are grown on this. The germanium concentration within the buffer may be graded to increase upwards, but is generally constant for the few hundred nanometres below the active channel. It is important to ensure that the buffer layer is fully relaxed (by growing it at a high enough temperature) and that the defects are not migrating up into the active channel.[9]

3.1. Principle Advantages In Using A Silicon-Germanium Heterostructure

There are three principle advantages to using a silicon-germanium heterostructure as opposed to a plain silicon-silicon dioxide (Si/SiO_2) system, arising from the lower

effective mass, the possibility of band-structure engineering, and the energy band splitting in strained layers.

3.1.1. Effective Mass

The mobility of electron and holes is inversely proportional to their effective mass (Equation 3.3) and the effective mass of holes is generally greater than that of electrons: some bulk properties of common semiconducting materials are summarized in Table 3.1.

TABLE 3.1: Transport properties of some bulk semiconductors at 300 K[10]

	Ge	Si	GaAs
m_e^*/m_e	0.22	0.33	0.067
m_h^*/m_e	0.29	0.55	0.62
$\mu_e/cm^2V^{-1}s^{-1}$	3900	1500	8500
$\mu_h/cm^2V^{-1}s^{-1}$	1900	450	400

However, the effective mass of holes in bulk germanium is smaller than in almost any other semiconducting material and in fact at room temperature the mobility of holes in bulk germanium (1900 $cm^2V^{-1}s^{-1}$) exceeds that of electrons in bulk silicon (1500 $cm^2V^{-1}s^{-1}$). Therefore, it might be expected that sandwiching a layer of germanium, or at least a layer of germanium-rich alloy, between layers of silicon to use as a p-channel would be advantageous.[11–14]

3.1.2. Band-Gap Engineering

The band-gap of germanium is smaller than that of silicon and so, by varying the concentration of germanium in the strained (active) layer* relative to that of the material it is lattice-matched to (either pure silicon in the case of a pseudomorphic system or an alloy in the case of a virtual substrate system) then offsets in the valence and conduction bands arise.vii The valence-band maximum is always in the material with the higher germanium concentration independently of which is the strained layer; for pseudomorphic systems where pure silicon is the substrate then the valence band offset is predicted to rise almost linearly to over 0.7 eV for an active channel of pure, fully strained germanium.[3,14] The conduction-band offset is less certain: if the active channel contains less germanium than the substrate (and is therefore under tensile strain) then the minimum in the conduction band will be in the active channel: a maximum offset of over 0.5 eV is reached for pure silicon on a virtual subtrate which is 85% germanium. However, if the active channel contains more germanium than the substrate then the conduction-band offset is almost always less than the error in the calculations.[3,14] If the conduction-band minimum and the valence band maximum are both located within the strained layer then the band alignment is designated Type I, if one or the other resides in the substrate material then the band alignment is designated Type II.[3,14]

If a layer of material either above ("normal") or below ("inverted") the active layer is doped, then the impurities supply carriers which diffuse into the active channel and are confined there in the triangular quantum well defined by the band offset and their own electrostatic potential to form a 2-dimensional carrier gas. The carriers are then free to move without being directly scattered by the ionized dopant impurity atoms.[1,3,5,15] A schematic example of a band profile for an inverted, pseudomorphic structure (such as will be investigated in Chapters 5 and 7) at $T \ll T_F$ is shown in Figure 3.1.

3.1.3. Strain

The in-plane strain itself generates an anisotropic structure which breaks the conduction and valence band directional degeneracies. In silicon the 6 degenerate conduction band Δ minima are split into a set of 2 in the growth direction and a set of 4 in the plane. Under tensile strain, the former are at the lower energy, which greatly reduces intervalley scattering.[3] In compressively-strained $Si_{1-x}Ge_x$ the light and heavy hole bands are split, substantially lowering the heavy hole mass to below that of the light holes with the heavy hole band lying lowest in energy. This reduces both the intersubband and intrasubband scattering.[3] Generally, when band profiles are drawn, as in Figure 3.1, the "Valence band" and "Conduction band" energies each show only the point of the band which is lowest in energy.

3.2. The 2-Dimensional Carrier Gas

Figure 3.1 is a schematic band profile of a pseudomorphic, inverted p-channel heterostructure. The upper panel shows the components of the system (from right to left as grown: boron-doped silicon, intrinsic silicon, fully strained alloy) and the lower panel shows how the energy bands arrange when an equilibrium is reached. Acceptor atoms at the edge of the dopant slab ionize,† and these free holes form a 2-Dimensional Hole Gas (2DHG) at the heterointerface. (In this system, there are no free charges in the system from, for example, contamination of the supposedly undoped layers. Such "depletion charge" N_{Depl} will be incorporated if necessary, in later chapters.) The 2DHG is examined more closely in Figure 3.2. The electric field between the heterointerface

* The band structure remains silicon-like (six conduction-band minima) for concentrations of germanium up to 85%, and is germanium-like (eight conduction-band minima) beyond this as far as the bulk unstrained alloy is concerned.

and the dopant slab is directly proportional to the charge density in the 2DHG:

$$F = \frac{e}{\epsilon_0 \epsilon_r}(n_s + N_{Depl}) = \frac{e n_s}{\epsilon_0 \epsilon_r} \text{ if } N_{Depl} = 0 \qquad (3.1)$$

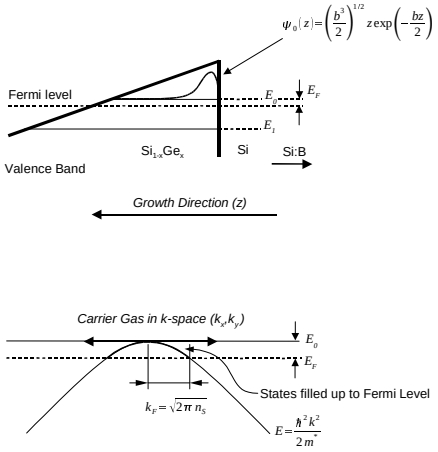

FIG. 3.2: The upper panel gives a closer view of the heterointerface, where the 2DHG forms. E_0 and E_1 represent the two lowest quantum energy levels of the triangular potential well formed by the valence band. The simple Fang–Howard wavefunction is shown, which assumes that the vertical potential barrier is effectively infinite. The lower panel shows how the Fermi Energy E_F depends on the sheet density in the simplest case of constant, isotropic effective mass.

The system is drawn with flat bands in the cap layer effectively meaning that there is no free charge anywhere else. Often, as will be seen in later chapters, the formation of an inversion layer at the top of the silicon cap layer is an important consideration. A triangular quantum well is formed in the valence band. Motion of carriers is constrained completely in the growth direction but free (within the usual effective-mass assumption) perpendicular to the growth direction. It is generally the case that only the ground state of this quantum well is populated (up to the Fermi Energy E_F, as shown in the lower panel of Figure 3.2) since the spacing between the energy levels is of the order of tens of electronvolts for sheet densities ($\sim 10^{12}\text{cm}^{-2}$) comparable to those discussed in later chapters. (In fact, the quantum well ceases to be approximately triangular on these energy scales.) The Fang-Howard approximation to the wavefunction in the ground state of this triangular well is shown[16]

$$\psi_0(z) = \left(\frac{b^3}{2}\right)^{\frac{1}{2}} z \exp\left(-\frac{bz}{2}\right) \qquad (3.2)$$

where b is a variational parameter. This wavefunction assumes that the band offset at the heterointerface is infinite, so the probability density is zero at the interface itself. A realistic finite valence band offset would lead there to be a small, finite component to the wavefunction across the heterointerface, and this has implications for some of the scattering mechanisms which are discussed in the following section.

3.3. Scattering

By considering the rate of change of momentum of the charge carriers, the mobility (as defined in equation 4.1) is found to be related to the transport scattering time τ by[17]

$$\mu = \frac{e\tau}{m^*} \qquad (3.3)$$

Matthiesen's rule says that the total scattering rate is the sum of the scattering rates due to each of the processes operating within the material:[17]

$$\tau^{-1} = \frac{e}{\mu m^*} = \sum_i \tau_i^{-1} \qquad (3.4)$$

At finite temperature, Equation 3.3 becomes an integral which takes into account the energy dependence of the relaxation rates τ_i^{-1}.[18] However, the qualitative result is the same: the mobility is essentially limited by the scattering mechanism with the highest relaxation rate. Scattering rates are generally calculated by considering the reaction of a charge-carrying particle to some sort of potential. This approach necessarily incorporates screening: a particular carrier sees not the potential itself, but rather the potential screened to some extent by the rest of the carrier gas.[18–20] Some typical scattering mechanisms will be mentioned here, with more specific (and mathematical) details to be presented in later chapters.

3.3.1. Impurities

Charge carriers may be scattered by impurities located at or close to the interface at which they are confined. This interface impurity scattering depends mainly on density of impurities at the interface and is stronger when the sheet density of the charge carriers is lower. (At the lowest sheet densities, multiple scattering becomes relevant and eventually the carrier gas becomes strongly localized.[20]) As sheet density increases, impurities are increasingly screened and the scattering effect is reduced.[18,21,22]

3.3.2. Interface Roughness

The roughness of the interface at which the charge carriers are confined leads to scattering which depends

† It is assumed that the dopant depletion width is small enough to be neglected. For typical modulation doping doses, it is roughly 2 nm.

on the depth of the fluctuations, and their characteristic correlation length. The strongest scattering occurs for deep fluctuations when the carrier density is such that the Fermi wavelength coincides with the fluctuation correlation length.[18,23]

3.3.3. Alloy Effects

Alloy scattering and scattering from strain fluctuations may be blamed for the poor performance of heterostructures where the active channel is an alloy with roughly equal germanium and silicon concentrations.The first mechanism exists because even though the silicon-germanium alloy has a crystalline structure, whether a silicon or germanium atom occupies a particular lattice site is random. There is controversy regarding the strength of this mechanism.[18,24] (Poor performance of structures which feature intermediate germanium concentrations may not be due to alloy scattering directly but due to issues regarding the growth of the heterostructures: low growth temperatures are needed to avoid the relaxation of the alloy as it is grown but lead to poor material with high crystal defect densities. Higher growth temperatures lead to the formation of germanium-rich islands on the growth surface and eventually result in a material with uneven germanium distribution and an undulating interface.[25]) The second mechanism is induced by deformation associated with interface roughness.

3.3.4. Phonons

The acoustic phonon scattering relaxation rate is directly proportional to temperature and is related to material and physical constants rather than electronic parameters. In a confined geometry, optical phonon scattering should be considerably weaker provided that only a single subband is occupied.[11,24] In any case, phonon scattering is not relevant unless the lattice temperature is greater than the Fermi temperature and the carrier gas becomes nondegenerate,[11] which for a 2-dimensional carrier gas with a circular Fermi line and a constant effective mass is:

$$T_F = \frac{E_F}{k_B} = \frac{1}{k_B}\frac{\hbar^2 k_F^2}{2m^*} = \frac{\hbar^2 \pi}{k_B}\frac{n_s}{m^*} \tag{3.5}$$

These analyses have generally neglected the interactions between carriers in the gas, and other effects which are most relevant at very low temperatures or high magnetic fields, and will be introduced in Chapter 4.

4. EXPERIMENTAL DETAILS

This chapter explains some of the general principles behind the characterization of the devices detailed in the following chapters.

4.1. Standard Characterization Methods

The simple process of attaching a pair of leads to either end of a piece of semiconductor material, passing a current through them, and measuring the voltage developed across them, is generally not sufficient for any accurate characterization. The value of the resistance of the leads and contacts is an unknown variable, but if it is known to be small compared to the resistance of the sample, then meaningful qualitative results can be obtained from this 2-terminal approach.

For quantitative results, it is necessary to use a 4-terminal approach: current is passed along the sample between two contacts, and the voltage across two other contacts is measured. The simplest arrangement for performing this measurement involves having a bar-shaped sample, with the current passing along the long axis: if the aspect ratio is sufficiently large, and the voltage probes are not too near the ends, then the resistivity can be found very easily.[26] It is important to make sure that the contact resistances are not too high, that the impedance of the voltage measurement equipment is as high as possible.

4.1.1. Drift Mobility

Reference has been made in the introductory chapters to the mobility of the charge carriers within the material in terms of the rate at which they undergo scattering events. In terms of the measurement process, however, the mobility is defined by the drift velocity that the carriers reach when subjected to a given electric field, and is therefore sometimes explicitly referred to as the drift mobility:

$$\mu = \frac{|\boldsymbol{v}|}{|\boldsymbol{E}|} \tag{4.1}$$

where $\boldsymbol{v}$ is the drift velocity of the carriers and $\boldsymbol{E}$ is the applied electric field. If the applied current is I, the measured voltage is V, the sheet charge density is qn_s,* the distance between the voltage probes is l and the width of the sample is w, then

$$\mu = \frac{I}{Vqn_s} \tag{4.2}$$

In simple cases, the sheet charge density can be calculated using the Hall effect.

---

* Since we are concerned with a 2-dimensional carrier gas, it is convenient to refer to a "sheet carrier density" n_s (with dimensions of number per area) rather than a volume carrier density. The use of n_s for the carrier density does not necessarily imply n-type conduction.

The most widely used test of electrical quality of semiconductor materials is the Hall measurement.[27] The simplest case concerns a material in which a single carrier is responsible for passing the current through a material with isotropic energy bands and energy-independent scattering mechanisms.

Single Carrier Classical Hall Effect

Using Ohm's law, we define the resistance along the direction of current as

$$R_{xx} = \frac{V_x}{I_x} \tag{4.3}$$

For a 2-dimensional system, this is converted into a "sheet resistivity" by correction for the aspect ratio of the device:

$$\rho_{xx} = \frac{V_x}{I_x}\frac{w}{l} \tag{4.4}$$

Even though w/l is dimensionless, sheet resistivity is generally specified as Ohms per square. The scalar ρ_{xx} ($= \rho_{yy}$) is the diagonal component of the resistivity tensor defined by $\boldsymbol{E} = \rho\boldsymbol{J}$, and the off-diagonal element ρ_{xy} (=-ρ_{yx}) becomes non-zero in the presence of a magnetic field $\boldsymbol{B}$ as summed up in the Lorentz force law $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$. In the Hall bar under consideration (with $\boldsymbol{B} = (0, 0, B)$ and $\boldsymbol{v} = (v, 0, 0)$) a stable condition is eventually reached ($\boldsymbol{F} = 0$) when $E_y = vB$. Then, the velocity v of the carriers is related to the current I by $I = nqvA = nqvwt = n_s qvw$.

The voltage produced across the Hall bar by the electric field is the Hall Voltage, $V_H = E_y w$, which can be related back to the magnetic field and current:

$$V_H = \frac{I}{n_s q w}B \tag{4.5}$$

Finally, the off-diagonal component of the resistivity tensor can be found, and the Hall coefficient R_H can be defined.

$$\rho_{xy} = \frac{V_H}{I} = \frac{B}{n_s q} = R_H B \tag{4.6}$$

From this, an expression for finding the sheet carrier density in terms of known or measured quantities can be deduced:

$$n_s = \frac{IB}{qV_H} \tag{4.7}$$

Since the mobility of the carriers (Equation 4.1) relates back to the conductivity by $\sigma = nq\mu$ then in low magnetic field where $\mu B \ll 1$:

$$\rho_{xx} = \frac{1}{n_s q \mu} \tag{4.8}$$

So, finally:

$$\mu = \frac{1}{n_s q \rho_{xx}} = \frac{\rho_{xy}}{B\rho_{xx}} = \frac{1}{B}\frac{V_H}{V_X}\frac{l}{w} \tag{4.9}$$

Based on these simple calculations, the sheet resistivity ρ_{xx} and Hall coefficient R_H are functions of carrier density and mobility and are constant with respect to the magnetic field. However, it is rarely the case that this simple model is directly applicable. In cases where this simple model has been applied without consideration, the properties calculated can be referred to as the *Hall* sheet density and mobility, sometimes denoted as n_H (or p_H for hole-like conduction) and μ_H.

Generalized Classical Hall Effect

The resistance ρ and Hall coefficient R_H vary with magnetic field if there are two or more distinct carrier gases present in the material, or if the carrier gases feature a spread of mobilities. In general,[28]

$$\sigma_{xx}(B) = \int_{-\infty}^{\infty} \frac{s(\mu)}{1 + \mu^2 B^2}\mathrm{d}\mu \tag{4.10}$$

$$\text{and} \quad \sigma_{xy}(B) = \int_{-\infty}^{\infty} \frac{\mu B s(\mu)}{1 + \mu^2 B^2}\mathrm{d}\mu \tag{4.11}$$

$$\text{where} \quad s(\mu) = n_s(\mu) q \mu \tag{4.12}$$

is a generalized conductivity function, and $n_s(\mu)$ represents the mobility spectrum of carriers throughout the material. The experimentally measured parameters are related to the elements of the conductivity tensor by:

$$\rho(B) = \rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \tag{4.13}$$

$$R_H(B) = -\frac{\rho_{xy}}{B} = -\frac{1}{B}\frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \tag{4.14}$$

The variation of resistance with magnetic field is known as magnetoresistance. In general, classical magnetoresistance is always positive, meaning that the resistance increases with magnetic field.

Assuming a single, ideal carrier gas is equivalent to using an $n_s(\mu)$ of the form of a single delta function, and the result of the previous section is recovered. By using a function for $n_s(\mu)$ whose form comprises a pair of delta functions, the two-carrier Hall coefficient can be found. The Hall coefficient for a two carrier system at low fields, $\mu B \ll 1$, is given by

$$R_H = \frac{\pm n_1 \mu_1^2 \pm n_2 \mu_2^2}{q(n_1\mu_1 + n_2\mu_2)^2} \tag{4.15}$$

$$\text{which becomes} \quad R_H = \frac{1}{q(\pm n_1 \pm n_2)} \tag{4.16}$$

at high fields, where $\mu B \gg 1$, unless $n_1 \sim -n_2$ (when the condition that $\sigma_{xy} \gg \sigma_{xx}$ breaks down). Positive and negative signs refer to hole and electron transport respectively. Hence, a single carrier model calculation based on the measurement of a two carrier system at low field would yield μ_H and n_H such that

$$\mu_H = \frac{|\pm n_1\mu_1^2 \pm n_2\mu_2^2|}{n_1\mu_1 + n_2\mu_2} \tag{4.17}$$

$$\text{and} \quad n_H = \frac{(n_1\mu_1 + n_2\mu_2)^2}{\pm n_1\mu_1^2 \pm n_2\mu_2^2} \tag{4.18}$$

Even single carrier systems with an isotropic effective mass can exhibit magnetoresistance, and this is most commonly due to energy-dependent scattering mechanisms. The Hall scattering factor r is given by

$$r = \frac{\langle\tau^2\rangle}{\langle\tau\rangle^2} \tag{4.19}$$

where τ is the scattering time, and the angle brackets refer to averaging over energy.

$$\text{at low fields} \quad R_H = \frac{r}{n_s q} \tag{4.20}$$

$$\text{but at high fields} \quad R_H = \frac{1}{n_s q} \tag{4.21}$$

as before. Equations 4.16 and 4.21 are essentially the same: at high fields, the Hall coefficient depends only on the total charge density in the system. If two carrier gases are present, each with some sort of anisotropy or energy dependent scattering mechanisms, then at low fields both Equations 4.15 and 4.20 apply, and the problem becomes intractable. In fact, the distinction between the existence of multiple carrier gases and the existence of a single carrier gas with a range of properties can become a matter for convention rather than physics in real materials; this distinction may be obvious when considering spatially separate carrier gases but not if, for example, multiple subbands are occupied within the same spatial region of the device at finite temperature.

Computationally intensive methods must be employed to invert Equations 4.10 and 4.11 without making assumptions regarding the form of $n_s(\mu)$ provided at least that μ itself is not a function of magnetic field.[28,29]

In the low temperature degenerate limit there is no broadening of the Fermi-Dirac occupation function in terms of energy, so the Hall scattering factor $r \sim 1$. As temperature increases, the carriers' energy spreads so the effects of energy-dependent scattering become apparent. In this case, the Hall factor tends to be greater than unity.

Alternatively, the Hall factor can deviate from unity if the energy bands are anisotropic, leading to anisotropy in the effective mass. This consideration is especially relevant in strained layers, such as a silicon-germanium alloy grown epitaxially on a silicon substrate, where the strain breaks the symmetry of the energy bands. In this case, the Hall factor is not necessarily greater than unity.

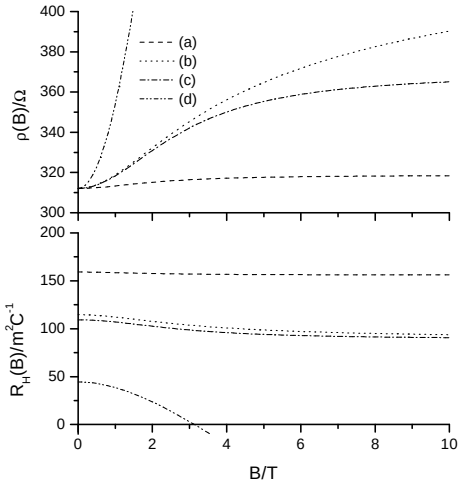

FIG. 4.1: Magnetoresistance calculated for four different systems. They are (a) a single carrier gas with a large mobility width, (b) two carrier gases with large width, (c) two "ideal" gases and (d) two gases, one of which contains carriers of the opposite charge (represented by a "negative" mobility and sheet density).

Figure 4.1 shows magnetoresistance calculated for a number of systems, summarized in Table 4.1. The carrier gases are represented by Gaussians $n_s(\mu)$ of full-width $\Delta\mu_i$ centred at μ_i normalized such that:

$$\int_{-\infty}^{\infty} n_s(\mu) q\mu \mathrm{d}\mu = n_i q \mu_i \tag{4.22}$$

Magnetoresistance features appear on a scale of $B \sim \frac{1}{B}$, so the presence of multiple high-mobility gases gives rise to magnetoresistance effects at low fields, whereas a range of mobilities within a single gas causes effects at much larger field scales.

Also, the presence of two carrier gases with opposite charge (that is, electrons and holes) causes very large magnetoresistance and the eventual reversal of the sign of R_H.

4.1.3. Quantum Effects at High Magnetic Fields

In a perfect (that is, non-scattering) semiconductor in the zero temperature limit, the application of a perpendicular magnetic field would cause the carriers to execute cyclotron orbits. The conductivity of the sample would be zero. Current will only flow if the cyclotron motion is interrupted by scattering events.

Expressing this more formally, the presence of a perpendicular magnetic field such that $\hbar\omega_c \geq k_B T$ where

TABLE 4.1: Parameters used to generate the magnetoresistance curves of Figure 4.1.

	μ_1 [$\mathrm{cm^2V^{-1}s^{-1}}$]	$\Delta\mu_1$ [$\mathrm{cm^2V^{-1}s^{-1}}$]	n_1 [$\mathrm{cm^{-2}}$]	μ_2 [$\mathrm{cm^2V^{-1}s^{-1}}$]	$\Delta\mu_2$ [$\mathrm{cm^2V^{-1}s^{-1}}$]	n_2 [$\mathrm{cm^{-2}}$]
(a)	5000	1000	4.0			
(b)	2000	1000	5.0	5000	1000	2.0
(c)	2000	10	5.0	5000	10	2.0
(d)	-2000	1000	-5.0	5000	1000	2.0

ω_c is the cyclotron frequency $\frac{qB}{m^*}$ leads to the formation of Landau levels. Landau levels are eigenstates of the Schrödinger equation for a 2D carrier gas in a perpendicular magnetic field, essentially the quantum expression of the semi-classical concept of cyclotron orbits. Ideally the sub-band structure created by Landau levels takes the form of a series of delta functions at energies $E_n = \hbar\omega_c(1 + \frac{1}{2})$, but in a real system in the presence of scattering the energy can only be defined to within $\sim \frac{\hbar}{\tau_q}$, where τ_q is the quantum lifetime, leading to broadened Landau levels which are only distinct if $\hbar\omega_c \geq k_B T$. Essentially, this means that for these effects to be relevant, each carrier must have enough time to execute at least one cyclotron orbit between scattering events,[30] ie. $\tau_q > \frac{2\pi}{\omega_c}$.

As the magnetic field increases, the separation between Landau levels grows, and so does the number of states each contains,[30] $\frac{eB}{h}$. The position of the Fermi level will move away from its zero-field value, so that the density of the carrier gas remains constant as the number of available Landau levels decreases. The filling factor ν is a measure of the number of filled Landau levels. At integer values of ν the Fermi level is half way between two Landau levels and there can be no scattering. For a constant carrier sheet density, $\nu \propto \frac{1}{B}$. In an imperfect system, impurities and crystal defects lead to random potential fluctuations which cause states at the edges of each Landau level to be localized so that they do not contribute to the conductivity of the device.

However, if a carrier begins an orbit very close to the edge of a spatially finite device, then it will hit the edge of the device (a potential barrier) and be scattered; this will lead to a net drift velocity which will be higher the closer the carrier is to the edge. The presence of an edge effectively bends the subband structure upwards in energy. These "edge states" behave as one-dimensional quantum wires and mean that there is conduction even when the Fermi level is positioned between Landau levels in the bulk of the device.

These considerations lead to a completely revised picture of magnetotransport, compared to behaviour in the classical regime.

Figure 4.2 shows data from the sample studied in Chapter 6, a 2-dimensional electron gas confined within a quantum well of pure silicon, sandwiched between relaxed $Si_{0.75}Ge_{0.25}$ layers. The temperature is 350 mK; the carriers have a mobility of around 25,000 $\mathrm{cm^2V^{-1}s^{-1}}$ and a sheet density of $8.1\times10^{11}\mathrm{cm^{-2}}$. For fields around 1 or 2 Tesla, Shubnikov–de Haas oscillations (which are periodic in inverse magnetic field) appear. Minima in resistivity appear at filling factors of $4n$, where n is an integer. For larger fields, the oscillations first become "spin-split" and then "valley-split," eventually showing minima for all integer n.

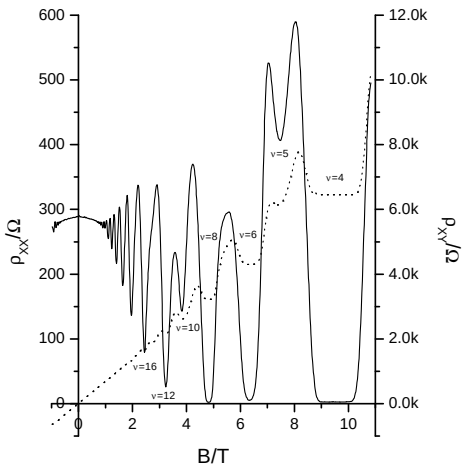

FIG. 4.2: A 2DEG showing Shubnikov–de Haas and Quantum Hall Effects at a temperature of 350 mK. The dotted line is the transverse resistivity and is associated with the y-axis on the right. Minima in resistivity are labelled with their corresponding filling factors. Conduction is dissipationless at $\nu = 4$.

The spin degeneracy is lifted (spin-splitting is observed) when the Zeeman energy between spin states is greater than the broadening of the Landau Levels. In Figure 4.2 spin-split resistance minima can be seen at $\nu = 10$ and $\nu = 6$. Thus at higher magnetic fields minima appear at $2n$.

Since the electron has spin $\frac{1}{2}$, in an applied magnetic field they can take a low energy orientation along the field or a high energy orientation against it. The energy difference between these two states is $E_{Zeeman} = g\mu_B B$, with $g = 2.0023$ for a free electron.[31] In the 2DEG the charge carriers are not free electrons, but rather the electron-like quasiparticles that result from the interactions of the electron with other electrons and the silicon lattice. The Landé g-factor of electrons in bulk silicon is 1.99[32] but

in a 2DEG it is larger and is in general anisotropic[33–36] and a function of magnetic field and sheet density.[37–39] (Values of between 2.6 and 4.2 are reported in Reference 38, found by tilting the magnetic field but ignoring possible anisotropy.) The Zeeman splitting energy can be compared to the Landau level separation $\hbar\omega_c$ and it is found that

$$E_{Landau} = \frac{2}{g}\frac{m_e}{m^*}E_{Zeeman} \tag{4.23}$$

The appearance of a minimum in ρ_{xx} for an odd filling factor ($\nu = 5$ at 7.4 T) shows that the two-fold valley degeneracy has also been lifted: there is additional degeneracy factor of 2, because in silicon the electrons in the 2DEG occupy two [001]-valleys. The precise origin of valley splitting is not entirely understood, but may be related to the strength of the electric field at the interface which defines the 2DEG.[38]

As the magnetic field reaches 9 T, ρ_{xx} drops to zero, but since ρ_{xy} remains finite (and large) then σ_{xx} is also (counterintuitively, but clearly from Equation 4.13) zero and conduction is dissipationless, in accordance with the theory given above which asserts that the conductivity drops to zero when the Fermi level is halfway between two Landau levels. The ρ_{xy} data show classical $\rho_{xy} \propto B$ behaviour for low fields, and integer quantum Hall effect plateaux where the transverse resistance is pinned at $\rho_{xy} = \frac{h}{\nu e^2}$ over a finite magnetic field range, for higher fields.

For holes, where both m^* and g are larger than for electrons,[33] the up-spin state of one Landau level can be closer in energy to the down-spin state of the next. Since there is no valley degeneracy (there is a single minimum in the band at $|\boldsymbol{k}| = 0$) then minima which are not spin-split would appear at odd filling factors.

Shubnikov–de Haas oscillations are very useful, because their period in inverse magnetic field can be used to determine the true carrier sheet density, in contrast to Hall measurements which may be invalidated by the anisotropy of the band structure, for example. Their decay with increasing temperature can be used to determine the effective mass of the carriers, and their amplitude with respect to magnetic field can be used to determine the ratio α between the quantum lifetime τ_q (the time taken for a particle to scatter elastically to any other state) and the transport lifetime τ (which recognises that small angle scattering has little bearing on the drift velocity, whilst large angle scattering does; referred to explicitly as τ_{tr} when necessary for clarity).

If the Landau levels are assumed to be broadened into Lorentzians each with a width $\frac{\hbar}{2\tau_q}$ then the resistivity of a 2-dimensional carrier gas subject to a perpendicular magnetic field at low temperatures is given by:[40]

$$\rho_{xx}(B,T) = \rho_0(T)\left[1 + 4\sum_{s=1}^{\infty}\frac{s\xi(B,T)}{\sinh[s\xi(B,T)]}e^{\frac{-\pi\alpha s}{B\mu(T)}}\cos\left(2s\pi^2\frac{\hbar n_s}{qB} - s\pi\right)\right] \tag{4.24}$$

$$\text{with} \quad \xi(B,T) = \frac{2\pi^2 k_B T m^*}{eB\hbar} \tag{4.25}$$

$$\text{and} \quad \alpha = \frac{\tau_{tr}}{\tau_q} \tag{4.26}$$

provided that the oscillations are small. The $s > 1$ terms in the summation are only relevant for very high mobility samples, so only $s = 1$ will be considered. The cosine term represents the oscillations themselves, and if their period in inverse magnetic field is $\Delta(1/B)$ then the sheet density n_s can be extracted using

$$n_s = \frac{q}{\pi\hbar\Delta(1/B)} \tag{4.27}$$

The exponential term controls the amplitude of the oscillations with respect to magnetic field (in the zero temperature limit) depending on how many cyclotron orbits each carrier executes between scattering events.

The term $s\xi(B,T)/\sinh[s\xi(B,T]$ thermally damps the oscillations.

In order to find m^* and α, first it is necessary to pick out the maxima and minima of the oscillations (using magnetic fields values between the onset of oscillations up to spin-splitting) for a number of temperatures, and subtract the slowly varying classical magnetoresistance background. This leaves a set of amplitudes of the oscillations themselves, $\delta\rho_{mm}(B,T)$.

Assume a value (for example, unity) for α and then consider $\delta\rho_{mm}(T)$ at a particular B: A plot of

$$\ln(\delta\rho_{mm}) \text{ vs. } \ln\left(\frac{\rho_0(T)\xi(m^*,T)}{\sinh\xi(m^*,T)}\right) - \left(\frac{\pi\alpha}{\mu(T)B}\right) \tag{4.28}$$

will have a gradient of unity and an intercept of $\ln(4)$ if the choice of m^* is correct. $\rho_0(T)$ is the Boltzmann resistivity: the value of $\rho_{xx}(B = 0, T)$ may be used, or the value of $\rho_{xx}(T)$ at the field just before oscillations become visible. If no temperature dependence is assumed, then

$$\ln(\delta\rho_{mm}) \text{ vs. } \ln\left(\frac{\xi(m^*,T)}{\sinh\xi(m^*,T)}\right) \tag{4.29}$$

is plotted instead. Data from maxima and minima at many magnetic fields should be averaged.

This effective mass can then be used in a plot of

$$\ln\left[\left(\frac{\sinh\xi}{\xi}\right)\left(\frac{\delta\rho_{mm}}{\rho_0}\right)\right] \text{ vs. } \frac{1}{\mu B} \tag{4.30}$$

for $\delta\rho_{mm}(B)$ at a particular temperature. These points should lie on a line with a gradient of $-\pi\alpha$ and an intercept again of $\ln(4)$. This α should be used to determine a better value for m^*, and iterations should continue until satisfactory fits for both parameters are found.[41]

Since τ_q considers all elastic scattering events to be equally important but τ_{tr} considers that the most important events are those that lead to the greatest change in the direction of the wavevector, it is generally the case that $\tau_q < \tau_{tr}$ and thus that $\alpha > 1$. For example, scattering from remote impurities is predominantly small-angle so leads to a value of α greater than unity (in GaAs/$Ga_{1-x}Al_xAs$ heterostructures the value of α can be of the order 10) whilst scattering from a rough interface tends to lead to $\alpha \sim 1$.[42]

4.1.4. Quantum Effects at Low Magnetic Fields

In a disordered semiconductor where $k_BT \ll E_F$, the transport properties of a 2-dimensional carrier gas may reflect the quantum nature of the charge carriers even in the absence of a field strong enough to form Landau levels. (These effects can usually be generalized to other dimensionalities).

Weak localization

Figure 4.3 shows a particle which is scattered around in a loop, due to disorder in the system: each scattering event corresponds mainly to a change in the direction of the momentum with only a small change in energy. By definition, these happen on a timescale of the elastic scattering time (or quasi-particle lifetime) τ. In quantum mechanical terms, this means the wavefunction of the particle interferes with itself at the intersection point. The wavefunction must be single valued at this point, so the probability current cannot cross itself. Instead, three paths are possible; part of the current will ignore the loop entirely and continue unscattered, part of the current will travel around the outside of the loop, and the remaining part will remain trapped within the loop.

If a scattering event occurs which changes the energy of the particle significantly (and therefore breaks the time-reversal symmetry of the loop) then the phase coherence of the wavefunction will be broken, the quantum mechanical constraint will be relaxed, and the particle will no longer be localized. Events like this occur on a timescale of the dephasing time τ_ψ which is related to the inelastic scattering time.[43] The maximum loop size is therefore set by τ_ψ whilst the minimum loop size is set by τ.

A magnetic field applied perpendicular to the plane of the carrier gas will serve to change the phase of the wavefunction around the loop. Larger loops will be the most affected since they will enclose more flux.

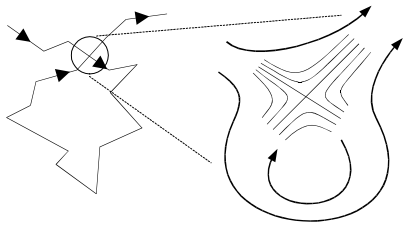

FIG. 4.3: A schematic representation of weak localization. On the left, a classical trajectory featuring a loop is shown. However, in quantum-mechanical terms the wave-function must be single valued at the intersection point (shown on the right: fine lines represent equal-probability contours) and part of the probability current (shown by arrows) becomes localized in the loop. Inelastic scattering will destroy the time-reversal symmetry and break this coherence, whilst a magnetic field through the loop will change the phase of each of the paths relative to each other.

The correction to the Boltzmann conductivity in the diffusive limit $\tau \ll \tau_\psi$ is given by[44]

$$\delta\sigma_{xx}^{loc}(B) = -\frac{e^2}{2\pi^2\hbar}\left[\Psi\left(\frac{1}{2}+\frac{\hbar}{4eBD\tau}\right) - \Psi\left(\frac{1}{2}+\frac{\hbar}{4eBD\tau_\psi}\right)\right] \tag{4.31}$$

with the diffusion constant $D = \frac{E_F\mu}{e}$ and the digamma function (related to the Gamma function $\Gamma(x)$) $\Psi(x) = (\ln[\Gamma(x)])' = \frac{\Gamma'(x)}{\Gamma(x)}$. In the limit of low field, this reduces to[45]

$$\delta\sigma_{xx}^{loc}(0) = -\frac{e^2}{2\pi^2\hbar}\ln\left(\frac{\tau_\psi}{\tau}\right) \tag{4.32}$$

The transverse component of the conductivity tensor is

then found to be

$$\delta\sigma_{xy}^{loc}(B) = -2\mu B\delta\sigma_{xx}^{loc}(B) \tag{4.33}$$

and this result is assumed to hold for all B.[46]

Weak localization is quenched when $\mu B \sim \hbar/E_F\tau$: the magnetic flux through the loop described by the weakly-localized carrier is such that the phase change around the loop reaches the order of π even for the smallest allowed loops.

Interactions and Spin-Splitting

Corrections also arise from particle-hole interactions (which are not directly dependent on magnetic field) and the manner in which these interactions are effected by the Zeeman splitting of spin states. The former is given by[43,45,47,48]

$$\delta\sigma_{xx}^{int} = \frac{e^2}{2\pi^2\hbar}\left(1 - \frac{3}{4}F^*\right)\ln\left(\frac{k_BT}{\hbar/\tau}\right) \tag{4.34}$$

Here, F^* is a renormalized screening parameter

$$F^* = \frac{8}{F}\left(1 + \frac{1}{2}F\right)\ln\left(1 + \frac{1}{2}F\right) - 4 \tag{4.35}$$

where F itself comes from considerations of the exchange and Hartree contributions to the self-energy correction (representing the interaction of a given eigenfunction with the nonuniform electron density in the ground state) and is given by[43,45]

$$F = \int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}\left[1 + 2k_F\frac{2\pi\epsilon_0\epsilon_r\hbar^2}{m^*e^2}\sin(\theta/2)\right] \tag{4.36}$$

The $-\frac{3}{4}F^*$ Hartree term in Equation 4.34 was initially calculated to be simply $-F$, under the assumption that and the addition of this first-order term was therefore valid.[45] The realization, however, that triplet and singlet states should feature in the considerations of particle-hole scattering lead to the replacement of F with $-\frac{3}{4}F^*$ in each of the formulae in which it appears. F^* is actually a different function of F depending on which kind of correction one is calculating and the dimensionality: Equation 4.36 is appropriate for 2-dimensional conductivity.[48] Since F can in theory only take values between 0 and 1 (decreasing with increasing sheet density) then F^* can only take values between 0 and 0.866.

The correction to the conductivity due to the Zeeman splitting of spin states (provided that this is greater than their broadening due to very high spin-orbit scattering or spin-flipping rates) is given by[43]

$$\delta\sigma_{xx}^{spin}(B) = -\frac{e^2}{2\pi^2\hbar}\left(\frac{F^*}{2}\right)G\left(\frac{g\mu_BB}{k_BT}\right) \tag{4.37}$$

Here, μ_B is the Bohr magneton, g is the Landé g-factor of the carriers and may in general be anisotropic or field or density dependent. The function G has been evaluated numerically.[49]

Inversion and implicit differentiation of the conductivity tensor leads to the following expression for the magnetoresistance $\Delta\rho_{xx}(B) = \delta\rho(xx)(B) - \delta\rho(xx)(0)$ which incorporates the corrections described above:[46]

$$\Delta\rho_{xx}(B) = \rho_0^2[(\mu^2B^2 - 1)\delta\sigma_{xx}^{loc}(B) + \delta\sigma_{xx}^{loc}(0) - 4\mu^2B^2\delta\sigma_{xx}^{loc}(B)] + \rho_0^2[(\mu^2B^2 - 1)\delta\sigma_{xx}^{spin}(B) + \mu^2B^2\delta\sigma_{xx}^{int}(B)] \tag{4.38}$$

Interaction effects should also lead to corrections in the Hall signal, whereas weak localization effects should not. Also, if the magnetic field is applied parallel to the plane of the 2-dimensional carrier gas then all magnetoresistance terms incorporating μB apart from the Zeeman term should vanish.[46] However, calculations and measurements of the anisotropy of the Landé g-factor suggest that it too vanishes in a parallel magnetic field.[33,34] Applications of this theory will be presented in Chapter5.

4.2. Measuring Equipment

4.2.1. Room Temperature IV Measurements

Making Contact to Unmounted Samples

It is possible to characterize a semiconductor device at room temperature without disecting a wafer or mounting the chip into a package, by using a needle probing station. The chip is positioned on a stage under a stereoscope. The stage can be moved in both the x and y directions so that the intended device can be brought into view.

Needle probes, which feature a needle that can be moved in the x, y and z directions mounted on a heavy magnetic base, can then be gently (and very carefully) lowered onto the contact pads of the chip. Each probe needle can then be connected to the measurement electronics and the device can be characterized.

IV Measurement with the Hewlett-Packard HP4145 Parameter Analyzer

A HP4145 Parameter Analyzer, specially designed for the characterization of electrically sensitive semiconduc-

tor devices, was employed extensively throughout. Applied currents and voltages were limited (generally to the order of microamps and millivolts drain-source, although higher voltages and smaller currents were necessary for gate terminals) so as not to subject the device under test to harmful extremes, whilst the high input impedance (at least $10^{12}\Omega$) ensured that voltages were measured accurately. Gate voltages were only applied once the source and drain voltages were set: this was particularly relevant to the sensitive Siemens devices described in Chapter5.

4.2.2. Low Temperature Measurement Issues

Sample Mounting

Initially, small, thin packages with a very small lead spacing were used for the low-temperature characterization of these samples. However, this led to problems with the connection of wires to the sample. Securing the package to the cryostat cold-finger was also unsatisfactory, and since space was not an issue within the closed-cycle cryostat larger 14-pin DIL (Dual InLine) packages were employed instead. These were based on the industry-standard 0.1" pitch, so cheap sockets were available from electronics suppliers. The device could be mounted into its socket once wires were already soldered and all the test equipment connected, so that the device was less at risk from stray and potentially damaging electrical signals. A cryostat block incorporating a 14-pin DIL socket was fabricated, which also solved physical mounting issues.

The process of electrically connecting the sample's contacts to the leads of the package, using an ultrasonic gold-wire ball-bonder, was not entirely reliable. As will be apparent in Chapter 7, not every contact worked at low temperatures. The bonding process may have been responsible for altering the IV characteristics of the contacts. However, provided that the resistances of the contacts remained reasonable any non-linearities should have been corrected for by the 4-terminal method.

Lock-In Amplifiers and Automated Computer Recording.

Lock-in amplifiers were employed mainly for low-temperature, low signal measurements. Computer programs existed for automating the measurement process and recording the lock-in values via the IEEE-488 General-Purpose Interface Bus (GPIB). Also, a configuration was developed where the HP4148 Parameter Analyser reliably modulated the gate voltage and recorded the analogue output voltages of the lock-in amplifiers, which were measuring the other voltages and currents in the system. This system was both more efficient and safer (for the device) than applying gate voltages manually with a standard Keithley voltage source. Also, when many devices are connected in an IEEE system, one temperamental device can halt the whole measurement and human intervention is required.

Current Heating

Care was taken at every stage to ensure that the current drive through the devices was not causing any significant heating, either of the carriers in relation to the lattice (bringing them out of thermal equilibrium with each other) or of the whole lattice. In general this was performed by checking for Ohmic behaviour in the IV characteristics, but considerations of, for example, the energy loss rates per carrier will be presented where especially relevant.

Cryostats

Two cryostats were employed for the experimental work presented in Chapters 5, 6 and 7. Initially, Hall-effect and low-temperature IV measurements were performed in a closed-cycle cryostat with a base temperature of 10 K and a permanent magnet with a field of 0.41 T. The permanent magnet was replaced by an electromagnet with a maximum field of 1.2 T during the course of these investigations.

More detailed Hall-effect, magnetoresistance and temperature-dependance measurements were made using an Oxford Instruments ^{3}He cryomagnetic system, which allowed more-or-less stable temperatures to be reached between 350 mK and ∼300 K and fields of up to 11 T.

5. P-CHANNEL DEVICES

5.1. Abstract

It is desirable, as explained in Chapter 2, to increase the performance of holes to match electrons at room temperature within a semiconductor system, and strained silicon-germanium heterostructures are promising in this regard. However, this promise is not always fulfilled. It would be useful to be able to find out how to optimize the design and growth of such structures rather than rely on trial and error, and low-temperature characterization of the p-type silicon-germanium system is investigated with this in mind. Even though the results in terms of fundamental 2-dimensional carrier gas physics may not seem directly applicable to room-temperature electronics (despite being interesting in their own right) information about scattering mechanisms and localization may lead to higher quality material for commercial device production.

5.2. Introduction

Two contrasting p-channel devices are investigated in this chapter: the first is a research-scale device grown by

MBE with a $Si_{0.8}Ge_{0.2}$ active channel and inverted doping; the second is a commercial-scale device grown by CVD with a $Si_{0.5}Ge_{0.5}$ active channel and n-type doping throughout. Both have gates so that the sheet carrier concentration can be modulated, and are in standard Hall-bar geometries.

5.3. Structure of Devices

5.3.1. Research Scale Device on Wafer 55/53

This device is a simple heterostructure with doping below the alloy and a metal gate. In contrast to the Siemens device (below) its large size makes it relatively uncomplicated to handle and measure. The structure of this device is shown in Figure 5.1.

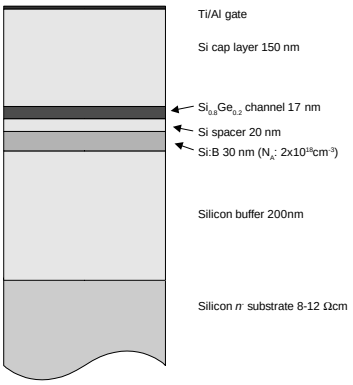

FIG. 5.1: The structure of sample 55/53, a research scale p-channel device featuring inverted doping and a pseudomorphic alloy layer.

This device is a heterostructure with the same alloy composition as the coupled-channel devices that will be considered in Chapter 7. The buffer layer was grown at 700°C, the dopant layer and the spacer were grown at 575°C, and the alloy and cap layers were grown at 610°C. Contacts were formed by Al sputtering and annealing at 470-520°C in a nitrogen atmosphere. There is a Ti/Al sputtered Schottky gate contact: the device is an inverted structure (there is modulation doping below the active channel) so the sheet density in the channel can be varied with gate bias.[15]

5.3.2. Commercial Scale Siemens Device

The common length scale of devices created for research is of the order of at least a few hundred microns (as with 55/53, and the devices investigated in following chapters) but devices fabricated in commercial integrated circuits are at most a few microns in size (depending on the application)[10,50] so it is important to investigate such small devices in a research context. Also, this second device was grown with a commercial reduced-pressure CVD process, not a research-laboratory MBE system.

The device is very sensitive, due to its small size, both in terms of the geometry of the Hall bar and the thickness of the heterostructure layers: there is a high degree of strain associated with a $Si_{0.5}Ge_{0.5}$ layer grown pseudomorphically on silicon which limits the thickness of the alloy layer.

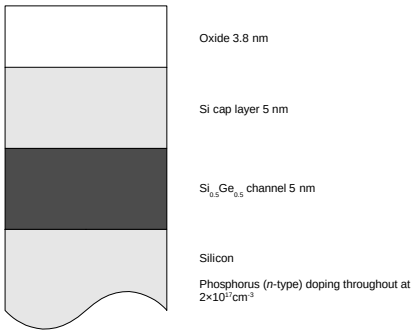

FIG. 5.2: The structure of the Siemens device.

The structure of this device is shown in Figure 5.2. The samples used to obtain the results presented here were grown using CVD, by Siemens, and took the form of gated Hall bars. A 5 nm layer of $Si_{0.5}Ge_{0.5}$ was grown pseudomorphically on Si, capped by 5 nm of Si. Whilst this alloy thickness exceeds the Matthews-Blakeslee equilibrium critical thickness of ∼4 nm,[51–53] and the presence of the cap stabilizes the epitaxial alloy only slightly against strain relaxation,[9] accounting for interactions between dislocations more correctly gives an equilibrium critical thickness of 8 nm.[54]

An n-type (phosphorus) dopant level of $2\times10^{17}\text{cm}^{-3}$ is present throughout: the structure is not modulation doped but the donor concentration may be higher than this deep below the active channel. The gate oxide is 3.8 nm thick. The Hall bars were created on a commercial device scale; they are 2.5 μm wide and 12.4 μm long overall, with the terminals spaced 6.6 μm apart symmetrically. This means that there is no appreciable correction to the Hall potential due to the device geometry.[26]

5.4. Experimental Considerations

5.4.1. Current Heating

Measurements of the Hall effect at currents of 20 nA, 50 nA and 100 nA (at 350 mK) were performed on device 55/53 (see Figures 5.8 and 5.9) and the results are

discussed in section 5.4.2. IV measurements showed no change in resistance of the device as a function of current, up to at least 100 nA, implying that there was no heating effect from the current.

During measurements on the Siemens device, a current of 10 nA was generally used, to avoid heating the carriers above the lattice temperature. The small size of the device implies a high current density. The energy loss rate per carrier was calculated to be of the order of 10^{-17}W, but this should be low enough that thermal carrier diffusion to the contact regions equalises the carrier and lattice temperatures.[55]

5.4.2. ESD Protection

The breakdown field of silicon dioxide is of the order of $10^8 \mathrm{Vm}^{-1}$.[56] This means that the gate oxide of the Siemens devices, which is a few nanometres thick, is in danger of being compromised when subject to potential differences of just a few tens of volts. Not only must great care be taken to protect the devices from electrostatic discharge (by storing the devices with the pins of the 8-pin DIL packages embedded in conducting foam or aluminium foil-wrapped polystyrene, for example, as is common practice with CMOS integrated circuits) but experimental procedures must be designed so as to keep voltages at safe levels at all times.

It is not sufficient merely to short-circuit the gate and a single other contact: since the devices operate in enhancement mode, when V_{GS} is zero all the Hall terminals are electrically isolated from each other, and a large enough potential between any one Hall terminal and the gate will destroy the device. Protection was also considered necessary during the measurement procedure, since otherwise devices were surviving no more than a few days.

The protection circuit shown in Figure 5.3 was built and incorporated into the junction box between the cryostat system and the measurement electronics. The gate was protected with a pair of Zener diodes (each rated at 2.7 V wired back-to-back) and a low-pass filter circuit with a cut-off frequency of around 200 Hz; each of the six Hall terminals was protected by a pair of silicon diodes wired in opposed parallel. The IV characteristics of the gate circuit are shown in Figure 5.4, and the IV characteristics of a sample contact circuit are shown in Figure 5.5. For voltage levels considered reasonable in terms of normal device operation and measurement (a few millivolts for each Hall terminal, and up to 2.5 V on the gate) the protection circuitry should have no bearing on the results. The gate protection circuitry does, however, prevent measurement of gate leakage current.

Additionally, all terminals were short-circuited to ground between measurements.

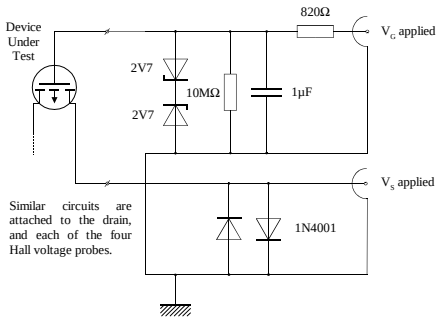

FIG. 5.3: The circuit diagram of the ESD protection system: the gate is protected from short spikes and sudden changes by a low-pass filter (with a time constant of the order of a few milliseconds) whilst the back-to-back Zener diode arrangement ensures that V_{GS} is limited to a safe value. The two current and four voltage contacts are each protected by a pair of silicon diodes in anti-parallel which prevent potentials of more than a few hundred millivolts from building up.

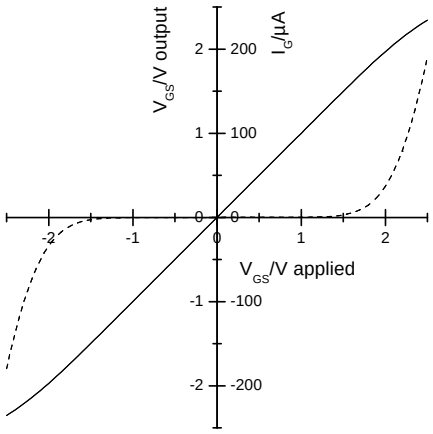

FIG. 5.4: The direct-current transfer characteristic of the gate protection circuit in Figure 5.3. It can be seen that as increasing gate voltages are applied, current (dashed line) begins to flow to earth through the Zener diodes, and the voltage presented to the gate of the device (solid line) is reduced. For small applied voltages, the leakage current is negligible.

5.5. Results from 55/53

As this device uses a Schottky gate contact, its operation relies on the silicon cap behaving as a non-conducting dielectric. However, at temperatures above around 60 K current can flow more or less freely between

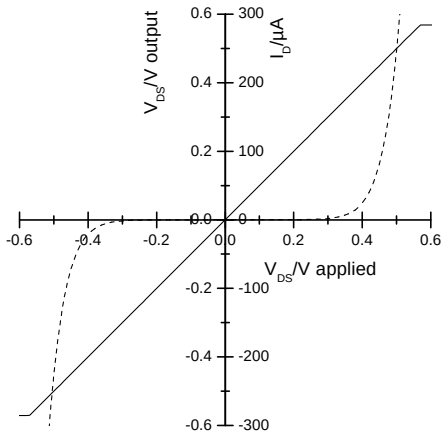

FIG. 5.5: The direct-current transfer characteristic of the contact protection circuit in Figure 5.3. As the applied voltage reaches a few hundred millivolts the current (dashed line) through the silicon diodes quickly reaches the 1 mA compliance limit of the HP4148 performing the measurement and the voltage presented to the device (solid line) saturates.For small applied voltages, the leakage current is negligible.

the gate and the active channel restricting the useful measurement range to below this temperature. At temperatures below 60 K the gate system behaves as a Schottky diode so the range of gate voltages is limited. The 350 mK characteristics of the gate are shown in Figure 5.6.

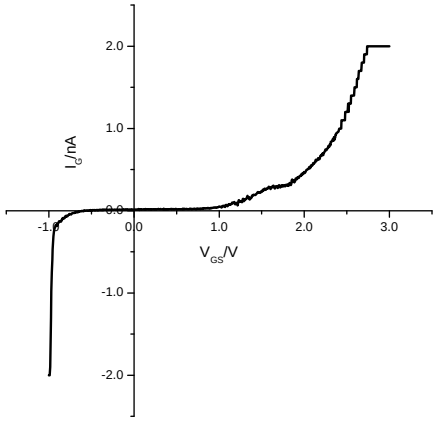

FIG. 5.6: The IV characteristics of the gate of the device on wafer 55/53 at 350 mK, showing leakage similar to that shown by a Schottky diode. At 2.7 V, the current reaches the set compliance limit of the HP parameter analyzer. Operation was limited to gate voltages in the range -1 V to +1 V.

On the basis of this, gate voltages were kept within the range -1 V to +1 V. The relationship between gate voltage and sheet density was roughly linear, as will be seen.

5.5.1. Resistivity as a Function of Temperature

A summary of the ρ(T) characteristics of this device (in the absense of magnetic field) is shown in Figure 5.7. For larger, "metallic" sheet densities, ρ(T) weakly saturates as the temperature goes to zero, but features a local maximum at $T_F/3$, where T_F is the Fermi temperature as defined in Equation 3.5. However, as will be discussed in terms of similar data from the Siemens device, current theories of the behaviour of ρ(T) are based on screening and interactions which should lead to a minimum in resistance at a temperature of the order of 1 K, and a divergence at zero temperature.[21]

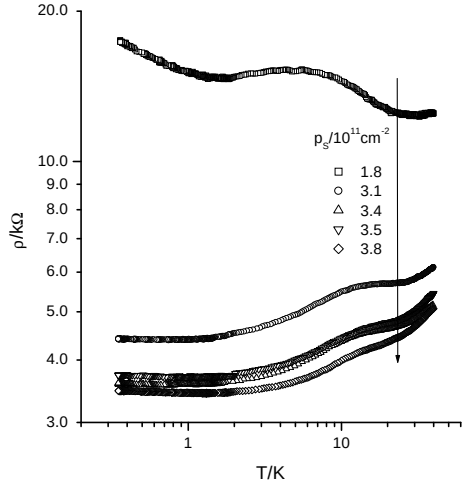

FIG. 5.7: A summary of the temperature dependence of the resistance of the device on wafer 55/53, at varying sheet density. High sheet densities (greater than $3\times10^{11}\text{cm}^{-2}$) are "metallic" in that the resistance increases with temperature, and in these cases the resistance seems to saturate as temperature decreases. At the lowest density (of $1.83\times10^{11}\text{cm}^{-2}$) the resistance is increasing weakly as temperature decreases.

5.5.2. Mobility

Hall measurements were performed, and the single-carrier analysis described in Chapter 4 was applied. There should be no complications from multiple carrier gases since, for the device's gate to operate, the unstrained silicon layers must be frozen out and this includes the doped layer. The only free carriers available for conduction should be those in the 2DHG. Results for the sheet density as a function of gate voltage at 350 mK are shown in Figure 5.8. A field of 0.5 T was used throughout: this was deemed large enough to produce a

useful Hall signal whilst also minimizing the quantum effects which will be discussed in a following section.

Figure 5.9 presents the (Hall) mobility as a function of the (Hall) sheet density at 350 mK. Lower current drives yield results which are noisier but apparently better: this may be due to current heating effects. However, since this is a DC measurement with the current only flowing in one direction along the Hall bar, there may be an absolute offset in measured voltages which remains uncorrected and therefore leads to a systematic error.

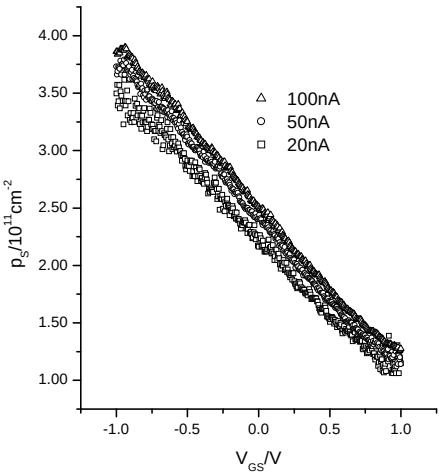

FIG. 5.8: The gate voltage dependence of the sheet carrier density in device 55/53, from the Hall effect at 350 mK, using a field of 0.5 T. Results from three different drive currents are shown: lower currents are noisier since voltage signals are smaller.

Since mobility is increasing with sheet density for the sheet density range explored, it is to be expected (from Chapter 3) that the dominant mobility-limiting scattering process is interface impurity scattering, with surface roughness beginning to have an impact towards the higher end of the range. However, the analysis described in Chapter 3 does not take into account the quantum corrections which are described in the following section so detailed calculations based on this are out of the question.

The 50 nA and 100 nA mobility results in Figure 5.9, for sheet densities greater than $2.0\times10^{11}\text{cm}^{-2}$, seem to show slight oscillations. (The noise in the 20 nA results presumably masks them.) This may be a feature of the quantum nature of the charge transport which can only be seen at millikelvin temperatures, when the mobility has been measured at a high number of sheet density values.[57,58] Oscillatory features can also be seen in Figure 5.8.

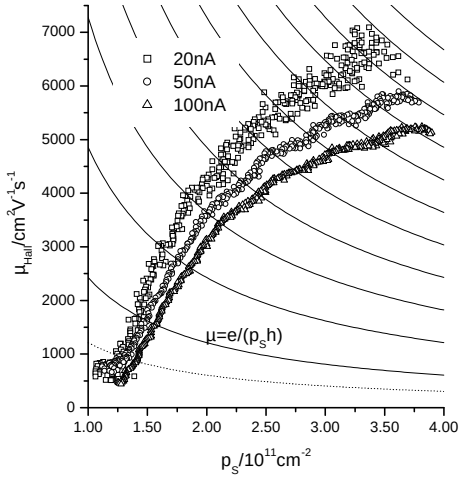

FIG. 5.9: The relationship between mobility and sheet density (calculated using the Hall effect at 350 mK with a field of 0.5 T) for device 55/53. As in Figure 5.8, lower current drives produce noisier results, but here mobility also appears to increase (for a given sheet density) as current drive decreases. Solid lines are $\mu = n\frac{e}{p_s h}$ where n is an integer, the dotted line is for $n = \frac{1}{2}$.

5.5.3. Low-Field Magnetoresistance

As described in Section 4.1.4, at low temperatures (such that $k_BT \ll E_F$) weak localization, particle-hole interactions (including the Zeeman effect) should generate characteristic low-field magnetoresistance behaviour.

In the absense of magnetic field, quantum interference effects will lead the path of a charge carrier to form a closed loop provided that the rate at which its direction changes around the loop (the elastic scattering rate τ^{-1}) is much greater than the rate at which its energy and therefore quantum phase changes (the dephasing rate τ_ψ^{-1}) and the coherence which leads to the quantum interference is broken. Part of the probability current of the charge carrying particle is localized in this loop, so the conductivity of the material is reduced.

The largest localization loops (essentially those which take a carrier τ_ψ to circumscribe) cannot form in the presence of a small magnetic field, since the flux enclosed by the loop leads to a phase shift which ruins the constructive quantum interference. As the magnetic flux density is increased, the maximum feasible loop size decreases until even the smallest localization loops (limited by the carrier requiring a few τ to scatter around the whole loop) enclose too much flux. This causes the conductivity of the material to increase with magetic field. Temperature is not an explicit feature of Equation 4.31 but the dephasing and elastic scattering times and also the transport scattering time (which features in the mobility, and therefore the diffusion constant) have temperature dependences which will be explored as the data is analyzed.

As the field continues to increase, the interaction correction due to Zeeman splitting between up and down spin states becomes comparable to k_BT (the up-spin and down-spin levels are resolved over thermal broadening) and the conductivity is reduced (Equation 4.37) by scattering between them.

Particle-hole interaction corrections to the conductivity (Equation 4.34) are temperature rather than field dependent, but would contribute slightly to the magnetoresistance when the corrected conductivity tensor is inverted. However, as seen from the data in Figure 5.7 and its discussion in a Section 5.5.1, interaction effects are not expected to be important.

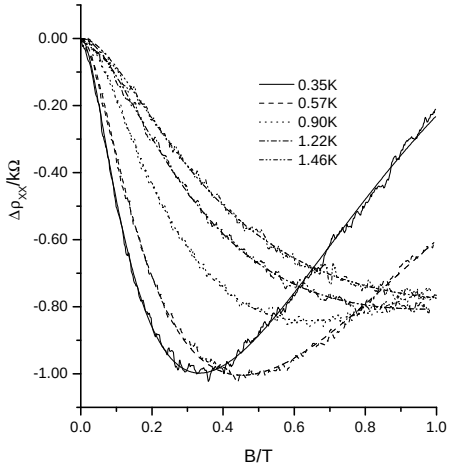

FIG. 5.10: Low-field magnetoresistance of 55/53 at a sheet density of $1.8\times10^{11}\text{cm}^{-2}$, offset by the zero-field value of the resistance (15 kΩ at 0.35 K falling to 13 kΩ at 1.46 K imply $k_Fl \sim 1.7$). The noisy lines are data, the smooth lines are fits to Equation 4.38 produced with the parameters given in Figure 5.11.

Some results in this regime are shown in Figure 5.10. The gate voltage is such that the sheet density is $1.8\times10^{11}\text{cm}^{-2}$. At the lowest temperatures, the suppression of weak localization by the magnetic field causes the resistivity to fall to a minimum before other effects cause it to rise again. As temperature increases, the minimum in resistivity becomes weaker and moves to higher fields, and then is lost.

The zero-field resisitivity ρ_0 for this sheet density ranges from 15 kΩ at 0.35 K to 13 kΩ at 1.5 K: the fact that ρ_0 is falling as temperature increases, coupled with its high value, means that the system is in a non-metallic regime. This can be characterized by

$$\rho_0 = \frac{h}{e^2}\frac{1}{k_Fl} \tag{5.1}$$

where h/e^2 is the von Klitzing quantum resistance standard (25812.8 Ω) and k_Fl is the product of the Fermi wavenumber and the mean free path of the carriers. Here, $k_Fl \sim 2$ but the truly metallic regime requires $k_Fl \sim 10$.[59] The solid lines on Figure 5.9 represent constant resistivity values such that $\mu = n\frac{e}{p_sh}$ where n is an integer. The dotted line is for $n = \frac{1}{2}$. This can be related to the metallic or insulating behaviour of the 2DHG, as described in Equation 5.1.

The zero-field resisitivity has been subtracted from each of the datasets in Figure 5.10 so that the correction to the resistance (as a function of magnetic field) can be clearly seen for each temperature. Noisy lines are actual data, and smooth lines are fits generated using Equation 4.38 which encapsulates corrections to the zero-field resisitivity due to weak localization, carrier interactions and the Zeeman effect at low fields. The fitting parameters which result are shown in Figure 5.11. (An effective mass of 0.2 m_e has been assumed throughout, on the basis of the Shubnikov–de Haas analysis in section 5.5.4. Generally, an increase in the assumed effective mass will lead to a proportional increase in the resultant relaxation rate from Equation 3.3.)

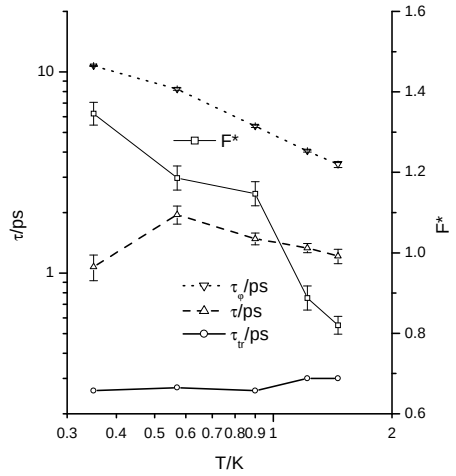

FIG. 5.11: Fitting parameters used to create the smooth lines in Figure 5.10.

Four fitting parameters are shown in Figure 5.11, but τ_{tr} is actually found from the Hall mobility. The low-field negative magnetoresistance is caused by suppression of weak localization, so τ and τ_ψ are found from fitting this part of the curve. F^* is found from the positive magnetoresistance contribution (mainly due to the effect of Zeeman splitting on interactions) which can be seen when weak localization is suppressed. The Landé g-factor is fixed at 4.5.[33]

This fitting has been repeated at other sheet densities, and the results are summarized as follows: Figure 5.12 shows how the dephasing time τ_ψ is roughly proportional to T^{-p}, where $p \sim 1$ as can be seen in Table 5.1.[21]

Figure 5.13 shows how, for most sheet densities, the elastic scattering time τ decreases with temperature. However, for the lowest sheet density (the pink curve)

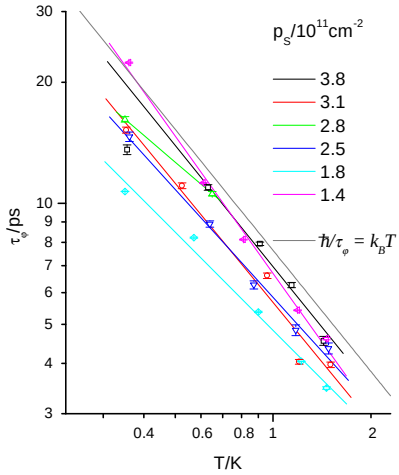

FIG. 5.12: The temperature dependance of the dephasing times summarized from data similar to that presented in Figure 5.10. The fits shown are for $\tau_\psi \propto T^{-p}$ and values for p are given in Table 5.1.

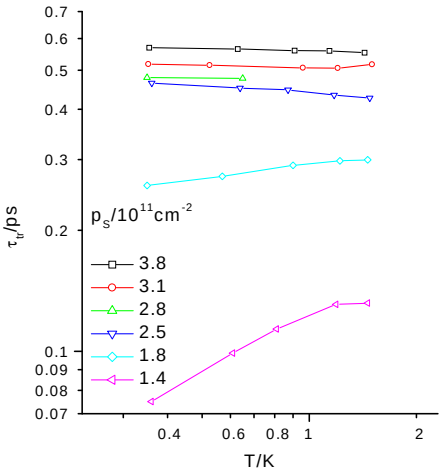

FIG. 5.14: These values of the transport scattering lifetime were calculated from the Hall mobility at a field of 0.5 T, and are provided for comparison with Figure 5.12 and Figure 5.13.

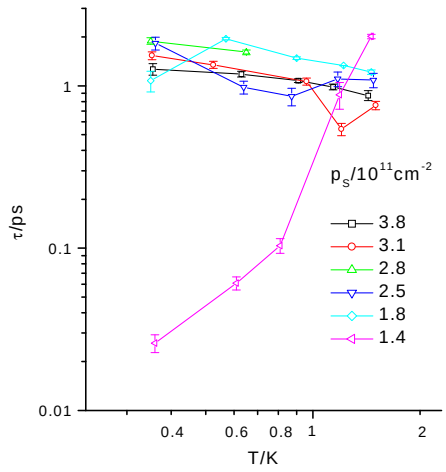

FIG. 5.13: A summary of the elastic scattering times from data similar to that presented in Figure 5.10.

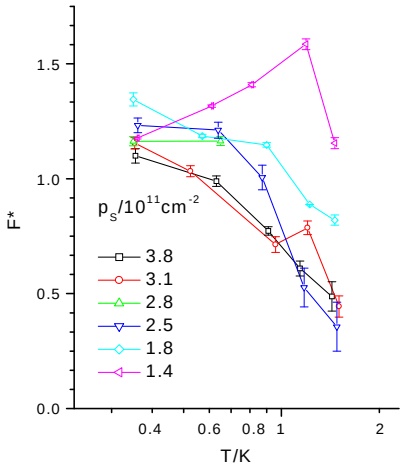

FIG. 5.15: Values of the renormalized screening paramter, summarized from data similar to that presented in Figure 5.10. According to its theoretical derivation, F^* can only take values between zero and ∼0.866.[59]

where $k_F l \sim 1$ and the 2DHG is on its way to becoming strongly localized, τ decreases sharply with decreasing temperature. The mobility of the hole gas was found at each sheet density and temperature, employing the Hall effect as described in the preceeding section. Using Equation 3.3, the transport scattering times τ_{tr} were obtained and plotted in Figure 5.14.

Lastly, the screening parameter F^* (defined in Equations 4.35 and 4.36) is shown in Figure 5.15. It is a cause for concern that F^* appears to vary with temperature: it is clear from its definition that it should only be a function of the sheet density (assuming factors such as the effective mass remain constant). In fact, its value should be at most 0.866 (in the limit of no carriers and therefore no screening) so the values returned from the fitting, being often greater than unity, are a sign of some deficiency in either the theory itself or its application.

The parameter F^* appears in both the temperature-dependent conductivity correction from interactions (Equation 4.34) and the magnetoconductivity correction from Zeeman splitting (Equation 4.37). Since these are both *conductivity* corrections, there is no reason to suggest that the forms used for F^* in each case should not

TABLE 5.1: Dephasing times obtained from fits (similar to those shown in Figure 5.10) were plotted against temperature for each sheet density in Figure 5.12. Fits to these data yield the index p which should be close to unity.[21]

p_s	p
[10^{11}cm^{-2}]	[$\tau_\psi \propto T^{-p}$]
3.79	1.0±0.2
3.09	1.0±0.2
2.53	0.9±0.1
1.83	0.8±0.1
1.41	1.1±0.1

be the same. (The conductivity screening parameter is sometimes referred to explicitly as $\tilde{F}_\sigma$ whilst the form of the screening parameter used to calculate, for example, corrections to the specific heat $\tilde{F}_c$ is different).[43] However, inconsistencies between the values of F^* produced by fitting different conductivity corrections to different results have been noted.[21,49]

Analysis of the Dephasing Time, τ_ψ

For a 2-dimensional system, theory predicts that $\tau_\psi \propto 1/T$ and that this should be generally independent of any other specific device parameters.[60,61] This is demonstrated in Table 5.1 and represented by the grey $\hbar/\tau_\psi = k_BT$ line in Figure 5.10.

This result is more usefully expressed as $\hbar/\tau_\psi \simeq k_BT/(k_Fl_\psi)$ where k_Fl_ψ is the dimensionless conductivity of the system if limited by dephasing alone (in the absence of magnetic field). However, since $l_\psi = \sqrt{D\tau_\psi}$ then a self-consistent solution is:[62]

$$\hbar/\tau_\psi = \frac{k_BT}{k_Fl}\ln(2k_Fl) \tag{5.2}$$

Dephasing times predicted in this case are larger than those in Figure 5.12 and if this were true then the data in Figure 5.12 for each sheet density should lie above the grey $\hbar/\tau_\psi = k_BT$ line, generally further above it the greater the sheet density.

In any case, these forms predict a divergence in τ_ψ as temperature decreases suggesting that at zero temperature there would be no dephasing at all, and no true metallic behaviour. Experimental results seem to suggest that this is not the case, and τ_ψ in fact saturates as temperature decreases.[63] Evidence for this can be seen most clearly in the results in Figure 5.12 for a sheet density of 3.8×10^{11}cm^{-2} (in black) where the data point at the lowest temperature value spoils the fit. This could be due to heating or external electromagnetic noise effects,[62] or it may be the case that a finite dephasing rate even at zero temperature is fundamental and results in the familiar (but theoretically forbidden according to traditional scaling theories of localization) metal-insulator transition in two-dimensional systems.[64]

Carrier Interactions

Two kinds of scattering are expected to contribute to the dephasing rate: (a) direct carrier-carrier interaction involving energy transfer of the order of k_BT and *large* momentum transfer, and (b) the interaction of one carrier with the time-fluctuating electromagnetic field (Nyquist noise) of all the other carriers, involving *small* momentum transfer.[65] The carrier-carrier relaxation rate is given by[66]

$$\frac{1}{\tau_{cc}} = \frac{F^2\pi(k_BT)^2}{2\hbar E_F}\ln\left(\frac{E_F}{k_BT}\right) \tag{5.3}$$

where the term inside logarithm is replaced by $\frac{E_F}{\hbar/\tau}$ if the carrier energy around the Fermi line is smeared by disorder rather than by temperature. The Nyquist relaxation rate is given by[67]

$$\frac{1}{\tau_N} = \frac{k_BT}{4E_F\tau}\ln\left(\frac{2k_BT}{\hbar/\tau_\psi}\right) \tag{5.4}$$

(compare Equation 5.8) and in this case the term inside logarithm is replaced by $\frac{E_F}{\hbar/\tau}$ in the special case that $\tau_\psi = \tau_N$. Mathiesen's rule (Equation 3.4) would then suggest that:*

$$\frac{1}{\tau_\psi} = \frac{1}{\tau_N} + \frac{1}{\tau_{cc}} \tag{5.5}$$

When p_s=3.8×10^{11}cm^{-2} at 1 K, Equation 5.3 gives a carrier-carrier scattering time of 63 ps. (The thermal energy k_BT is greater than the level broadening $\hbar/\tau$ so the relevant cutoff is the thermal energy.) Equation 5.4 gives a Nyquist scattering time of 330 ps. The total dephasing time from Equation 5.5 is 53 ps, much greater than the measured dephasing time of 7.2 ps. Calculations of the dephasing rate at other temperatures and sheet densities confirm that carrier-carrier and Nyquist scattering do not account for most of the dephasing.

This suggests that the actual, measured dephasing time $\tau_{\psi m}$ is given by

$$\frac{1}{\tau_{\psi m}} = \frac{1}{\tau_N(\tau_\psi)} + \frac{1}{\tau_{cc}} + \frac{1}{\tau_x(\tau_\psi)} \tag{5.6}$$

where $(\tau_x)^{-1}$ represents the dephasing rate (which in general depends on the overall dephasing time due to lifetime broadening effects) from other processes. These other processes would seem to dominate in this device, contradicting the traditional theories which state that carrier-carrier scattering (in particular, electron-hole pair production) is the only way in which a carrier can lose energy at low temperatures in a semiconductor.

Two alternative theories follow, in which the dephasing rate is finite even at zero temperature.

---

* It may be argued that since the dephasing time is a function of itself then Mathiesen's rule should not strictly have meaning.

Zero Point Fluctuations

It has been suggested that this saturation of the dephasing time in the limit of zero temperature is not due to heating effects or magnetic impurities (as had previously been suggested) but rather due to zero-point fluctuations of phase coherent electrons. This leads to[63,68]

$$\tau_\psi = \tau_0 \tanh\left(\alpha\pi\sqrt{\frac{\hbar/\tau_0}{k_B T}}\right) \tag{5.7}$$

where α is a constant of order unity and τ_0 is the zero-temperature saturation value of the dephasing time, which has so far been calculated only for the one-dimensional case.[68] Equation 5.7 apparently produces successful fits to data from a wide range of one and two-dimensional systems (with relevant modifications) once electron-phonon scattering is added $\tau_{\psi ep} = [\tau_\psi^{-1} + \tau_{ep}^{-1}]^{-1}$ with a form $\tau_{ep} = A_{eq}/T^3$.[63]

This fits well to the p_s=3.8×10^{11}cm^{-2} (in black) data in Figure 5.12 with the values τ_0=16.5±1.3 ps, α=0.35±0.05 and A_{ep}=25±3 psK^3. This value for the strength of the electron-phonon interaction is anomalous: at the temperatures discussed here (∼1 K) electron-phonon scattering is not important in silicon or silicon-germanium.† The fact that in the limit of temperatures greater than $\hbar/(k_B\tau_0)$ (0.4 K in this case) $\tau_0 \propto T^{-\frac{1}{2}}$ unless an extra term is added to Equation 5.7 would seem to be a problem. Fits to data at other sheet densities are less successful: the uncertainties on the returned parameters are greater than the values of the parameters themselves.

High Frequency Fluctuations

Alternatively, the dephasing rate can be calculated within existing weak localization theory but incorporating high frequency ($\omega \gg k_B T/\hbar$) quantum fluctuations to give:[69]

$$\tau_\psi = \frac{e^2}{2h\sigma\tau}\left[1 + 2\frac{k_B T}{\hbar/\tau}\ln\left(\frac{k_B T}{\hbar/\tau_\psi}\right)\right] \tag{5.8}$$

The saturation value of the dephasing time in this case is therefore (from Equation 5.1) simply $\tau_0 = 2k_F l\tau$. For p_s=3.8×10^{11}cm^{-2}, inspection of Figure 5.13 suggests that a zero temperature elastic scattering time of 1.2 ps is reasonable, and since $k_F l = 8$ at this sheet density then τ_0 is found to be ∼20 ps. At other sheet densities, though, a good value for $\tau(T = 0)$ seems less obvious. The temperature dependence of Equation 5.7 is weaker than $1/T$ and, as with Equation 5.7, electron-phonon scattering could be explicitly added. However, as noted below, since the dephasing time is a function of itself in Equation 5.8 then Mathiesen's rule (Equation 3.4) may not strictly be applicable. Equations 5.2, 5.7 and 5.8 are compared in Figure 5.16.

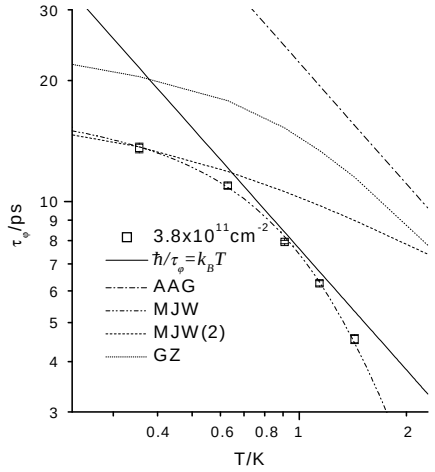

FIG. 5.16: A summary of dephasing times, calculated by various theories and compared to data from Figure 5.12. The AAG line is based on Equation 5.2. MJW was produced by fitting Equation 5.7 including phonons, but as the text suggests requires unreasonably strong electron-phonon scattering; MJW(2) is Equation 5.7 alone. Equation 5.8 was used to produce the GZ line.

Whether it is necessary to modify the fundamentals of weak localization theory is controversial.[70–73] It is argued that the zero-point fluctuations which lead to Equation 5.7 (or the unoccupied high-frequency modes which lead to Equation 5.8) cannot cause any dephasing, since the energy of a harmonic oscillator may not be available to be transferred in situations where the level spacing is much greater than $k_B T$, and that the saturation of the dephasing time is due to external electromagnetic noise of microwave frequencies.[72] It is further suggested that only particle-particle scattering is relevant to dephasing (the Coulombic interaction being an elastic process) and any finite dephasing at zero temperature predicted theoretically is based on a profoundly incorrect calculation.[62] Alternatively, it is argued that the zero-temperature dephasing time is lost in calculations which are performed incorrectly.[73]

The form of the mobility as a function of sheet density (Figure 5.9) and the calculations described in Chapter 3 suggest that the scattering that limits the transport is due to screened interface impurities (at low sheet densities) and surface roughness (as sheet density increases) plus quantum effects which will be discussed in Section 5.6.2.

These contribute only to elastic scattering, and therefore momentum relaxation, rather than the dephasing through inelastic processes. (It is puzzling, though, that the elastic scattering times seem to be 2 or 3 times larger

† See, for example, the acoustic-phonon limited mobility at 25K in Figure 5.28

than the momentum relaxation (transport) time). Recent experimental studies confirm that in high mobility silicon-germanium systems, scattering is dominated by potential fluctuations and the loss of screening rather than by carrier-carrier interactions.[74]

With finite elastic scattering rates but no dephasing, at zero temperature a 2D system would be completely localized. The tendency of the resisitivity to saturate at a finite value as temperature decreases, for large enough sheet density, contradicts this. (One caveat is that insulating behaviour could set in at a temperature an order of magnitude lower than has been explored in this study.)

The implication is that the dephasing rate is proportional to the sheet density: in high density systems which remain metallic as temperature decreases, the dephasing rate must be prevented from continuing to increase as T^{-1}. In Equations 5.3 and 5.4, the dephasing time is proportional to the Fermi energy, meaning that high density systems would have longer dephasing times than low density systems. The Nyquist scattering rate (Equation 5.3) also incorporates the elastic scattering time which increases as impurities are screened by an increasingly dense carrier gas. The carrier-carrier scattering rate (Equation 5.4) incorporates the screening parameter F which decreases as sheet density increases, increasing the carrier-carrier scattering time further.

The alternative theories presented, which give a finite dephasing rate at zero temperature from either zero-point or high-frequency quantum fluctuations, do not seem to reproduce the T^{-1} result in any limit. The density dependence of the dephasing time (due to zero-point flucuations) in Equation 5.7 is presumably incorporated into τ_0 which remains uncalculated for 2D systems. Equation 5.8 does ensure that the dephasing time (due to high-frequency fluctuations) increases with the conductivity of the system.

5.5.4. High-Field Magnetoresistance

The magnetoresistance of the device in the high-field, low temperature regime (where $\mu B \sim 1$ and $\hbar\omega \sim k_B T$ as described in section 4.1.3) is shown in Figure 5.17. The former shows conventional Shubnikov-deHaas oscillations and the Quantum Hall effect. A filling factor of unity is reached at the highest available fields. At low fields, minima in resistivity appear at odd values of the filling factor. Spin splitting occurs at around 5 T, so $\nu = 2$ is also a minimum for all sheet densities but $\nu = 4$ only features a weak minimum at the highest sheet density.

For lower sheet densities than Figure 5.17, where the 2DHG is insulating in character, Figure 5.18 shows how a high resistance state appears between filling factors 1 and 3. This can be compared with the last ρ_{xx} maximum visible in Figures 5.17 and 5.18 (where the 2DHG is metallic) which is not particularly pronounced. The horizontal scale is linear in applied field but scaled so

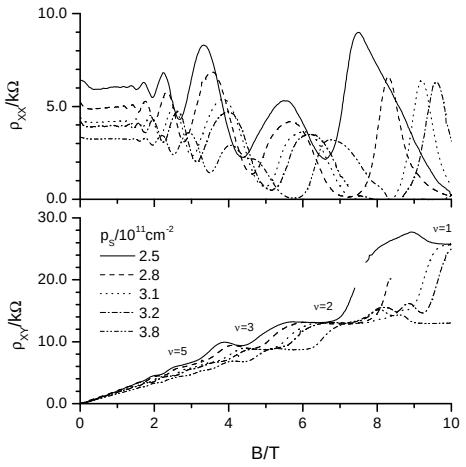

FIG. 5.17: Magnetoresistance of 55/53 in the low temperature, high field regime where Shubnikov–de Haas oscillations and the Quantum Hall Effect are visible. Sheet densities (calculated from the Hall effect) are such that the 2DHG is metallic in character. In general, the sheet density calculated from the positions of the minima and maxima in the oscillations will not exactly match that from that from the Hall effect.

that the low-field oscillations coincide:

$$\frac{1}{\nu} = \frac{eB}{n_{SdH} h} \tag{5.9}$$

where ν is the filling factor described in section 4.1.3. In general, the sheet density n_{SdH} found through Equation 5.9 will not perfectly match that found from the low-field Hall effect. This can be seen in Figure 5.19.

The magnetic-field-induced metal-insulator transition is well known in bulk semiconductors where it is caused by squeezing of the electronic wavefunction at localization centres.[75] In a 2-dimensional system the off-diagonal conductivity σ_{xy} is proportional to the filling factor but a divergence in ρ_{xx} implies vanishing σ_{xy}. This means that the large maximum in ρ_{xx} at a filling factor of $\sim$1.6 cannot be explained in terms of the Quantum Hall Effect.

This has been taken as evidence for Wigner crystallization, but if $\nu > 1$ then at least one totally occupied magnetic level should exist below the Fermi level with a band of delocalized states located at the magnetic level centre, giving a non-zero σ_{xy}.[75,76]

The insulating phase between filling factors of 1 and 2 has been seen in p-type silicon-germanium systems[77–80] but is not predicted in the global phase diagram for 2D systems.[81] Magnetic-field-induced phase transistions have been observed in the GaAs/AlGaAs system, but insulating states only tend to form at $\nu \ll 1$.[82–85] The unusual energy level degeneracy in p-type silicon germanium may be central to its formation.[80]

More recently, it has been suggested that the similar transition from the insulating state to the metallic state

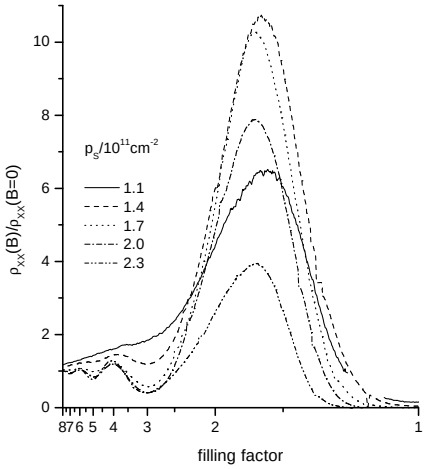

FIG. 5.18: Magnetoresistance of 55/53 in the low temperature, high field regime where sheet densities are such that the 2DHG is insulating in character. The vertical axis has been rescaled by the magnitude of the zero-field resistance for each sheet density: in absolute terms, the resistance peak for the lowest sheet density is of the order of 1 MΩ, nearly two orders of magnitude greater than the peak in the highest sheet density (just over 20 kΩ). The horizontal axis has been scaled using the Shubnikov–de Haas rather than the Hall sheet density.

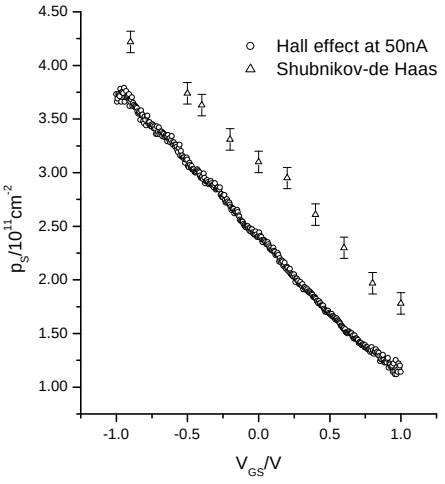

FIG. 5.19: A comparison of the sheet density extracted from the Hall effect (from Figure 5.8) with that found from Shubnikov–de Haas oscillations.

in silicon MOSFETs at zero field has a percolation nature, indicating electron localization to be the origin of the former state.[75,77,86,87]

Figure 5.20 shows how Shubnikov–de Haas oscillations (in the region where the field is low enough to avoid spin-splitting) decay with temperature. As described in Section 4.1.3, this data can be used to find the effective mass m^* of the carriers, and also the ratio between the quantum and transport scattering lifetimes, $\alpha = \tau_{tr}/\tau_q$ (from Equation 4.26). However, analysis of these results (or results at other sheet densities) does not yield self-consistent values for m^* or α. With $\alpha \approx 1$ the plot described in Equation 4.28 gives $m^* \approx 0.3m_e$. The gradient of the plot described in 4.30, however, yields $\alpha \approx 0.7$ and if the iterative process is continued then the values of both m^* and α diverge.

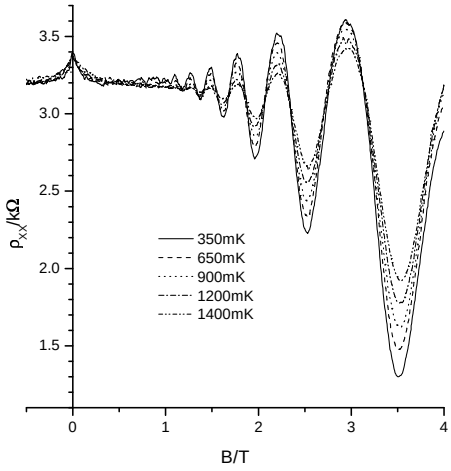

FIG. 5.20: Shubnikov–de Haas oscillations as a function of temperature, for fields low enough to avoid spin-splitting. Sheet density from the Hall effect is $3.79\times10^{11}\text{cm}^{-2}$ but from the period of the oscillations is $4.22\times10^{11}\text{cm}^{-2}$. The drive current is nominally 50 nA.

5.6. Results from the Siemens Device

5.6.1. Room-Temperature IV Characterization

Room temperature measurements of this device, shown in Figure 5.21, demonstrate p-channel MOSFET-like behaviour at drain-source voltages generally small enough to avoid pinch-off. The threshold voltage V_{TS} is around -1.0 V: in this region, at higher V_{DS}, the drain current increases quadratically as the gate-source voltage becomes more negative. Away from the threshold, the drain current is linear with both gate-source voltage and drain-source voltage.

5.6.2. Mobility as a Function of Sheet Density and Temperature

The results shown in Figure 5.22 and Figure 5.23 were taken during preliminary tests of a device, using the HP parameter analyzer and closed-cycle cryostat. (A

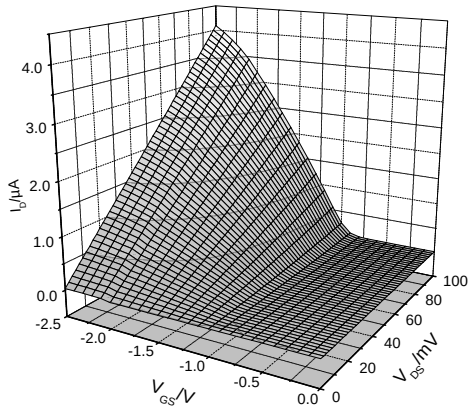

FIG. 5.21: Room temperature IV results for the Siemens device. The gate-source threshold Voltage V_{TS} is around -1.0 V. Drain current rises linearly with drain-source voltage for gate-source voltages more negative than this. Behaviour is consistent with that of a p-channel MOSFET.

constant drain-source potential of 10 mV was applied, whereas the usual method is to apply a constant drain current. This means that the drain current was of the order of a few microamps at times, which is rather high.)

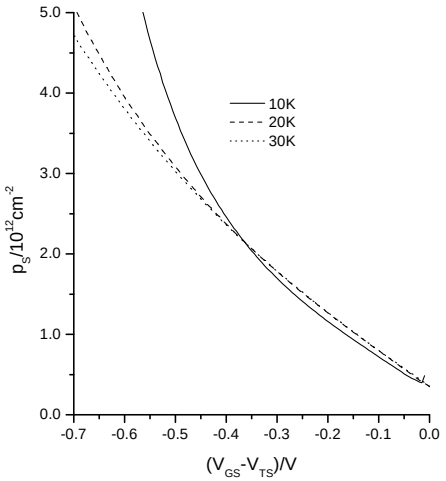

FIG. 5.22: Early Hall-effect results at low temperature: sheet density increases linearly as the gate voltage becomes more negative, but the linearity breaks down eventually. (The gate capacitance is ∼0.2 pF).

Later results are worse in terms of the actual peak mobility value as can be seen from the 350 mK results shown in Figure 5.24 and Figure 5.25: 1700 cm^{2}V^{-1}s^{-1} at 1.3×10^{12}cm^{-2} as opposed to 3300 cm^{2}V^{-1}s^{-1} at 1.2×10^{12}cm^{-2} at 10 K in Figure 5.23. (A current of 10 nA was used for these measurements, to avoid heating the carriers above the lattice temperature.[55]) Whilst

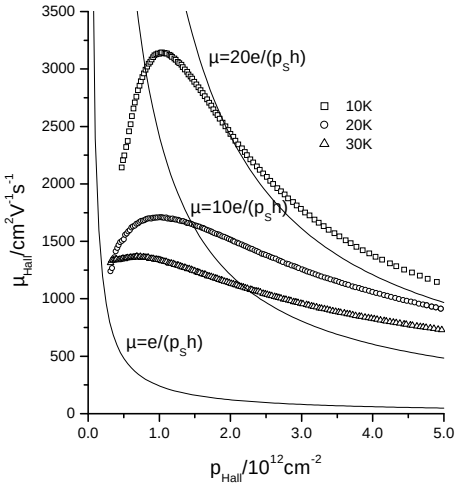

FIG. 5.23: More early Hall-effect results, showing typical behaviour for the mobility as a function of sheet density.

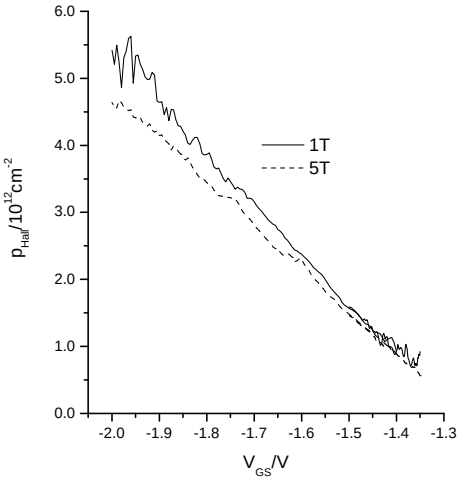

FIG. 5.24: Hall-effect results at 350 mK (with a drain current of 10 nA) showing the dependence of sheet density on gate voltage. The good agreement between the results at two different values of magnetic field suggest either that the Hall results are very close to the true ("drift") results, or that the mobility of the 2DHG is very small. From this, the capacitance of the gate is ∼0.3 pF. (The gate voltage was swept in both directions, and the results averaged, for each field.)

this mobility does not seem impressive compared with some of the best in p-type strained silicon-germanium alloys (for example, 15000 cm^{2}V^{-1}s^{-1} at 1.8×10^{12}cm^{-2} at 0.35 K in a normally-doped $Si_{0.87}Ge_{0.13}$ alloy grown at 950°C)[4] it is worth noting the relatively high sheet density (giving a sheet resistivity of 2.6 kΩ/□) and high germanium content of the pseudomorphic alloy layer.

A 4K mobility of 2500 cm^{2}V^{-1}s^{-1} at 6.2×10^{11}cm^{-2}

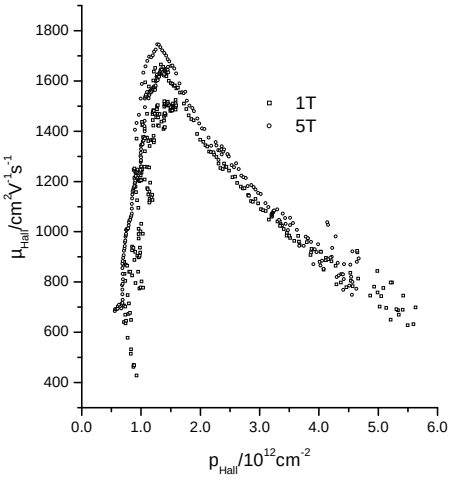

FIG. 5.25: The mobility at 350 mK peaks at more than 1700 $cm^2V^{-1}s^{-1}$ but this value seems worse than that seen in Figure 5.23. Again, values at two different fields match closely.

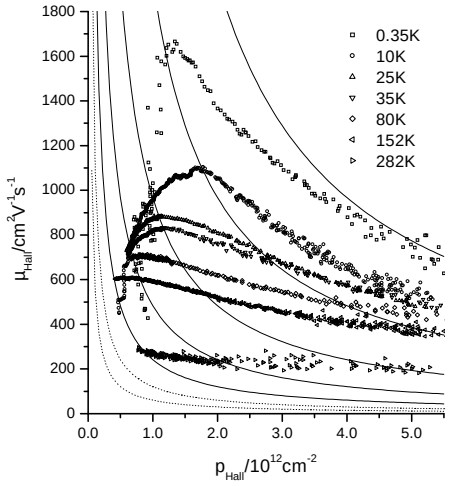

FIG. 5.27: Hall mobility as a function of sheet density, using magnetic fields of ±1 T, at temperatures from 350 mK up to almost room-temperature. Lines are shown at $\mu = 2^n \frac{e}{p_s h}$ with solid lines at n=0,1,2,3 and 4 and dotted lines at n=-1 and -2.

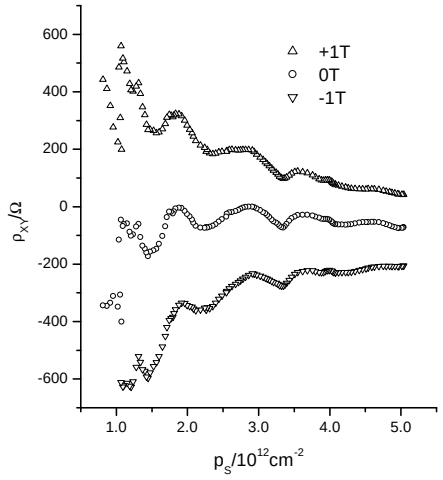

FIG. 5.26: From Equation 4.6, ρ_{xy} should be inversely proportional to p_s at a constant field. The x-axis has been scaled from gate voltage to sheet density using a straight-line fit to Figure 5.24, but the data clearly show anomalous quasi-oscillatory features. Importantly, these features appear not to depend on the magnetic field.

(and therefore a sheet resistivity of around 4kΩ/□) has been reported in a high-quality pseudomorphic $Si_{0.64}Ge_{0.36}$ p-channel device,[25] 77K hole mobilities of up to 3500 $cm^2V^{-1}s^{-1}$ at $3\times10^{12}cm^{-2}$ have been reported in a system with a $Si_{0.2}Ge_{0.8}$ channel grown on a virtual substrate.[7]

Some sort of noise can be seen in Figure 5.24, especially at the low and high sheet density ends of the ±1 T data. This may be related to the oscillations seen in Figure 5.8 and Figure 5.9, discussed in Section 5.5.2. The off-diagonal component of the resistivity tensor, ρ_{xy}, is shown in Figure 5.26 as a function of sheet density (assuming a linear dependence on gate voltage as in Figure 5.24). The relationship should be $\rho_{xy} = \frac{B}{p_s q}$ (Equation 4.6) but quasi-oscillatory features can be seen. These features clearly do not depend on the magnetic field (so are not, for example, due to the formation of Landau levels) and are repeatable (so are fluctuations as a function of gate voltage, not time). These fluctuations in ρ_{xy} may be a sign of fluctuations in ρ_{xx};[57,58] no fluctuations were visible in the ρ_{xx} data, but this may be because the absolute value of ρ_{xx} tended to be much larger than the scale of the fluctuations, drowning them out. These oscillations are largely suppressed in Figure 5.24 since they cancel out when positive and negative field results are combined.

Comprehensive results for the mobility as a function of sheet density from 350 mK to 282 K are presented in Figure 5.27. (There is essentially no variation between 350 mK and 1.4 K.) These are the subject of the following calculations.

Calculations of Mobility as a Function of Sheet Density

In Chapter 3, the mobility of a 2-dimensional carrier gas as a function of sheet density is discussed in terms of the mechanisms which may limit it. Calculations have been performed in similar studies.[18,24]

Experimental data at 25 K is presented in Figure 5.28, along with calculated values for the mobility produced using a C program written by A. I. Horrell* using the forms presented in References 18 and 24. The mobility

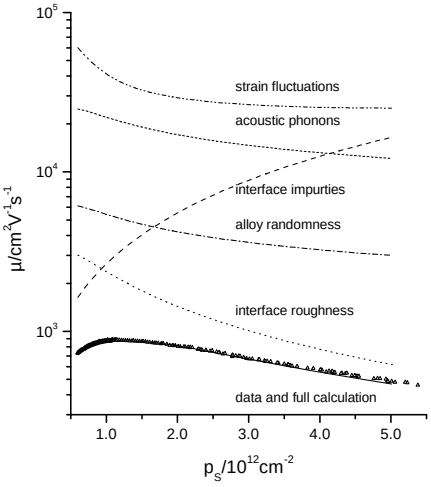

FIG. 5.28: Calculated mobility at 25 K, showing contributions from each of the relevant mechanisms: It can be seen that interface impurities limit the mobility at the lowest sheet densities, with surface roughness becoming dominant over most of the rest of the range. The optical-phonon-limited mobility is off the scale.

limit set by each scattering process is shown (apart from optical phonons, which are off the scale at this temperature) so that it is clear how interface impurities limit the mobility at low sheet densities and interface roughness becomes dominant as sheet density increases. At the lowest sheet densities, where the carriers are becoming increasingly localized, multiple scattering may need to be taken into account.[20]

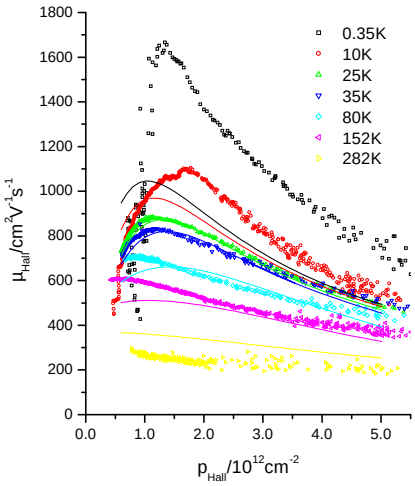

FIG. 5.29: Comparison of mobility calculated and measured across the whole temperature range. The parameters used for the calculation are those from the fit at 25 K, shown in Table 5.2.

The parameters used to produce these calculated mobilities are shown in Table 5.2. Those in italics were taken from the literature and held constant whilst those in normal type were varied to create a fit to the experimental data, based on reasonable initial values from the literature.

The fit was performed at 25 K because this temperature is well below T_F in this sheet density range (meaning that the carrier gas is degenerate and therefore that the Hall mobility needs no correction due to energy-dependent scattering mechanisms) but should be high enough that the transport is not be influenced by any of the factors described in Section 4.1.4.

The fit in Figure 5.28 is quite successful, demonstrating that the factors which limit the mobility are those which are open to improvement through better growth. Fundamental limits on mobility set by alloy scattering, for example, are several times greater than those measured. However, Figure 5.29 shows how calculations of the mobility at other temperatures measure up (with these same parameters) against experimental data and the agreement is less successful.

Using Equation 3.5 and the value of effective mass quoted in Table 5.2, the Fermi temperature in Kelvin is given by:

$$T_F = 93 p_s \tag{5.10}$$

with p_s in units of $10^{12}\mathrm{cm}^{-2}$.

This means that, for example, at 152 K the carrier gas would become non-degenerate as the sheet density decreases below $1.6\times10^{12}\mathrm{cm}^{-2}$ and the mobility and sheet density measured with the Hall effect would become increasingly affected by energy-dependent scattering and band-structure anisotropy.

The data and calculated values in Figure 5.29 at 152 K do indeed coincide reasonably well at high sheet densities (bearing in mind that no attempt has been made to fit any of the parameters in Table 5.2 which directly relate to the temperature dependence of the mobility) but diverge as the sheet density drops below around $2\times10^{12}\mathrm{cm}^{-2}$. At 282 K, where the carrier gas is more or less non-degenerate throughout the whole (Hall) sheet density range (for a T_F of 282 K, according to Equation 5.10, p_s would have to be $3.0\times10^{12}\mathrm{cm}^{-2}$) the Hall scattering factor can be found. From Equation 4.8, $\sigma = p_{Hall} q \mu_{Hall} = p_{Drift} q \mu_{Drift}$ so

$$r = \frac{p_{Drift}(\sigma)}{p_{Hall}(\sigma)} \tag{5.11}$$

This analysis yields a Hall scattering factor at 282 K of 0.68±0.04, averaged over the sheet density range. The influence of energy-dependent scattering can only lead

---

* Electronic Device Engineering Group, Department of Electronic & Electrical Engineering, Loughborough University.

TABLE 5.2: Table 5.2 Parameters used in the calculations presented in Figures 5.28, 5.29 and 5.30. Those in italics were taken from the literature and held constant whilst those in normal type were varied to create a fit to the experimental data, based on initial values from the literature.

Depletion charge density[18,21]	N_D	20.0	$\times 10^{11}\text{cm}^{-2}$
Interface impurity charge density[18,21]	n_i	2.8	$\times 10^{11}\text{cm}^{-2}$
Effective mass in the growth direction[14]	m_z	0.3	m_e
Effective mass for in-plane transport[14]	m_*	0.3	m_e
Relative permittivity[14]	ϵ_r	14.0	ϵ_0
Interface roughness height[18]	Δ	0.46	nm
Interface roughness correlation length[18]	Λ	1.20	nm
Alloy scattering interaction strength[18,24]	δE	0.6	eV
Lattice mismatch factor[18,88]	f	0.0125	
Poisson's ratio[18]	ν	0.28	
Acoustic phonon deformation potential[18,55,89]	Ξ_u	4.5	eV

to Hall scattering factors greater than unity (Equation 4.19); this value is consistent with other work in the field and is suggestive of an anisotropic Fermi surface as is believed to be the case for hole transport in strained silicon germanium alloys.[90–92]

The value found in this work is actually slightly larger than values found elsewhere, but this may be due to additional contributions from energy-dependent scattering mechanisms. In Reference 92 the Hall scattering factor is found to generally increase with sheet density from around 0.4 at $3\times10^{11}\text{cm}^{-2}$ to 0.8 at $4\times10^{12}\text{cm}^{-2}$. (This behaviour is independent of the composition of the alloy layer.) The tendency of the Hall factor to unity as sheet density increases is unsurprising, since the Fermi temperature is directly proportional to the sheet density; as sheet density increases (at any given temperature) the carrier gas will gradually become degenerate. The Hall scattering factor is unity when $T_F \gg T$.[91]

From Equations 4.20 and 5.11, the 282 K drift mobility of this device is therefore around 350 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at 10^{12}cm^{-2}. A drift mobility of 330 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at $5\times10^{11}\text{cm}^{-2}$ has been measured in a 6 nm-thick modulation-doped p-type $Si_{0.6}Ge_{0.4}$ alloy layer at 295 K;[92] for comparison, in a pure germanium channel grown on a $Si_{0.3}Ge_{0.7}$ virtual substrate a mobility at room temperature of 1700 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at $2.3\times10^{12}\text{cm}^{-2}$ has been recently reported.[93]

At 10 K and 0.35 K, once again the calculations fail to fit to the experimental data for mobility as a function of sheet density. In Figure 5.30 experimental data for the resistivity as a function of sheet density is compared to calculations at certain sheet density values, and the form fits quite badly. Clearly, there are effects which augment the mobility of metallic carrier gases and diminish the mobility of insulating carrier gases, which are not included in the method and limit its application at low temperatures where these effects are significant. This will be explored in the following section.

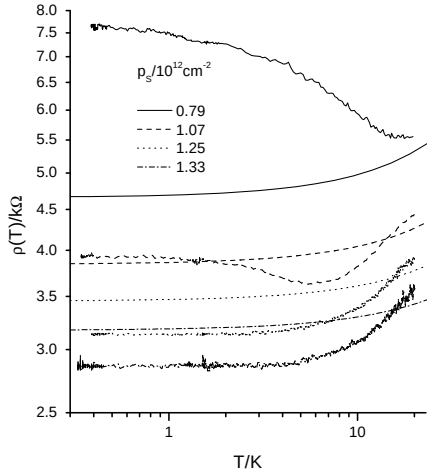

FIG. 5.30: Resisitivity as a function of temperature, comparing experimental data with calculations similar to those demonstrated in Figure 5.28 and Figure 5.29.

5.6.3. Resistance as a Function of Temperature and Sheet Density

Figure 5.31 shows how the resistivity of the Siemens device varies as a function of temperature at a few different sheet carrier concentrations around the peak in mobility. The resistivity appears to be saturating as the temperature decreases; for low sheet densities resistivity increases with decreasing temperature (insulating behaviour) but for higher densities it decreases (metallic behaviour).

Resistivity (or conductivity) at ∼1 K is usually discussed in terms of weak localization, interactions (Sections 4.1.4, 5.5.3 and 5.6.4) and screening.[19,21] Equation 5.12 is based on finite-temperature screening theory, if

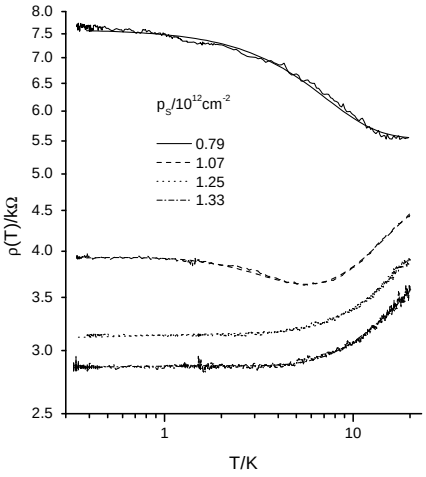

FIG. 5.31: Temperature dependence of the resistivity around the metal-insulator transition. At the lowest sheet density shown, resistivity increases as temperature decreases but may be saturating. At larger sheet densities, the resistivity decreases to a saturation value as the temperature decreases. (The behaviour changes little as sheet density continues to increase.)

$k_BT \ll \hbar/\tau$ and $T < T_F$:[94]

$$\sigma(T) = \sigma(0)\left[1 - \beta\left(\frac{T}{T_F}\right)^2\right] + \frac{Ae^2}{\pi h}\ln\left(\frac{k_BT}{\hbar/\tau}\right) \quad (5.12)$$

And Equation 5.13 on the temperature dependence of the static dielectric function:[95]

$$\sigma(T) = \sigma(0)\left[1 - C\frac{T}{T_F}\right] + \frac{Ae^2}{\pi h}\ln\left(\frac{k_BT}{\hbar/\tau}\right) \quad (5.13)$$

The second term in Equations 5.12 and 5.13 is the correction to the conductivity due to interactions, Equation 4.34;[48] τ is the elastic scattering time which in this system should be roughly equal to the momentum relaxation (transport) lifetime and $A = \left(1 - \frac{3}{4}F^*\right)$. The constants β and C should be of order unity, and $\sigma(0)$ is the zero-temperature Boltzmann conductivity. If the low-field magnetoresistance showed any evidence for weak-localization, then a term for the temperature dependence of the dephasing would be incorporated in A.[19,21]

However, for sheet densities of the order of 10^{12}cm^{-2} and an effective mass of the carriers around 0.3 m_e the Fermi temperature T_F (Equation 3.5) is over 100 K, a factor of three greater than in Reference 21.

For temperatures which are high (but still such that $T < T_F$) the interaction term contribution weakly increases with temperature but the first term in both Equations 5.12 and 5.13 dominates, causing the conductivity to decrease. As the temperature decreases toward zero, however, the interaction term dominates and the conductivity decreases. Hence, there should be a maximum in the conductivity and this is indeed seen in References 21 and 19 (at 2 K and 0.8 K respectively) where forms similar to Equations 5.12 and 5.13 are successfully applied. (Eventually the interaction term in Equations 5.12 and 5.13 breaks down, tending to negative infinity: in this limit, k_BT must be replaced by a temperature-independent upper cut-off.[47] This avoids an unphysical negative overall conductivity.[96])

A maximum in conductivity (that is, a minimum in resistivity) may be seen in Figure 5.31 in the data at a sheet density of $1.07\times10^{12}\text{cm}^{-2}$ but not at other sheet densities, and the resistivity is clearly not increasing as temperature decreases further. Analogous data in device 55/53 (Figure 5.7) also fail to clearly show a minimum in resistivity at metallic sheet density values. Another approach is needed to evaluate these data.

Work on the metal-insulator transition in 2-dimensional samples suggests that the resistivity can be approximated by the following expression:[59,97]

$$\rho(T) = \rho_0 + \rho_1 \exp\left[-(T_0/T)^n\right] \quad (5.14)$$

ρ_0, ρ_1 and T_0 are dependent on sheet density but not temperature. In p-type silicon-germanium heterostructures, good fits have been acheived with $n \approx 0.5$. The exponential term becomes decreasingly important as sheet density increases.[97,98]

In order to fit the data in Figure 5.31, the following form for $\rho(T)$ is introduced:

$$\rho(T) = \rho_0 + \rho_c \exp\left[-(T/T_0)^n\right] + \rho_l \ln\left[1 + \frac{k_BT}{\hbar/\tau}\right] \quad (5.15)$$

The first term, which is constant with respect to temperature, represents elastic scattering from impurities[43] and can be estimated using $\rho_0 = (p_s q \mu_0)^{-1} - \rho_c$ where $\mu_0 = \mu(T \to 0)$. (The possibility of metallic transport at zero temperature is discussed in section 5.5.3, in the context of quantum dephasing mechanisms at zero temperatures which may play a role in the metal-insulator transition.[64,99])

The second term represents a thermally activated process which enhances the conductivity as temperature increases. An exponential form is used, but since the resistivity does not vary by more than an order of magnitude this may not be necessary (and the very high conductivity state which may feature a re-entrant insulator-metal-insulator transition as seen in gallium arsenide[100] is never reached). In any case, $1 \leq n > 2$ keeps this term well behaved as $T \to 0$.

The third term appears superficially similar to the interaction correction discussed above, but in this case causes the resistance to rise weakly with temperature. The use of the $\ln(1 + x)$ form again ensures good behaviour as $T \to 0$, and $\ln(1 + x) \approx x$ for low x.

Fitting parameters are summarized in Figure 5.32: n is fixed at 1.5. At low sheet densities, for the temperature range considered, the third term in Equation 5.13 becomes redundant and Equation 5.14 is recovered.

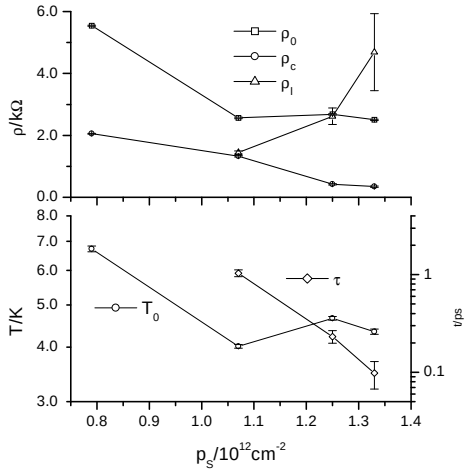

FIG. 5.32: A summary of parameters for fitting Equation 5.15 to the data in Figure 5.31: at low sheet densities the third term does not contribute at these temperature so $\rho_l \to 0$ and τ is undefined.

While this form successfully fits the data, it has no direct theoretical justification and the behaviour of at $\rho(T \to 0)$ high sheet densities remains unresolved. Low-field magnetoresistance may clarify the influence of quantum corrections to the conductivity,[101] as it has in the case of the previously discussed devices.

5.6.4. Magnetoresistance

For temperatures between 25 K and 300 K, magnetoresistance data (taken across the sheet density range shown in Figure 5.24) shows an essentially constant ρ_{xx} and R_H up to 11 T. This makes it clear that, firstly, no weak localization or Landau level formation is present at these temperatures (which is to be expected) and secondly that there is no parallel conduction in, for example, an inversion layer at the interface between the silicon cap layer and the silicon dioxide gate dielectric.

The form of the 350 mK magnetoresistance data in Figure 5.33 constrasts strongly with that of Figure 5.17 and Figure 5.18. The magnetoresistance is always positive, showing no evidence of the low-field negative magnetoresistance that is the signature of weak localization. Features at high field in some of the data at lower sheet densities may be related to the formation of Landau levels (at 10^{12}cm^{-2}, 8 T corresponds to a filling factor of 5) but ρ_{xy} (not shown) does not show signs of Quantum Hall plateaux. High field magnetoresistance may be due mainly to the effect of Zeeman splitting on interactions. Spin-orbit scattering gives rise to positive magnetoresistance at low fields.[74,102] Data taken by other workers on another device from the same Siemens wafer is shown in Figure 5.34. Shubnikov–de Haas oscillations are clearly visible, but the device was destroyed before more data at other temperatures and gate voltages could be obtained.

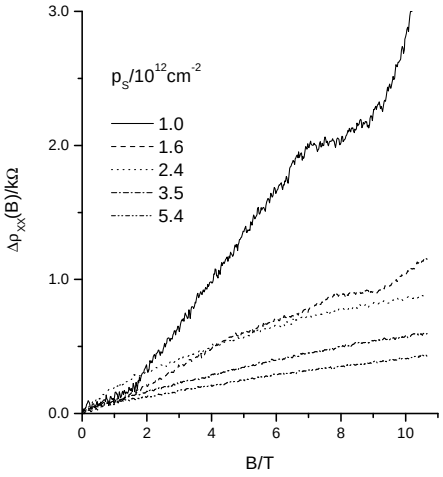

FIG. 5.33: Magnetoresistance data at 350 mK, with the zero field value of the resistance subtracted in each case. The magnetoresistance is always positive, in contrast to Figure 5.17.

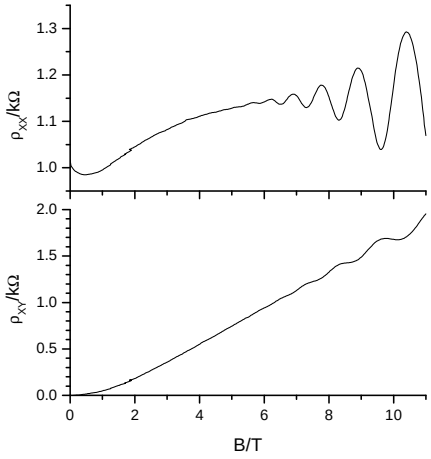

FIG. 5.34: This magnetoresistance data was obtained during an early phase of device characterization by other workers. There are clear Shubnikov–de Haas oscillations at this temperature (1.47 K) but the device was destroyed before further measurements could be performed.

It is not strictly possible to apply the analysis described in section 4.1.3 when data at only one temperature are available; the effective mass is normally found from the temperature decay of the oscillations. However, the variation of amplitude of the oscillations with respect to magnetic field can be used to find α if an effective mass is assumed. In fact, if Equation 4.28 is plotted

as described then the data points representing the minima and maxima at this temperature should all lie on the line $y = x + \ln(4)$ if the choices of m^* and α are correct. It was found that the best fit to a line of unity gradient from a plot of Equation 4.28, with a consistent value of α from the gradient of the plot of Equation 4.29, was produced with $m^* = 0.3\ m_e$ and $\alpha = 3$. However, the intercepts of both of these plots was much greater than the value $\ln(4)$ and attempts to correct this destroyed the consistency between the two plots. In other words, the oscillations are much larger in amplitude than the theory allows: the numerical factor before the summation in Equation 4.24 would have to be several times larger and there is no theoretical reason for this to be the case.

5.7. Conclusions

5.7.1. 55/53

This device has proved to be an interesting and useful testbed for aspects of 2-dimensional physics. Its resistivity as a function of temperature weakly saturates at finite (sheet density dependent) values as temperature drops to zero, which would seem to contradict conventional theories of 2-dimensional transport based on scaling described in terms of the results from the Siemens device.

Whilst the mobility values measured were themselves impressive, fluctuations can be seen in Figure 5.8 and Figure 5.9 which as yet remain unidentified: they may or may not be related to the effects seen in the Siemens device described in section 5.5.2 and References 57 and 58. Generally, data for the mobility as a function of sheet density is not as detailed as that presented in Figure 5.9 so subtle features will not be seen.

The low-field magnetoresistance seems to be explained well within the framework of weak-localization; the contribution from interactions and Zeeman splitting, however, is less clear due to the anomalous values of the screening parameter F^* returned by the fitting process. The temperature dependence of the dephasing time was considered in terms of traditional weak-localization, where the dephasing rate drops to zero at zero temperature, and controversial new ideas which suggest that the dephasing rate remains finite at zero temperature. The results are inconclusive, in that the traditional theory tends to predict dephasing times which are too long but the newer ideas do not fit the data very well unless unreasonable freedom is given to their free parameters. The question of finite dephasing at zero temperature is important and fundamental.[103] It is suggested that the theory most needs modification so that the dephasing rate increases with increasing carrier density, leading to metallic behaviour at high densities and low temperatures.

The behaviour of the elastic scattering time at low temperatures, however, does not appear to be of such theoretical interest even though it is of more direct relevance to the momentum relaxation time and therefore the mobility. From Figure 5.13, the elastic scattering time seems to be generally increasing with decreasing temperature in all but one case.

High-field magnetoresistance in the metallic regime is conventional, but for lower sheet densities a magnetic-field-driven transition from the quantum Hall effect state to an insulating state with a very high resistivity occurs at a filling factor ∼1.6. This is often seen in silicon and silicon-germanium systems, and is related to the zero-field metal-insulator transition seen in gated devices. Both this and the results of low-field magnetoresistance characterization may contribute to the understanding of localization and interactions in 2-dimensional semiconductors, and therefore the metal-insulator transition and the nature of mobility-limiting scattering mechanisms.

5.7.2. Siemens Device

It is clear from the long term characterization of Siemens devices that results during the earliest phases were more promising than those from the main investigation detailed in this chapter. However, during these early phases devices would last no more than a few days before they would be destroyed by stray voltages; only the protection circuit described in section 5.3.2 has made it possible to perform systematic and detailed characterization of a single device. As research focus moves towards the characterization of devices for commercial applications such considerations will become increasingly important.

It may be the case that the drop in performance between Figure 5.23 and Figure 5.25 is due to the full or partial relaxation of the active channel. Results obtained by H. E. Fischer at Siemens[104] at 4 K show a peak mobility of 1800 $\mathrm{cm^2V^{-1}s^{-1}}$ at $5{\times}10^{11}\mathrm{cm^{-2}}$ (the peak mobility in Figure 5.25 is 1700 $\mathrm{cm^2V^{-1}s^{-1}}$ at $1.3{\times}10^{12}\mathrm{cm^{-2}}$). The fact that this peak is at a lower sheet density in the data of H. E. Fischer than in Figure 5.25 suggests that interface impurities are a comparatively serious issue in the device studied here.[18]

The peak mobility in Figure 5.25 is consistent with calculations for a relaxed $Si_{0.5}Ge_{0.5}$ alloy.[105] On research-scale devices this can be verified by X-ray or micro-Raman measurements, or transmission electron microscopy. However, the small scale of the devices makes such analysis impossible in this case.

Degradation of the oxide may instead be the cause of the drop in performance. Since the carrier gas exists in the alloy layer only a few nanometres away from the oxide interface, the quality of this interface can limit the hole mobility in the strained channel.[106] Experience suggests that ESD damage is catastrophic rather than progressive, so this is not the cause. Ionic contamination (with, for example, Na^+) would lead to shifting of the threshold voltage and possibly changes in the transconductance characteristics,[10] which were not seen. A remaining possibility is that the oxide interface has been degraded by hot charge carriers.[107–109]

It is possible, also, that the small size of the device is itself the issue: a 5 nm $Si_{0.5}Ge_{0.5}$ alloy grown on and capped with pure silicon should remain fully strained (up to at least 500°C)[9] but the small lateral scale of the mesa (the Hall bar is 2.5 μm wide) may have lead to significant relaxation towards the edges of the device.[110] This would have the strongest implications for high-field quantum magnetotransport and may explain the absence of Shubnikov–de Haas oscillations in Figure 5.33 (since conduction is concentrated towards the edges in the regime where $\hbar\omega_c \sim k_B T$) but leave the room-temperature propeties almost unscathed. The calculations of mobility as a function of sheet density, and the extracted value for the Hall scattering factor at room temperature, should therefore not be invalidated. Further investigations into the long-term stability of micron and sub-micron scale devices with a high level of grown-in strain is certainly necessary.

The resistivity as a function of temperature has been analyzed, and shown not to agree with existing theories of weak-localization and screening. A new functional form has been proposed as an empirical fit, but now needs theoretical justification.

6. TRANSPORT IN AN N-CHANNEL STRAINED SILICON DEVICE

6.1. Abstract

Previous chapters have stressed the importance of silicon-germanium in terms of improving the mobility of holes in semiconductor materials. However, it is also important to study the characteristics of electron gases in strained silicon-germanium heterostructures if both n- and p-channel devices are to be integrated. The physics of 2-dimensional electronic systems is also simpler and more fully understood in some respects than it is for holes. Additionally, the mobility of electrons is generally better than that of holes, in a given system, so transport in a more metallic regime can be investigated and classical magnetoresistance is more easily analyzed.

6.2. Introduction

Pure, unstrained silicon has 6 conduction band minima. Tensile strain (perpendicular to the growth direction, as occurs when pure silicon is grown on relaxed silicon-germanium alloy) causes the two minima in the growth direction to lie at a lower energy than the four minima in the plane. This greatly reduces intervalley scattering.[3]

Since the structure of the device (Figure 6.1) is "normal" the heterointerface at which the 2DEG is defined should be of high quality, the electron channel should have a high mobility.[15] Quantum effects which rely on disorder and impurities (for example, weak localization) should not visibly influence the transport.

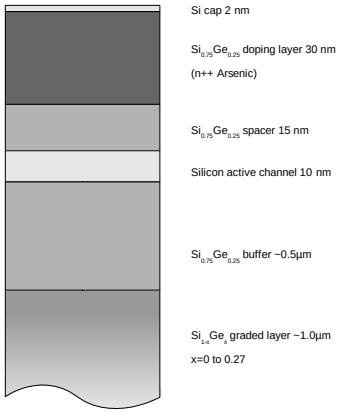

FIG. 6.1: The structure of the n-type device. There is a thick, graded layer followed by an alloy buffer layer. There is heavy n-type doping above the active channel and there is no gate.

6.3. Structure

The structure of the n-channel device is shown in Figure 6.1. It was grown by Gas Source Molecular Beam Epitaxy at Imperial College, by R. Ferguson. The dopant layer is above the active channel making this a "normal" structure and so there is no gate and the electron concentration is not variable. The active channel of pure silicon was grown on a "virtual substrate" of relaxed $Si_{0.75}Ge_{0.25}$ and capped by the same material, causing it to be under tensile strain and creating a quantum well for electrons. (This is in contrast with the p-channel devices considered in Chapters 5 and 7, in which the active $Si_{1-x}Ge_x$ channels are pseudomorphically grown on pure silicon leading them to compressively strained, forming quantum wells for holes.)

Calculations for this heterostructure yield a band gap in the active channel of 0.90 eV (compared to 1.17 eV for unstrained silicon or 1.09 eV for the unstrained $Si_{0.75}Ge_{0.25}$ alloy) and an effective mass for the electrons in the two Δ out-of-plane minima of 0.199 m_e. The conduction band offset between the active channel and its surrounding layers is 250 meV and the valence band is 60 meV lower in the active channel making this a Type II structure. The splitting between the in-plane and out-of-plane conduction band minima is 200 meV.[14]

Calculations similar to those described for a p-type system in Equation 3.1 and Figure 3.1 suggest that the sheet carrier concentration of electrons in the active channel (at the upper heterointerface) will be $\sim 10^{12}\text{cm}^{-2}$ at $T \ll T_F$; for this sheet density and effective mass (and

taking into account the double valley degeneracy) the Fermi temperature will be ∼70 K.

6.4. Classical Magnetoresistance and Mobility Spectrum Analysis

Data for the variation of resistivity with magnetic field, at various temperatures, are shown in Figures 6.2 to 6.4. Chapter 4 describes both the utility of measurements of magnetoresistance in the classical regime and the problems associated with their interpretation. Equations 4.10 and 4.11 show how elements of the conductivity tensor $\boldsymbol{\sigma}(B)$ can be found in terms of the function $s(\mu)$ (Equation 4.12) through integral transformation. (Equations 4.13 and 4.14 relate $\boldsymbol{\sigma}(B)$ to the more usually measured quantities of the longitudinal resistivity $\rho(B)$ and the Hall coefficient $R_H(B)$.)

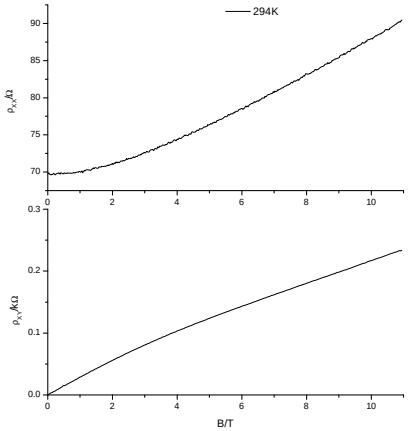

FIG. 6.2: Magnetoresistance at 294 K. It is clear from the gradient of the ρ_{xy} data that the Hall coefficient is decreasing as magnetic field increases.

This function $s(\mu)$ is known as the mobility spectrum. This gives the contribution to the conductivity of the system due to carriers with mobility μ. It is assumed in this investigation that this function does not change with magnetic field; this will not be true when $\hbar\omega_c \geq k_B T$ (Landau levels are formed) or when the carriers become weakly localized since both these conditions involve magnetic-field dependent mobility. Under this assumption, magnetoresistance is always positive; negative magnetoresistance can be analyzed using the assumption of a magnetic-field dependence of the mobility.[111] Significant intersubband scattering can also invalidate this method.[112]

To find the magnetoresistance given a form for the mobility spectrum is straightforward. Finding the mobility spectrum from magnetoresistance data, however, is an inverse transform problem and therefore more difficult

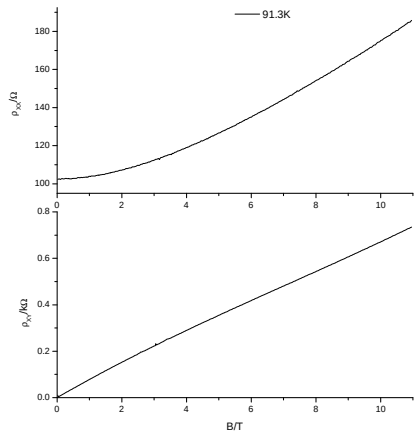

FIG. 6.3: Magnetoresistance at 91 K. the scales of the vertical axes have changed from Figure 6.2, but the forms of the curves are similar.

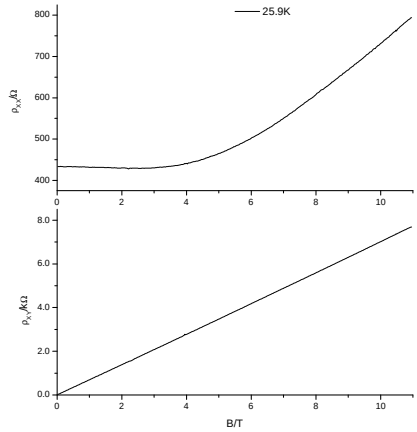

FIG. 6.4: Magnetoresistance at 26 K. The Hall coefficient is effectively constant with respect to magnetic field. For fields up to 4 T, there is little longitudinal magnetoresistance: this corresponds to the limit of $\hbar\omega_c \sim k_B T$. Eventually, as temperature decreases, Shubnikov–de Haas oscillations will appear.

and the subject of much investigation.[28,29,113–116] Some of these approaches will now be briefly reviewed, before considering the method of maximum entropy in more detail.

(i) Reference 28, which first introduced the concept of the mobility spectrum, involves a mathematically complicated solution method, the numerical details of which will not be explored. The number of data points used to create a solution is equal to the number of distinct

carriers gases within the sample, plus one. The solution method may either proceed by searching for an imposed number of carrier gases or be free to start by attempting a fit for a single gas, and then a two-gas system, and so on. For each search for n gases, the method must try each combination of $n+1$ data points from the whole data set, keeping the sets which lead to "physical" solutions, and from these sets generate an "envelope function" within which the mobility spectrum is found.

Once a set of carrier gases has been found, a least-squares fit on the original data is performed with the carrier gas parameters as variables. If this fit is satisfactory and the parameters emerge virtually unscathed then the results of the original analysis can be considered reasonable.

If only a limited number of data points ($\sim$10) are available and the system is expected to contain only two or three distinct carrier gases, then this method is quite efficient. However, the data collected in this part of the investigation generally comprises resistivity data at about 500 magnetic field values and the system simply cannot cope; to choose points or create averaged data by hand is to impose artificial constraints on the final solution. Also, this method does not respond well to errors in the experimental data.[29]

(ii) Reference 113 introduces and describes the Reduced Conductivity Tensor method for extracting the carrier concentration and mobility of each component of a multilayer semiconductor system. Assumptions are made that the carrier gases are essentially degenerate and that the effective mass is isotropic, but nevertheless, agreement between theory and experiment is good.

(iii) In Reference 114, it is assumed that the mobility for a mixed scattering can be approximated by the phenomenological expression:

$$\mu = \mu_0 x^\alpha \text{ where } x = \frac{E}{k_B T} \tag{6.1}$$

and that the mobility spectrum of non-degenerate carriers in a spherical band can be approximated by:

$$s(\mu) = S_0 e^{x(\mu)} x(\mu)^{2.5} \tag{6.2}$$

where S_0, μ_0 and α are coefficients depending on the density of states and on the parameters of all the scattering mechanisms. Then, the integration over the whole mobility spectrum in Equations 4.10 and 4.11 is converted to summation of discrete spectra for electrons and holes and an iterative transformation procedure is performed. The results are decomposed into high-mobility carrier and low-mobility carrier contributions, the division chosen arbitrarily, with results for the former seeming much more satisfactory than results for the latter.

(iv) References 115 and 116 describe the "quantitative mobility spectrum technique" (QMSA) and the "improved quantitative mobility spectrum technique" (i-QMSA) respectively. These, again, are iterative techniques but with no initial assumption about the solution (although the method from Reference 28 is used to create the first trial solution to iterate, and conductivity data is extrapolated to higher magnetic field values to extend the available mobility range). The i-QMSA method introduces a few extra tricks for improving the fits whilst smoothing the spectra and making them more physically reasonable.

A serious issue (regarding inverse transformation problems in general) is that while a particular spectrum $s(\mu)$ may, upon the action of an integral transform over the "kernel function" $K(B, \mu)$, produce a magnetoconductivity $\sigma(B)$, the inverse problem is ill-conditioned and the solution obtained for $s(\mu)$ by inverting the kernel is extremely sensitive to small changes or errors in the $\sigma(B)$ data. Also, a particular $s(\mu)$ is not a unique solution, within the uncertainty bounds of the original noisy and incomplete $\sigma(B)$ data.

6.4.1. Maximum Entropy Method of Mobility Spectrum Analysis

The technique of "Maximum Entropy" has been employed to solve this and other problems involving inverse integral transformation where the result is a positive, additive function.[29,117–119] Generally, if two solutions (found by any means) are of equal merit (in terms of, for example, their least-squares fits to the original results) then the solution with the larger entropy is to be favoured, since it is maximally noncommittal with regard to missing (unmeasured) information in the original data.[29,120] In other words, the solution favoured by the maximum entropy method extracts the most information out of the original data without making unreasonable assumptions about information which is unavailable.

This method embodies Bayes' theorem: the probability of a particular hypothesis given some data (and any background assumptions) is proportional to the probability of that data given the hypothesis and the assumptions and the probability of the hypothesis given the assumptions alone.[117]

The entropy S of a discrete mobility spectrum $\{s\}$ can be defined as:[29]

$$S\{s_j\} = -\sum_{j=1}^{n} p_j \ln p_j \tag{6.3}$$

$$p_j = \frac{s_j}{\sigma_0} \tag{6.4}$$

where p_j is the probability of s_j and σ_0 is the conductivity at zero magnetic field. This imposes the condition that the mobility spectrum is non-negative and normalizable, which is physically realistic. However, if any prior information about the form of the spectrum is available then this may be incorporated as a so-called default model $\{m\}$:[118,119]

Fitting is now a matter of minimizing the function Q:[119]

$$Q = \chi^2 - \alpha S \tag{6.5}$$

where $\chi^2 = \sum_i (\sigma(B_i) - \sigma_c(B_i))^2/\delta_i^2$; $\sigma_c(B)$ is magnetoconductivity calculated from the fitting mobility spectrum, δ^2gives a measure of the error in the data (this parameter is conventionally referred to as σ^2 but here is renamed to avoid confusion with conductivity) and α is a hyper-parameter which controls the relative importance of the least-squares and entropic constraints.[119] Full technical details are given in Reference 118 and Chapter 5 of Reference 119. Equation 6.6 can be meaningfully compared to the free energy, $F = E - TS$, which is often to be minimised to find the solution in thermodynamic or statistical-mechanic systems.[120]

There are various flavours of the maximum entropy method, which differ in the way in which they approach the hyper-parameter α. In the case of Historic maximum entropy, α is set so that at the minimum of Q (Equation 6.6) $\chi^2 = M$ where M is the number of "observations" or data points. In the case of Classic maximum entropy, the most probable value of α is found, given the data and the default model.[121,122] Alternatively, the method described in Reference 120 does not use Bayesian inference but treats the maximization of the entropy as a natural starting-point for derivations of other results of statistical mechanics.

The approach of Bryan's algorithm is to calculate solutions for a range of α values and evaluate the probability of each being correct given the data and the default model. These solutions are then averaged, weighted by their probabilities.[119] However, this method does not work very well if α is very small: in the method developed in Reference 119 a solution produced by Bryan's algorithm maximum entropy is used as the default model for the next solution. This procedure is repeated until the the most probable solution corresponds to an α value well within the range of applicability of Bryan's algorithm.

Since the integral transform from $s(\mu)$ to $\sigma(B)$ actually involves two integral equations (Equations 4.11 and 4.12) then we work with[29]

$$\sigma(B) = \sigma_{xx}(B) + \sigma_{xy}(B) = \int_{-\infty}^{\infty} \frac{1 + \mu B}{1 + \mu^2 B^2} s(\mu) \mathrm{d}\mu \tag{6.6}$$

and since the forms for $s(\mu)$ and $\sigma(B)$ are discrete, then Equation 6.7 can be rewritten $\sigma_i = K_{ij} s_j$. The mobility spectrum may be computed as:

$$s_j = K_{ji}^{-1} \sigma_i \tag{6.7}$$

Since the magnetoconductivity $\sigma(B_i)$ is relatively insensitive to the details of the mobility spectrum $s(\mu_j)$ then the kernel contains a large amount of repeated information; many of the linear equations described by the kernel will be (almost) identical.[119] In addition, elements of the kernel matrix which are dominated by small signal noise or computational rounding errors will invert to very large values which will be overpowering in Equation 6.8. If the kernel is an $M \times N$ matrix written as $\boldsymbol{K} = UWV^T$ with $\boldsymbol{U}$ an $M \times N$ column-orthogonal matrix, $\boldsymbol{W}$ as an $N \times N$ diagonal matrix with nonnegative values (the "singular values") and $\boldsymbol{V}^T$ the transpose of an orthogonal $N \times N$ matrix $\boldsymbol{V}$, then the inverse of the kernel can be written as $\boldsymbol{K}^{-1} = VW^{-1}U^T$.[123]

If one of the elements of $\boldsymbol{W}$ is zero then the corresponding element of $\boldsymbol{W}^{-1}$ will be infinite and the matrix will be singular. However, by invoking the concepts of *nullspace* and *range* for singular matrices an element of $\boldsymbol{W}^{-1}$ is set to zero if the corresponding element of $\boldsymbol{W}$ is zero (or smaller than some set noise floor). This technique of *singular value decomposition* effectively reduces the dimension of the space that must be searched for a solution and discards noise which would corrupt the solution. It is therefore a very important feature of the following analysis. A widely available algorithm for performing singular value decomposition in the C programming language is svdcmp, from Numerical Recipes.[123]

A simple, powerful and successful maximum-entropy method for finding a mobility spectrum is described in Reference 29, based on the method of Reference 120. The described procedure does not, however, use singular value decomposition to remove overspecification and noise and so the calculation time increases with the square of the product of the number of mobility and magnetic field points, $(NM)^2$; nor does it make explicit a hyper-parameter (α in Equation 6.6) which controls the relative importance of a result which fits the original data versus a result with the maximum the entropy or deal with error in the data in an obvious way.

In the Bryan's Algorithm Mobility Spectrum method described below, the calculation time scales as $N^n M$ where n falls from unity to around 0.7 as N increases. This is a direct result of the singular value decomposition; the time to perform this scales as $N^2 M$ but tests and calculations suggest that the number of input points would have to exceed tens of thousands before the SVD was taking as long to perform as the actual maximum entropy inversion. For input data at 500 magnetic field values, and an output spectrum of 1000 points, the BAMS result is produced in 7 hours on a Solaris 360 MHz Ultra SPARC-IIi or 3 hours on a 450 MHz Pentium III running Linux.

6.4.2. "Bryan's Algorithm Mobility Spectrum" Results

In Chapter 5 of Reference 119, the techniques of singular value decomposition and iterated Bryan's algorithm maximum entropy are employed to extract a spectral function (related to the density of states) from a static correlation function. The computer code used to perform this task was written by J. P. Hague and then modified in collaboration with the Author to extract a mobility spectrum from magnetoconductivity data.

Mobility spectra produced by application of this

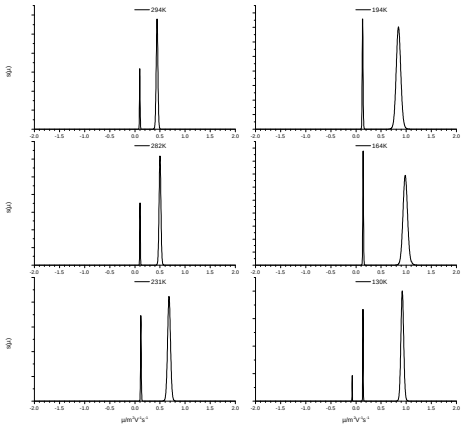

FIG. 6.5: Mobility spectrum results produced using maximum entropy analysis. The temperature is shown on each plot. In five of the spectra there are two peak; the low mobility peak is ascribed to conductino in the arsenic dopant slab, the higher mobility peak to the 2-dimensional electron gas itself, in the strained silicon active channel. The mobility of the 2DEG peak generally increases as temperature decreases. The third peak in the 130 K result will be discussed in the text. The vertical scale in each case is irrelevant: it is the area under each peak that is of importance, not their heights.

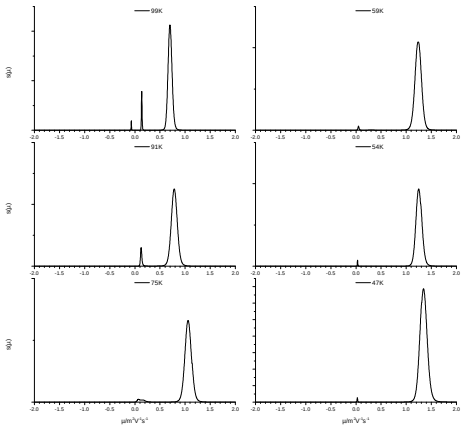

FIG. 6.6: (continuation from Figure 6.5) As temperature decreases further, the mobility of the 2DEG peak continues to increase while the dopant slab peak diminishes.

method are shown in Figures 6.5, 6.6 and 6.7. Summaries of the mobilities and sheet densities found from these spectra are shown in Figures 6.8 and 6.9 along with values obtained by simple application of Equations 4.7 and 4.9 at low (±0.5 T) and high (±10 T) magnetic fields. The manner in which these results have been extracted

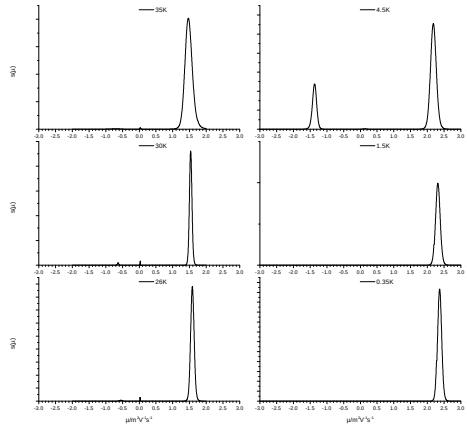

FIG. 6.7: (continuation from Figure 6.6 but with an extended x-axis scale) The 2DEG peak reaches a high mobility, and only faint vestiges and artifacts remain in the rest of the spectrum. The exception is the 4.5 K result, which features a significant peak in the "negative mobility" region. If this were genuine (see text) it would signify high-quality p-type conduction. Data used for these low-temperature results were limited to lower magnetic fields so as to avoid the Shubnikov–de Haas regime.

from the spectra will be explained in the following section. Results for 4.5 K, 1.5 K and 0.35 K are not shown; the 4.5 K result is discussed below, the 1.5 K and 0.35 K result agree on a mobility of 23,000 $\mathrm{cm^2V^{-1}s^{-1}}$ at a sheet density of $9.0\times10^{11}\mathrm{cm^{-2}}$.

Figure 6.10 shows how the 2DEG and dopant layer contribute to the overall resistivity of the device, as a function of temperature. It can be seen that, since the resistivity of the dopant layer is several times larger than that of the the 2DEG, it makes little contribution even at high temperatures. The temperature dependence of the resistivity of the 2DEG is the subject of section 6.4.

The Dopant Layer

The carrier concentration n0 in the dopant layer as a function of temperature is:[17]

$$\frac{n_0(N_A+n_0)}{N_D-N_A-n_0}=\beta N_C\exp\left(\frac{-E_D}{k_BT}\right)\quad n_0\ll N_C \quad (6.8)$$

where N_D and N_A are the concentrations of donors and acceptors, N_C is the effective density of states in the conduction band and β is the degeneracy. N_C is given by:[10]

$$N_C=2\left[\frac{2\pi m^*k_BT}{h^2}\right]^{\frac{3}{2}} \quad (6.9)$$

If the $T^{3/2}$ dependence of N_C (and the temperature variation of the width of the ionized region of the dopant

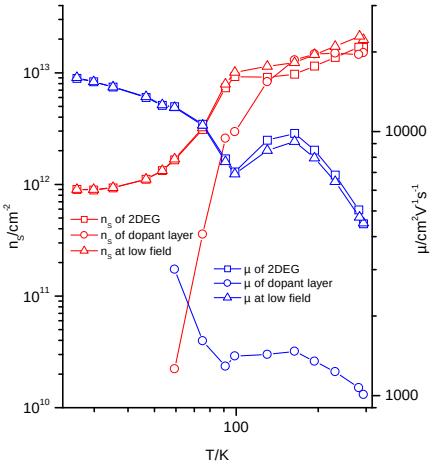

FIG. 6.8: A summary of results extracted from magnetoresistance measurements. The mobilities of the two main peaks in the spectra are shown (ascribed to the 2DEG and the dopant) as are the effective sheet densities obtained by their integration. Mobilities and sheet densities measured directly using assumption of the single-carrier Hall effect at low magnetic field are also shown. At low temperatures, the low-field Hall effect gives the correct 2DEG parameters; at high temperatures where there is parallel conduction there is a slight discrepancy. The mobility spectrum analysis yields unphysically high sheet densities for the 2DEG at high temperatures.

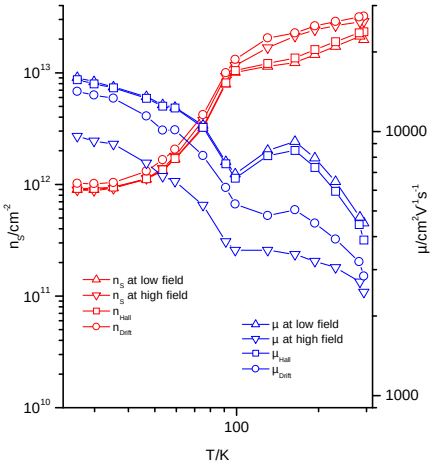

FIG. 6.9: A summary of results extracted from magnetoresistance measurements. Mobilities and sheet densities measured directly using assumption of the single-carrier Hall effect at low magnetic field (where Equations 4.15, 4.17 and 4.18 should apply) and at high magnetic field (where 4.16 should apply, not applicable in the quantum limit) are shown, alongside results from the derived analyses of the spectrum, Equations 6.17, 6.18, 6.20 and 6.21. The calculated drift mobility should not be larger than the measured high field value: this is another sign that something may be wrong.

layer) is assumed to be much less important than the exponential temperature dependence in Equation 6.9, $n_0 \ll N_D$ and $N_D \gg N_A$ (as is likely since the concentrations used in modulation doping are very high compared to traditional device doping levels)[10] then Equation 6.9 reduces to:

$$n_0 \simeq AT^{3/4} \exp\left(-\frac{E_D}{2k_BT}\right) \tag{6.10}$$

The temperature dependence of the sheet carrier density of the dopant layer (the red circles in Figure 6.8) in the region of 60–100 K* suggests an activation energy of the arsenic donors of the order of the accepted value in pure silicon of 54 meV,[124] but since only a few data points are available a rigorous analysis is not possible. Since the dopant slab is only significantly ionized in the region nearest the active channel at these temperatures,[21] this analysis cannot necessarily be used to infer the donor concentration.

* Data between 100 K and 300 K do not fit Equation 6.11 and cannot be used, for reasons which will be explored in the remainded of this chapter.

Some variation of sheet density with temperature of the 2DEG is to be expected:[29] when $T \ll T_F$ the Fermi level is pinned in the dopant slab and the occupied quantum-well subband, similar to the case for p-type conduction shown in Figure 3.1. The charge density in the channel must be consistent with the gradient of the energy bands between the channel and the dopant slab (as discussed in section 6.3). At higher temperatures, the spreading of the Fermi occupation function means that the conduction band does not need to cross below the Fermi level to be significantly populated. However, such a large increase in 2DEG density as temperature increases may be a cause for some concern, and will be discussed in sections 6.5 and 6.8.2.

In strained silicon, the conduction band minimum is doubly degenerate. These two conduction valleys have identical mobilities and occupation densities, so in each case the 2DEG peak in the mobility spectra represents both sets of electrons. However, the density according to Shubnikov–de Haas oscillations will be half this, as will be seen.

"Mirror" Peaks

A very strong peak can be seen in the 4.5 K spectrum, at a mobility of around -14,000 cm^{2}V^{-1}s^{-1}. The effective sheet density of this peak is (-)3.4×10^{11}cm^{-2}. This compares to the 2DEG peak, which has a mobility of 22,000 cm^{2}V^{-1}s^{-1} at 6.8×10^{11}cm^{-2}. It can be seen

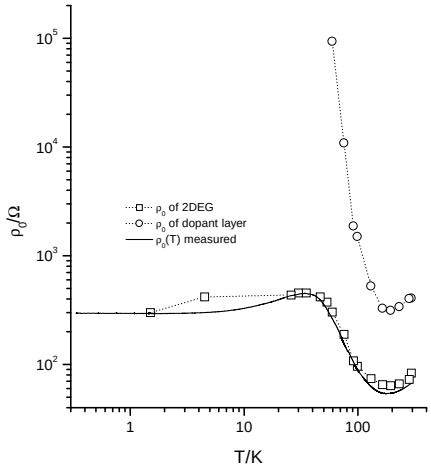

FIG. 6.10: Dependence on temperature of the resistivity of the n-type device; directly measured data (at 150 nA) is the solid line, symbols are results from mobility spectrum analysis. The Fermi temperature of the electron gas (calculated from its zero-temperature carrier concentration) is 70 K, and it can be seen that the resistance is changing dramatically in this region. As the temperature decreases to a few Kelvin, the resistivity saturates. Freeze-out of the dopant layer can clearly be seen.

immediately that this density for the 2DEG is too low (there is no reason for it to deviate from its value at other low temperatures of $9.0\times10^{11}\mathrm{cm}^{-2}$) and that the missing conductivity is provided by the "mirror peak" which would, if it were genuine, represent p-type conduction. Spectra at 130 K and 99 K show very small peaks at low, negative mobilities; at other temperatures, similar peaks are present but with areas (that is, effective sheet densities) between two and four orders of magnitude lower than the genuine peaks. In fact, such very small mirror peaks result even from synthetic data generated to test the software.

Mirror peaks (normally at exactly the opposite mobility value, unlike the peak seen in the 4.5 K spectrum) are a general feature of mobility spectrum analysis, and have been described either as a computational artefact, a sign of band anisotropy or a result of inaccuracies in the measurement or processing of the ρ_{xx} signal.[29,116,125] Only the first suggestion is consistent with mirror peaks even being present in spectra generated from synthetic data, but Reference 125 notes that they are a feature of analysis performed with different computational methods. Consider a mobility spectrum similar to the 4.5 K result, with a well-defined and sharp real peak at μ_1 (with a total integrated conductivity of σ_1 and an effective carrier concentration of n_1) and a mirror peak at $\mu_2 = -\mu_1$. The magnetoconductivity is then given by

$$\sigma_{xx} = \int_{-\infty}^{+\infty} \frac{s(\mu)\mathrm{d}\mu}{1+\mu^2B^2} = \frac{\sigma_1}{1+\mu_1^2B^2} + \frac{\sigma_2}{1+\mu_2^2B^2} = \frac{\sigma_1+\sigma_2}{1+\mu_1^2B^2} \tag{6.11}$$

$$\sigma_{xy} = \int_{-\infty}^{+\infty} \frac{\mu Bs(\mu)\mathrm{d}\mu}{1+\mu^2B^2} = \frac{\mu_1B\sigma_1}{1+\mu_1^2B^2} + \frac{\mu_2B\sigma_2}{1+\mu_2^2B^2} = \frac{(\sigma_1-\sigma_2)\mu_1B}{1+\mu_1^2B^2} \tag{6.12}$$

If $n_2 \ll n_1$then the conductivity, and therefore the resistivity, reverts to the single-carrier model. However, if $n_2 \sim n_1$ then the transverse component is reduced greatly. Since the longitudinal component is always normalized to match the data at zero field, the result of the mirror peak is to change the scale of σ_{xy} in relation to σ_{xx} without changing its form. This suggests that the mirror peak is a consequence of mis-match between the numerical values of the longitudinal and transverse elements of the conductivity tensor. In the case of real data, this may arise due to mis-calibration between the equipment that measures the longitudinal and transverse resistivities, numerical errors when inverting this to generate the conductivity, or rounding errors (data was exported as tab-separated text, with numbers in standard form to 6 significant figures). This latter applies even if the data is synthetically generated.

In conclusion, the presence of mirror peaks is a sign of suspect data that must be treated with care: calibration of experimental equipment and the methods in which the data is processed should be checked.

6.4.3. Obtaining Semiconductor Properties from a Mobility Spectrum

Sheet Density and Mobility

Traditionally, the extraction of semiconductor properties from the magnetoresistance is treated in terms of simple analytical functions such as are described in section 4.1.2. The Hall sheet density, for example, is defined as:

$$n_{Hall} = \frac{1}{qR_H(B\to 0)} \tag{6.13}$$

where $R_H(B \to 0)$ is the Hall coefficient (Equation 4.14) at low magnetic field. It is acknowledged that the Hall coefficient at low fields is at the mercy of energy-dependent scattering within a carrier gas, or the presence of multiple carrier gases with differing mobilities.[126] In the limit of high magnetic field, however, the Hall coefficient should reflect only the total number of carriers within the system (Equations 4.16 and 4.21):

$$\sum |n| = n_{Drift} = \frac{1}{qR_H(B \to \infty)} \tag{6.14}$$

The zero-field conductivity is $\sigma_0 = n_{Hall} q \mu_{Hall} = n_{Drift} q \mu_{Drift}$ and so the Hall and drift mobilities can be found.

A mobility spectrum $s(\mu)$ generally comprises one or more peaks. The centre of a peak i can be simply interpreted as the mobility μ_i of the carrier gas represented by the peak, and the concentration of carriers within the gas n_i can be found by integrating over the peak:

$$n_i q \mu_i = \sigma_i = \int s(\mu) \mathrm{d}\mu \tag{6.15}$$

with the normalization of the spectrum is such that $\sigma_0 = \sum_i \sigma_i = \int_{-\infty}^{\infty} s(\mu) \mathrm{d}\mu$. The total number of carriers found by this method is the drift concentration:

$$n_{Drift} = \sum_i |n_i| \tag{6.16}$$

And so the drift mobility is then:

$$\mu_{Drift} = \frac{\sigma_0}{n_{Drift} q} = \frac{\int s(\mu) \mathrm{d}\mu}{q \sum |n_i|} \tag{6.17}$$

However, since the Hall coefficient in the limit of zero magnetic field (from Equations 4.11 and 4.14, assuming $\sigma_{xy}^2 \ll \sigma_{xx}^2$ and $\mu B \ll 1$) is:

$$R_H(B \to 0) \simeq -\frac{1}{B} \frac{\sigma_{xy}}{\sigma_0^2} \simeq -\frac{1}{\sigma_0^2} \int \mu s(\mu) \mathrm{d}\mu \tag{6.18}$$

then the Hall concentration as defined in Equation 6.14 is found to be

$$n_{Hall} = \frac{\sigma_0^2}{q \int \mu s(\mu) \mathrm{d}\mu} = \frac{\left(\int s(\mu) \mathrm{d}\mu \right)^2}{q \int \mu s(\mu) \mathrm{d}\mu} \tag{6.19}$$

The Hall mobility is then

$$\mu_{Hall} = \sigma_0 R_H(B \to 0) = \frac{1}{\sigma_0} \int \mu s(\mu) \mathrm{d}\mu = \frac{\int \mu s(\mu) \mathrm{d}\mu}{\int s(\mu) \mathrm{d}\mu} \tag{6.20}$$

Hall Scattering Factor

The Hall scattering factor r is defined such that

$$R_H(B \to 0) = \frac{1}{n_{Hall} q} = \frac{r}{q \sum |n_i|} \tag{6.21}$$

which leads to:

$$r = \frac{n_{Drift}}{n_{Hall}} = \frac{q \sum |n_i| \int \mu s(\mu) \mathrm{d}\mu}{\left(\int s(\mu) \mathrm{d}\mu \right)^2} \tag{6.22}$$

In fact, a scattering factor can now be defined for each peak in the spectrum by limiting the integration to only the peak in question, rather than the whole spectrum:

$$r_i = q n_i \left(\int_{peak} \mu s(\mu) \mathrm{d}\mu \right) \left(\int_{peak} s(\mu) \mathrm{d}\mu \right)^{-2} \tag{6.23}$$

Using the definition of n_i (Equation 6.16) this becomes:

$$r_i = \frac{1}{n_i q \mu_i^2} \int_{peak} \mu s(\mu) \mathrm{d}\mu \tag{6.24}$$

Alternatively, since at high fields $\sigma_{xy}^2 \gg \sigma_{xx}^2$ and $\mu B \gg 1$:

$$R_H(B \to \infty) \simeq \frac{1}{B \sigma_{xy}} = \frac{1}{B} \left[\int \frac{\mu B s(\mu) \mathrm{d}\mu}{1 + \mu^2 B^2} \right]^{-1} \simeq \left[\int \frac{s(\mu)}{\mu} \mathrm{d}\mu \right]^{-1} \tag{6.25}$$

from Equation 6.15 the drift concentration may be described as

$$n_{Drift} = \frac{1}{q} \int s(\mu) \mu \mathrm{d}\mu \tag{6.26}$$

provided that $s(\mu = 0) = 0$, as it physically should be. The drift mobility is now

$$\mu_{Drift} = q \left(\int s(\mu) \mathrm{d}\mu \right) \left(\int \frac{s(\mu)}{\mu} \mathrm{d}\mu \right)^{-1} \tag{6.27}$$

and the Hall scattering factor is:

$$r = \int \frac{s(\mu)}{\mu} \mathrm{d}\mu \frac{\int \mu s(\mu) \mathrm{d}\mu}{\left(\int s(\mu) \mathrm{d}\mu \right)^2} \tag{6.28}$$

(The integrals in Equations 6.27, 6.28 and 6.29 may be performed over individual peaks rather than the whole spectrum, as before.)

A Hall scattering factor greater than unity in a carrier

gas is the result of energy-dependent scattering mechanisms, and according to Equation 6.25 a peak at μ_i will have r_i greater than unity if $\int \mu s(\mu) \mathrm{d}\mu > \mu_i \int s(\mu) \mathrm{d}\mu$. This will be the case for a symmetric, broad peak in the mobility spectrum. In order for a peak to have a Hall scattering factor less than unity, it is necessary for it to be asymmetric with most of its weight towards the origin. However, anisotropy in the energy band which leads to $r < 1$ is expected to produce "harmonics" in the mobility spectrum and Equation 6.25, which only considers a single peak, is not applicable.[28]

All of the significant peaks in the spectra shown Figures 6.5, 6.6 and 6.7 have a scattering factor of unity; small peaks which are thought to be artefacts tend to have unphysical values of r which are far from unity and sometimes even negative (and divergent between values calculated from Equations 6.25 and 6.29). Experience suggests that the shape of the peak in a mobility spectrum may owe as much to the quality of the data as to the scattering processes and band shape of the material, so while the mobility of the 2DEG can be found despite parallel conduction in the dopant slab, little can be said about the link between energy-dependent scattering and mobility-dependent conductivity and therefore about the scattering mechanisms in non-degenerate semiconductors.

6.5. Resistivity and Mobility as a Function of Temperature

Figure 6.10 shows how the resistivity of the n-type device varies as a function of temperature at zero magnetic field, which should be compared with the variation of mobility and sheet density shown in Figure 6.8. The resistivity at room temperature is just over 60Ω/□; at first the resistivity decreases slightly as temperature decreases from 300 K, reaching a minimum around 200 K, and then increases by almost an order of magnitude in the region of T_F to a peak at 30 K. Resistivity then decreases to saturation at just under 300Ω/□ as the temperature reaches 350 mK. This data was obtained using a drive current of 150 nA: the behaviour of the resistivity as a function of current at low temperatures is the subject of Figures 6.13 and 6.14 which are discussed in the next section.

Information from the mobility spectrum analysis (Figure 6.8) of the previous section suggests that the dominant contribution to the conductivity in this device is the 2DEG, even at high temperatures where the dopant slab has not frozen out.

There is a minimum in the resistivity at 160 K which, since the sheet density is only increasing slowly in this range, corresponds to a maximum in the mobility. The decrease in mobility with temperature from here to 300 K can be ascribed to acoustic phonon scattering: $\mu \propto T^{-\gamma}$ with $\gamma \sim 1$.[127] The room temperature value of the mobility in the 2DEG is 4,500 $\mathrm{cm^2V^{-1}s^{-1}}$ (better than the room temperature electron mobility of bulk silicon of 1,400 $\mathrm{cm^2V^{-1}s^{-1}}$)[124] but the sheet density is unphysically high at $1.7{\times}10^{13}\mathrm{cm^{-2}}$. The Hall scattering factor of the 2DEG (from Equations 6.25 and 6.29) is 1.0, so nothing can be inferred regarding scattering mechanisms.[128] Whilst some variation of sheet density with temperature is to be expected,[29] such a large variation may be a sign of a problem with the magnetoresistance data, mobility spectrum theory, or the computer code performing the calculation. These possibilities will be discussed in section 6.8.2.

The change in resistivity between 160 K and 40 K is associated with the change in sheet density as the 2DEG becomes degenerate (and conduction in the dopant layer freezes out). However, inspection of the mobility spectrum results in Figure 6.8 shows how the mobility is at a minimum at 100 K; the change in resistivity seen in Figure 6.10 appears to be, for temperatures between 100 K and 160 K, the result of changing mobility with a roughly constant sheet density. As the temperature falls below 100 K, the mobility increases but the sheet density decreases such that the resistance variation is maintained: there is no significant feature visible in Figure 6.10 to mark the minimum in mobility. This minimum in mobility may be a real physical effect due to the temperature dependence of screening in the 2DEG,[129] but is more likely to a sign that some part of the experimental procedure is awry. The mobility should probably continue monotonically decreasing as $\mu \propto T^{-1}$; the mobility and the density of the 2DEG are overestimated whenever the density in the dopant layer is significant.

The low temperature behaviour of the resistance is the subject of Figure 6.11. The mobility enhancement as the temperature decreases below 20 K is most likely to be a quantum transport effect: the carrier gas is fully degenerate and the phonon scattering rate is very low.

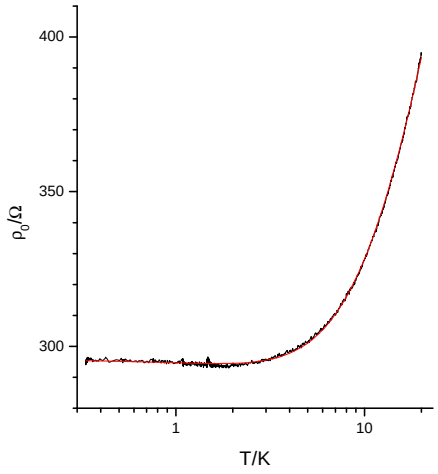

FIG. 6.11: Calculated and measured resistivity as a function of temperature, to be compared with similar results in the previous chapter. Fitting parameters are given in the text.

There is little variation in the ∼1 K region and no suggestion that the resistivity would tend to either zero or infinity at absolute zero. This can be compared with Figure 5.31 and the analysis in Section 5.5.3: $\rho(T)$ in a 2DHG with a sheet density of $1.07\times10^{12}\text{cm}^{-2}$ also decreases as temperature decreases below 20 K, reaching a minimum and then increasing slowly and saturating.

Equation 5.15 was introduced to fit the $\rho(T)$ data shown in Figure 5.31: the results of fitting Equation 5.15 to the data in Figure 6.11 with n fixed at unity are $\rho_0 = 116 \pm 1\ \Omega/\square$, $\rho_c = 179 \pm 1\ \Omega/\square$, $\rho_l = 164 \pm 2\ \Omega/\square$, $T_0 = 4.91 \pm 0.04$ K and $\tau = 1.65 \pm 0.04$ ps. This scattering rate corresponds to a mobility of 15,000 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ which is close to the value of the mobility of the 2DEG at the peak in resistivity at 30 K. Whether these fitting parameters relate to those shown in Figure 5.32 for a 2DHG is a subject for debate and investigation: the values of T_0 and τ from this 2DEG (at a density of $0.90\times10^{12}\text{cm}^{-2}$) would not look out of place in the lower panel of Figure 5.32, but the resistivity parameters are clearly at least an order of magnitude smaller. Further discussion is presented in the conclusions of this chapter, section 6.8.2.

The low temperature maximum value of the mobility is 23,000 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at a sheet density of $9.0\times10^{11}\text{cm}^{-2}$. This is typical for n-type strained silicon, not superlative.[127,130] Possible low-temperature mobility limiting mechanisms are discussed throughout the remainder of this chapter.

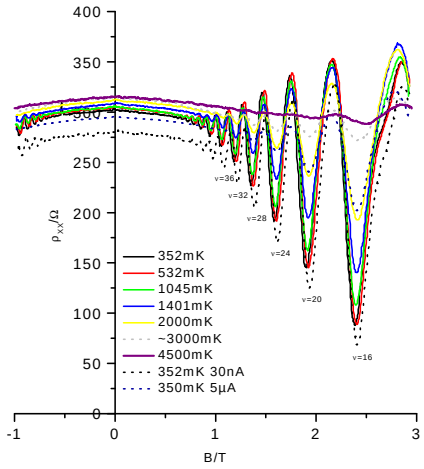

FIG. 6.12: Shubnikov–de Haas oscillations as a function of temperature, at a current of 150 nA. For comparison, two results at 350 mK using different currents are shown. The sheet density found from the period of the oscillations in inverse magnetic field is $4.7\times10^{11}\text{cm}^{-2}$, which must be multiplied by two because of valley degeneracy. Spin degeneracy also applied at these field values, which is why the minima are at filling factors that are multiples of four. (See Section 4.1.3 and Figure 4.2)

6.6. Quantum Hall Effect

6.6.1. Shubnikov–de Haas

Figure 6.12 shows a set of Shubnikov–de Haas oscillations at various temperatures, using a current of 150 nA. For comparison, two additional results are shown at 350 mK with different currents. Since the amplitude of Shubnikov–de Haas oscillations is the same at 150 nA (black solid line) as at 30 nA (black dotted line), there is no heating effect at 150 nA. A current of 5 μA (blue dotted line) heats the electron gas to over 2 K. Using Equation 4.27 to extract the sheet density from the period of the oscillations in inverse magnetic field gives a result of $4.7\times10^{11}\text{cm}^{-2}$ which is half the true value, found from the classical magnetoresistance, due to unresolved valley degeneracy.

A set of oscillations at 350 mK at various currents is shown in Figure 6.13. There is a large y-axis offset between the 3 nA and the 150 nA data, but no major reduction in the amplitudes of the oscillations until the current reaches 1 μA. The oscillation amplitude as a function of current can be compared with the oscillation amplitude as a function of temperature to estimate the effective electron temperature as a function of current at a lattice temperature T_l of 350 mK: at 1 μA, T_e ∼1.2 K and at 5 μA, T_e ∼2.3 K. The energy loss rate per electron as a function of electron temperature can be calculated:[65]

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}t} \right\rangle = \frac{k_B(T_e - T_l)}{\tau_e} = \frac{\rho_{xx} I^2}{l^2 n_s} \qquad (6.29)$$

From this, the energy relaxation time τ_e and thence electron-phonon interaction strength can be found.[55,65,127,131,132] Though this will not be pursued here, the energy relaxation time is relevant to the breakdown of the Quantum Hall state investigated in section 6.6.2. The data of Figure 6.13 reassures us that electron heating has not affected the Shubnikov–de Haas oscillation amplitudes in Figure 6.12.

The zero-field value of the resistivity as a function of current is plotted in Figure 6.14, which should be compared with Figure 6.11. Even though the resistivity seems lower at lower currents (which might suggest a heating effect) the fact that the variation in resistivity with temperature at these lower currents matches that shown in Figure 6.11 refutes this; the effect is more likely due to the difficulties with measuring such small signals. There is no variation of the low-field Hall coefficient with current.

The effective mass found from the temperature decay of the Shubnikov-deHaas oscillations in Figure 6.12 (using the method described in Section 4.1.3, particularly Equations 4.28 to 4.30) is $0.2m_e$ which is in agreement with the value of $0.199m_e$ from the calculations in Reference 14, presented in Section 6.3. The ratio of transport and quantum lifetimes, α, is 3.1.

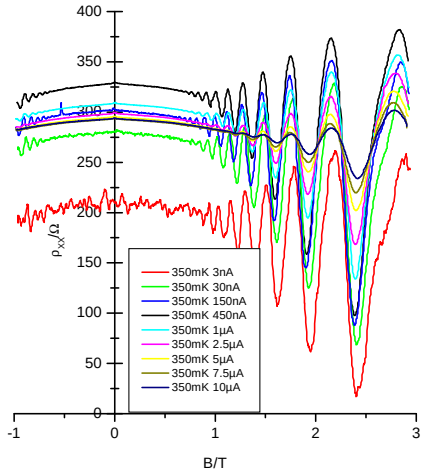

FIG. 6.13: Shubnikov–de Haas oscillations as a function of current, at a temperature of 350 mK. The zero-field value of the resistivity as a function of current is shown in Figure 6.14; the important notion here is that the amplitude of the oscillation is related to the temperature of the carrier gas (Equations 4.24 and 4.25) and there is no major decay in amplitude until the current reaches 1 μA.

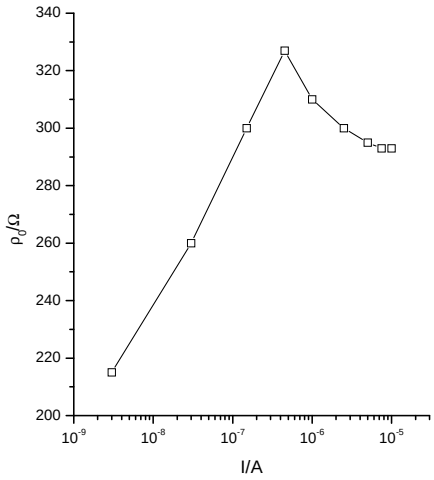

FIG. 6.14: The zero-field value of the resistivity as a function of current. When compared to Figure 6.11 it is clear that the forms for the current and temperature dependent resistivity do not obviously map to each other.

In p-type silicon-germanium where interface impurities limit the mobility (commonly to around an order of magnitude less than the mobility of this 2DEG at a comparable sheet density, as seen in Chapter 5) the value of α tends to be around unity, indicating that scattering is short-range.[133] In heterostructures where the heterointerface is of very high quality and mobilities are over an order of magnitude higher than the mobility of this 2DEG, the value of α can be much larger: this indicates that most of the scattering in the system is small-angle, caused by ionized impurities which are remote from the 2DEG.[133] α values of up to 20 have been reported in very high mobility n-type silicon-germanium systems,[6,134] comparable to that seen in high-quality heterostructures fabricated from other materials.[135]

This value of α of 3.1, then, intuitively indicates a system where impurities both at and remote from the interface influence the mobility. Calculations in Reference 134 suggest that for a spacer of 15 nm between the dopant slab and the active channel the value of α should be over 10 if interface impurities and roughness are ignored, which supports this conclusion. Calculations of the temperature dependence of the mobility, also from Reference 134, show that mobility increases slowly with decreasing temperature in the presence of interface impurities but that interface roughness causes the mobility to decrease with decreasing temperature. Figure 6.8 shows that therefore, interface impurities are limiting the mobility at low temperatures in this system.

Figure 6.15 (repeated from Figure 4.2) shows magnetoresistance data at 350 mK up to 11 T, which reaches a dissipationless state at the $\nu = 4$ quantum Hall plateau. A full discussion is presented in Section 4.1.3. At around 3 T the Shubnikov-deHaas oscillations with minima at $\nu = 4n$ begin to break down. A minimum appears at $\nu = 10$ as valley degeneracy is lifted, and a minimum appears at $\nu = 5$ as spin degeneracy is lifted.

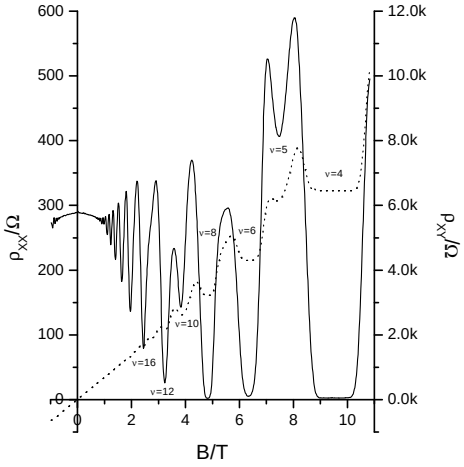

FIG. 6.15: Shubnikov–de Haas and Quantum Hall Effects at a temperature of 350 mK. The dotted line is the transverse resistivity and is associated with the y-axis on the right. Minima in resistivity are labelled with their corresponding filling factors. Conduction is dissipationless at $\nu = 4$. The feature at $\nu = 4$ corresponds to valley splitting, that at $\nu = 10$ corresponds to spin splitting.

6.6.2. IV at a Quantum Hall plateau

In Figure 6.16, data are presented which explore the current-voltage characteristics of the quantum Hall plateau state at 350 mK, around the filling factor of 4. The dissipationless conduction ($\rho_{xx} = 0$) makes this seem superficially similar to a superconducting state (in fact, there are similarities with the phase diagram of a Type II superconductor)[136,137] but since the transverse element of the resistivity tensor has a finite value, $\rho_{xx} = h/(\nu e^2)$, then $\sigma_{xx} = 0$ also. At the centre of this region, the dissipationless conduction breaks down at a critical current of around 3 μA (the device is 1.75 μm wide so the current density is 1.7 Am^{-1}). For the highest filling factor shown, where the behaviour is almost Ohmic, the resistivity is around 500$\Omega/\square$.

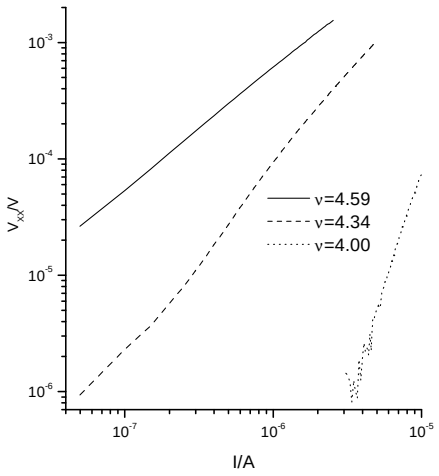

FIG. 6.16: IV data taken close to the dissipationless state at the $\nu = 4$ quantum Hall plateau. At $\nu = 4$ itself, there is essentially no voltage drop (that is discernable above the noise floor) along the sample until a critical current of 3 μA is applied. At higher filling factors (lower fields) there are decreasing deviations from perfect Ohmic behaviour.

The general characteristics and breakdown of the (integer) Quantum Hall Effect have been studied (mainly in n-type gallium arsenide and related systems) since it is exploited as quantum standard of resistance for use in metrology, and is also a system in which quantum phase transitions and interactions can be explored.[138–144]

In Reference 142, the breakdown current density in n-type GaAs/AlGaAs at $\nu = 2$ is around 0.5 Am^{-1} (a current of 210 μA through a Hall bar of width 0.4 mm). Reference 141 finds that at $\nu = 4$ the breakdown current density is 1.7 Am^{-1} (60 μA through 35 μm) decreasing to half this at $\nu = 3.9$ and $\nu = 4.1$. (Reference 144 finds that the breakdown current density in an InAs/GaSb crossed-gap electron-hole system is smaller by three orders of magnitude). The breakdown current decreases with filling factor and temperature in a manner similar to the phenomenological Gorter-Casimir two-fluid model for superconductivity.[136,137]

There is no clear consensus on the mechanism for the breakdown of the Quantum Hall Effect. Indeed, it may be the case that there are two or more unrelated mechanisms.[138,139,142] Breakdown can occur as a sequence of regular steps in longitudinal voltage, with heights of a few millivolts, and this has been linked to tunnelling between Landau level edge states.[142] No such steps were observed in this system. A "bootstrap" self-heating mechanism is proposed in Reference 139, where the Quantum Hall Effect undergoes avalanche breakdown when the transverse electric field reaches a critical value of $E_b = \sqrt{\frac{2\hbar}{m^*\tau_e}}B$. Using (from section 4.1.3) $\rho_{xy} = \frac{V_{xy}}{I} = \frac{E_y w}{I} = \frac{h}{\nu e^2}$ and $n_s = \nu B/h$ the breakdown current density is:

$$\frac{I_b}{w} = n_s e\sqrt{\frac{2\hbar}{m^*\tau_e}} \tag{6.30}$$

The energy relaxation time τ_e (Equation 6.30, found by comparing the temperature and current dependence of the amplitude of Shubnikov–de Haas oscillations in Figures 6.12 and 6.13) is $\sim 10^{-8}$s so the breakdown current density is estimated to be 0.5 Am^{-1}. To obtain the measured value of 1.7 Am^{-1} τ_e would need to be less than 10^{-9}s. Nevertheless, this agreement is good and may be improved by more rigorous determination of τ_e.

6.7. The Effects of Infra-Red Radiation

6.7.1. Shubnikov–de Haas

Figure 6.17 demonstrates the effect of illuminating the sample with a gallium-arsenide infra-red light emitting diode at 350 mK. It can be seen that the illumination increases the resistance at zero field, and changes the Hall coefficient and the period of the oscillations. The amplitude of the oscillations does not decrease, showing that there are no heating effects. The ratio of quantum and transport lifetimes, α, is 3.6 during the illumination (slightly larger than before) and the changes in sheet carrier concentration are summarized in Table 6.1. While the sample is being illuminated, carriers appear to be driven out of the conducting channel. Following illumination some, but not all, return. The mobility appears to change in line with the density.

The low-field magnetoresistance (before oscillations begin: high-field magnetoresistance can not be used to explore the mobility spectrum in the presence of Landau levels) does not suggest that any parallel conduction channels are formed by the action of the light: even though the zero-field value of the resistivity is changed, the form of the magnetoresistance is identical. Parallel conduction would be apparent as a change in the gradient of $\rho_{xx}(B)$ with magnetic field, as can be seen in Figure 4.1.

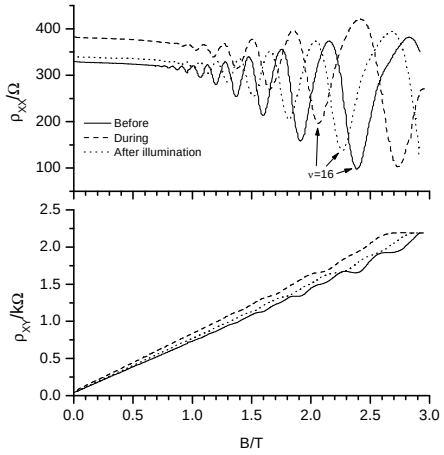

FIG. 6.17: Magnetoresistance results at 350 mK, showing the effects of illumination by infra-red radiation.

TABLE 6.1: A summary of the changes in carrier concentration in response to infra-red light, from Figure 6.17. The Hall results are calculated from the gradient of $\rho_{xy}(B)$ using Equation 4.7 and the Shubnikov–de Haas results are calculated from the period of the oscillations of $\rho_{xx}(B)$ in inverse magnetic field using Equation 4.27, including a factor of 2 for valley degeneracy.

Sheet Density [$10^{11}\mathrm{cm}^{-2}$]	Hall	Shubnikov–de Haas
Before:	8.8	9.4
During:	8.0	7.6
After:	8.5	8.8

6.7.2. Resistance as a Function of Temperature

Two results are presented in Figure 6.18: the continuous set shows how the resistivity varies with temperature for the device under illumination compared to the device in the dark, as it was cooled. The points joined by lines were taken at stable temperatures, with the device subjected to periods of illumination to observe how its resistivity changed with time. The transient at the beginning of the period of illumination was always very sharp, but the fall-off at the cessation of illumination showed a finite decay time, particularly at low temperatures. As Figure 6.19 shows, at 4.2 K a decay time of the order of 100 seconds was identified, this falls rapidly to the order of 10 seconds at 10 K and then becomes too short to discern behind the 3 second time constant of the measuring equipment. The energy involved is therefore roughly the value of k_BT at 10 K, $\sim$1 meV.

At temperatures less than 10 K, the irradiation simply increases the resistivity of the device by reducing both the sheet carrier concentration and the mobility. Since there is no evidence from the magnetoresistance for the formation of a parallel conduction channel of a mobility comparable to that of the 2DEG, and such a channel would tend to decrease the resistivity of the whole device, then the simplest interpretation is that the radiation causes 10% of the carriers to become unavailable for transport. The energy from the radiation clearly perturbs the equilibrium position (which then takes of the order of at least a few hundred seconds to become re-established) so it is

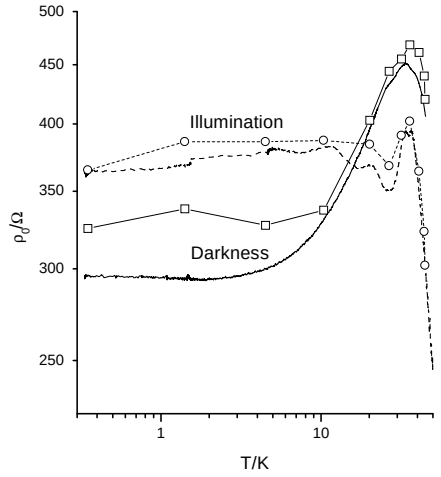

FIG. 6.18: The resistivity as a function of temperature, demonstrating the effects of infra-red radiation. The continuous traces were taken by cooling the device under constant (illuminated or dark) conditions; the points joined by lines show results taken at stabilized temperature values with intermittent illumination.

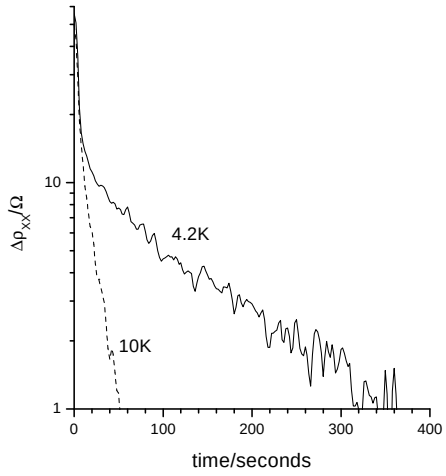

FIG. 6.19: The resistivity as a function of time, demonstrating the effects of infra-red radiation. At time zero, illumination ceases, and the value of resistivity decayse back to its dark value. At 4.2 Km the decay is much more persistent than at 10K.

possible that carriers are caused to move out of the quantum well and into the alloy spacer layer, where they are drawn towards ionized donor atoms. Upon cessation of the radiation, equilibrium is re-established.

The associated slight reduction in screening may be the cause of the increase in scattering rate. The slight increase in α suggests that the screening effects are more effective for the remote impurities than the local (interface) impurities, since large α values suggest remote impurity scattering, whereas $\alpha \sim 1$ is characteristic of scattering from impurities at the heterointerface where the 2DCG is defined.

Above 20 K, the radiation tends to reduce the resistivity. The sharp fall-off which begins at around 40 K parallels that which occurs in the device, in the dark, around the Fermi temperature; this is consistent with the reduction in the 2DEG carrier concentration and the consequent reduction in T_F.

The region between 15 K and 40 K is more interesting, showing what appear to be local minima and maxima in resistivity, and may be worthy of more investigation, particularly since the formation of Landau levels will be suppressed at this temperature so that mobility spectrum analysis may discover a low-mobility parallel conduction channel. At room temperature, there is no noticeable response to illumination.

6.8. Conclusion

6.8.1. Mobility and Resistivity

The mobility of the electrons in the 2DEG of this n-channel device is understandably superior to that of the holes in the p-channel devices discussed in the previous chapter. At room temperature, their mobility is also superior to the electrons in an n-channel MOSFET or indeed in bulk silicon, according to both low-field Hall effect and mobility spectrum analysis.

Since the effective mass difference alone cannot account for this (that of the electrons here is $0.2m_e$ whereas the holes in the previous chapter were found to have an effective mass of around $0.3m_e$) the scattering time must be mainly responsible.

The low resistivity of the device at low temperatures leads (via Equation 5.1) to a very high $k_F l$ value of almost 100, which (via Equation 5.2) suggests a dephasing time at 1 K of over 100 ps. The low-field magnetoresistance at very low temperatures shows no evidence for weak-localization (as seen in Figure 5.10) which suggests that $\tau_p si \sim \tau$.

An elastic scattering time as long as this could only be the result of an extremely low impurity density in the active channel, and a very smooth interface, compared to characteristic p-channel devices. This is backed-up by the ratio of quantum and transport lifetimes found from the field dependence of the amplitudes of the Shubnikov–de Haas oscillations being larger than one. Also, the fact that when infra-red radiation is used to reduce the sheet carrier density the mobility also decreases, suggests that the mobility is limited by screened interface impurities rather than by interface roughness.

The high quality of the interface may be due to the normal structure of the device, with the dopant slab grown after the active channel and therefore with less time to diffuse into it compared to inverted structures. However, a gated HMOS device would either be inverted or not be modulation doped at all.

The form of the resistivity at low temperature, shown in Figure 6.11, is similar to the result in Figure 5.31 for the Siemens device at a sheet density of $1.1{\times}10^{12}\text{cm}^{-2}$ (which is the closest sheet density to the 2DEG). There is a weak minimum in resistivity at 2 K but as temperature continues to decrease the resistivity appears to be saturating. Equation 5.15 has been fit to the data, but with $n = 1$. Much more work would be necessary to establish any physical justification for the applicability of Equation 5.15 to the data in Figure 6.11.

6.8.2. Mobility Spectrum Analysis

Methodology

Mobility spectrum analysis is a general technique that should be applicable to any system where the mobility of the charge carriers is not a function of magnetic field. The lower limit of mobility than can be reliably found is roughly the reciprocal of the maximum magnetic field that the data extends to; the upper limit is the reciprocal of the spacing between the data points in inverse magnetic field.

In semiconducting samples and cryomagnetic systems, the upper limit is practically set so that the resulting mobility spectrum is detailed enough in the region where carriers are expected to be found but does not waste points elsewhere. However, the lower limit is often a serious issue. Hence, this analysis method is much easier to test on an n-channel device such as this, compared to a p-channel device which may have a mobility smaller by a factor of 5 or 10.

Many mobility spectrum analysis methods do not respond well to large numbers of data points (requiring human choices of which data points to keep and which to discard) or employ tricks, assumptions, interpolations or extrapolations in order to recondition the data or guide the solution. A new method is presented here (Bryan's Algorithm Mobility Spectrum: BAMS) which uses singular value decomposition to remove noise and overspecification from the data set and the principles of Bayesian inference and Bryan's-algorithm maximum entropy to find a mobility spectrum. (It is quite distinct from the maximum entropy mobility spectrum method in Reference 29).

Singular value decomposition carries out two essential task: firstly, by removing overspecification from the data

(that is, by finding lines within the matrix which are linear combinations of each other) the basis space which must be searched for a solution is reduced dramatically. Secondly, by discarding elements of the matrix which are smaller than the noise floor, the inversion is protected against the tendency to fit very well to features of the data which are noise at the expense of the underlying, physical data.

The maximum entropy method also performs two functions. Firstly, since the use of a default model allows information about the expected solution to be added in, setting a positive and normalizable default model which has a constant value across the whole spectrum ensures (due to the logarithmic form of the entropic regularization) that the resulting mobility spectrum will be positive and normalizable. Secondly, the hyper-parameter α nominates the relative importance of a very good least-squares fit to the data versus a physically reasonable spectrum which contains as much information as it is reasonable to extract from the data without introducing spurious artefacts.

Bryan's algorithm chooses solutions which are the most probable regarding the value of the hyper-parameter, and the refinement of this method by Hague (use of a solution as the default model to find a better solution) ensures that the final value of the hyper-parameter is well within the bounds of the applicability of Bryan's algorithm.

One drawback with the maximum entropy technique is that it is very important to have an accurate estimation of the error in the experimental data. Tests with synthetic data, however, lead to the conclusion that overestimation of the error acts on the solution in an intuitive way, broadening peaks and washing out fine detail; underestimation of the error leads to solutions which can be rejected due to their obvious unphysicality. Whilst this is clearly not satisfactory (we would like to always achieve the best result possible and trust what we find) there has been no cause to worry that results which appear physical but are wrong are likely to be produced by mistakenly assuming a certain error level. A more analytical treatment of the level of error is an important line for further work, but the human process of rejecting data which is errant because of equipment mis-calibrations or failures will always be necessary.

Methods of obtaining classically measured quantities from a mobility spectrum have been described, and although mobility spectra are rarely of such good resolution that the Hall scattering factor can be unambiguously found, these analyses at least provide a check for unphysicality. It has been argued that a common feature in mobility spectra, the "mirror peak," is an artefact due to discrepancy between the relative values of the longitudinal and transverse values of the conductivity tensor.

Results

The results produced by the mobility spectrum analysis, shown is Figures 6.5 to 6.10, are feasible at low temperature but seem to become anomalous as temperature increases. The low temperature 2DEG sheet density agrees with calculations (see section 6.3) but the room temperature sheet density (nearly 20 times greater than this) is not supported by simulation of the band profile using a numerical solution of the Poisson equation. In fact, the room temperature 2DEG density should be much closer to its low temperature density.[29] However, to maintain the same conductivity at such a density would require ridiculously high mobility.

The sheet density of the dopant layer was found to reach $1.5\times10^{13}\text{cm}^{-2}$ at room temperature. Since the doping slab is 30 nm thick, this corresponds to an average carrier concentration of $5\times10^{18}\text{cm}^{-3}$ and this value is typical for the donor concentration in modulation doping. However, the mobility of the dopant layer was found to be 1000 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ meaning that overall its resistance was 400 $\Omega/\square$. For a 30 nm thick layer, this corresponds to a resistivity of $10^{-3}\Omega\text{cm}$. To achieve such low resistivities the donor concentration must be almost 10^{20}cm^{-3} since the mobility of a doped semiconductor generally decreases with the doping dose.[10,124,145] In other words, the mobility of the dopant layer returned by this mobility spectrum analysis is probably too high.

However, to ascribe less of the device's conduction at room temperature to the dopant layer (by reducing its mobility without increasing its sheet density) means that either the mobility or density of the 2DEG must be increased and these two parameters are already worryingly large. According to BAMS there are no other conducting channels within the range $|\mu| \leq 10\ \text{m}^2\text{V}^{-1}\text{s}^{-1}$.

Other established analyses yield similar results for mobility and sheet density of the 2DEG and dopant layer; BAMS analysis of other magnetoresistance data from p-channel systems yields more reasonable results, with the 2DHG density varying by less than 20% between 50 K and 300 K. This means that it is unlikely that there is a problem with the BAMS code. Also, even though mobility spectrum theory is possibly not as useful in a quantitative way as was originally hoped[28] (since it says little about the energy dependence of the relaxation time) it has proved its worth in a qualitative or semi-quantitative way, finding feasible mobility and sheet density results in many cases.

The issue seems to be that the room temperature resistance of the device is only 70 $\Omega/\square$, and the only way in which the resistance can be so low is for there to be a large number of mobile carriers. Since the low temperature results are feasible it is unlikely that the measurement system itself is at fault, and so there is no reason to suspect that the observed drop in resistance of almost an order of magnitude between 30 K and 100 K is not real. Also, BAMS tends to reject unphysical data outright, without returning any solution.

There may be, then, bulk conduction throughout much of the substrate through current paths which undermine the simple conversion of V_{xx} and V_{xy} to elements of the resistivity tensor of the device.

As for finding the energy dependence of the scatter-

ing time, more directly related to the scattering mechanisms limiting transport than the mobility on its own, a method based on the sheet density and/or temperature dependence of the mobility (obtained by BAMS if necessary) may probably be more promising than attempting to analyze the mobility spectrum peak shape.

6.8.3. Other Results

The IV characteristics of a dissipationless $\rho_{xx} = \sigma_{xx} = 0$ quantum Hall state have been briefly investigated. At exactly $\nu = 4$ the dissipationless state breaks down at a current of 3 μA; as ν is increased linear Ohmic behaviour is eventually recovered. The response of the device to infra-red radiation is also explored: at very low temperatures the free carrier density is reduced, the mobility drops, and α increases. This correspondence is seen as evidence that it is screened interface and remote impurities that limit the mobility, not interface roughness.

7. COUPLED CHANNEL DEVICES

7.1. Abstract

Chapters 5 and 6 explored the transport properties of holes and electrons respectively in silicon-germanium heterostructures. In this chapter, devices are presented in which electron and hole gases can both be induced by suitable gate bias conditions. The electron gas is formed in an inversion layer at the interface between silicon and silicon dioxide in the conventional n-channel MOSFET manner; the hole gas is formed by inverted modulation doping of a quantum well of pseudomorphically-grown $Si_{0.8}Ge_{0.2}$ alloy.

The heterointerface at which the hole gas forms is either approximately 100 nm (on wafer 50/51) or 40 nm (on wafer 50/53) away from the oxide interface at which the electron gas forms. The electron and hole systems are contacted separately by sets of n^+ and p^+ contacts, the former being implanted to a depth of around 25 nm and the latter essentially reaching down through the whole heterostructure.

These devices were initially conceived and designed to explore the novel transport phenomena which interactions between co-existing electron and hole gases in silicon-germanium may invoke. The discussion will show that the structures investigated cannot support co-existing electron and hole gases, and alternative designs will be presented.

7.2. Introduction

There are two contrasting aims behind attempts to create a semiconductor structure that features spatially separated, interacting, 2-dimensional charge carrier gases. There is the desire to study the fundamental interaction between carriers, in order to develop a better understanding of the physics of semiconductor materials. Then, there is also the desire to create a device with operation analogous to a CMOS integrated circuit, but with p-type and n-type channels arranged vertically rather than laterally.

7.2.1. Interactions

Lozovik and Yudson suggest a new mechanism of superconductivity based on the pairing of spatially separated electrons and holes due to their Coulombic attraction.[146] Their calculations indicate that the critical temperature may be as high as a few hundred Kelvin for reasonably realistic experimental parameters. This possibility has been considered in more detail since then.[147–150] Thakur, Neilson and Das estimate that the superconducting transition temperature for experimentally accessible carrier densities and layer separations is 100 mK.[150]

Electrons and holes, when they are both present in a semiconductor, can pair up by way of their Coulombic attraction. A bound electron-hole pair (an exciton) is boson-like, since electrons and holes are both fermions. In analogy with Cooper pairs in a superconductor (boson-like pairs of electrons, bound together by phonons) a collection of excitons is expected to undergo Bose-Einstein condensation under appropriate conditions.[149] One of these conditions is that the recombination lifetime of the electron-hole pair is longer than the thermal energy relaxation time (see Equation 6.30) so that the condensed state can form and be observed, usually by photoluminescence.[151,152] In indirect-gap semiconductors, such as bulk silicon and germanium, the recombination lifetime satisfies this condition but it has been established that the exciton condensate can not exist due to competition with a more stable electron-hole plasma droplet phase.[149]

Heterostructures provide a method of extending the recombination lifetime: by suitably arranging the materials of the structure (and if necessary applying an appropriate electric field) the valence and conduction bands can be bent into a profile which localizes electrons and holes seperately in their own quantum wells, as shown in Figure 7.1. Coulombic interaction can still lead to the formation of bound electron-hole pairs but, provided the potential barrier between the carrier gases is high and/or wide enough to prevent significant tunnelling, recombination is prevented.

Experimental work on interacting, spatially separated carrier gases has been performed using GaAs/AlGaAs heterostructures, using photoluminescence techniques to directly probe for the presence of excitons.[151,152] However, the transport properties of the interacting electron-hole layers are of the most interest and there is experimental work of this nature on GaAs/AlGaAs[154,155] and

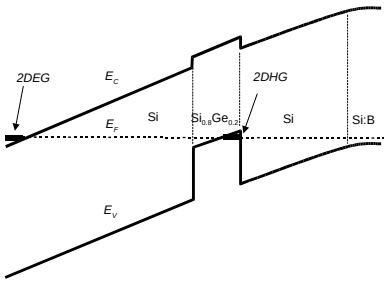

FIG. 7.1: Proposed band profile of coupled-channel devices.[153] Electron and hole gases co-exist within ∼100 nm of each other.

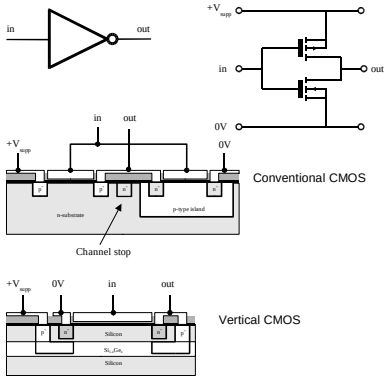

FIG. 7.2: Conventional CMOS versus vertical CMOS: the case for a simple inverter is shown, with the circuit schematic at the top right. In the case of familiar conventional CMOS, the p-channel and n-channel devices sit side by side and the conducting channels are formed at the interface between the silicon and the oxide.[157] In the case of vertical CMOS, a positive gate bias forms an n-channel at the oxide interface and a negative gate bias forms a p-channel at the upper heterointerface, as the band profiles in Figure 7.3 show.

InAs/GaSb/AlSb[156] structures (interesting because the conduction band of InAs falls at a lower energy than the valence band edge in GaSb) but not Si/SiGe. These structures must be designed so that independent electrical contact to the electron and hole gases can be made.

7.2.2. Vertical CMOS

A device which features vertically separated 2-dimensional electron and hole gases, where their carrier densities can be modulated, presents itself as an alternative to the standard CMOS architecture which features modulated electron and hole gases arranged laterally as MOSFETs as part of a silicon integrated circuit.[10,157] The use of silicon-germanium technology within conventional CMOS architecture would seem to require complicated heterostructure design in any case (if the full benefits of compressively strained germanium rich layers for p-channel conduction and tensile-strained pure silicon for n-channel conduction are to be reaped)[1,2,158–161] so it is worth investigating the possibility that novel design may find additional benefits to silicon-germanium CMOS technology.

An inverter is shown in Figure 7.2, in terms of circuit schematics, conventional CMOS structure and proposed vertical CMOS structure. In this case, interaction between the two carrier gases is not desired but this is unlikely to be a problem for a device working at room temperature. As the band profiles in Figure 7.3 show, suitable gate bias causes either an electron gas or a hole gas to form, but the two gases do not co-exist as in Figure 7.1.

7.3. Structure and Fabrication

The heterostructure and device design and processing work was completed by Dr. C. J. Emeleus, in conjunction with Dr. M. A. Sadeghzadeh.[15]

The heterostructures, described by Figure 7.4, were grown by MBE on three 4" silicon n^- 8-12 Ωcm substrates. 300 nm of intrinsic silicon (at 830°C) was followed by 40 nm of silicon (at 500°C) doped with boron to a concentration of 2?1018cm-3 giving an integrated dose of 8×10^{12}cm^{-2}. A 34 nm intrinsic spacer layer was then grown (at 500°C) followed by 15 nm of $Si_{0.8}Ge_{0.2}$ alloy at 650°C. This is double the equilibrium critical thickness of this alloy concentration but well below the metastable thickness for this growth temperature.[9] The intrinsic cap layer was either 90 nm (for wafers 50/51 and 50/52) or 30 nm (for wafer 50/53).

The 2-dimensional hole gas (2DHG) areas were defined at this point by etching this heterostructure away between devices on each chip, and the wafers were subject to an RCA clean which grows 1 or 2 nm of oxide.

A 9 or 10 nm native oxide was further grown on wafers 50/51 and 50/53, consuming 4 or 5 nm of the cap layer. 300±50 nm of low temperature oxide (LTO) was then deposited using a CVD gas-flow technique at Southampton University. The native oxide should provide a better silicon-oxide interface compared to that formed when LTO is deposited directly onto silicon. However, the native oxide process required one hour at 720°C which may

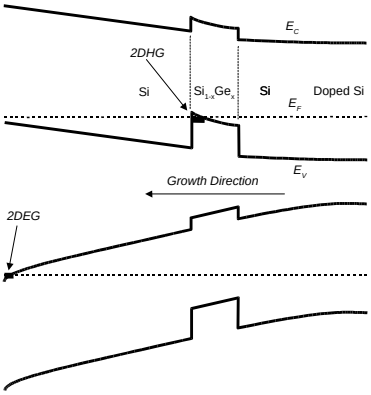

FIG. 7.3: Proposed band profiles of a vertical CMOS device. The substrate (here shown to be slightly positively doped) is to the right, and the left edge of the profile is the silicon–silicon dioxide interface. Negative gate bias (upper profile) forms a hole gas at the top of the $Si_{1-x}Ge_x$ alloy and positive gate bias (lower profile) forms an electron gas at the oxide interface. The gases do not co-exist, as in Figure 7.1. If the gate bias in the upper panel becomes slightly more negative, then the valence band will reach the Fermi level and a hole gas will form at the oxide interface.

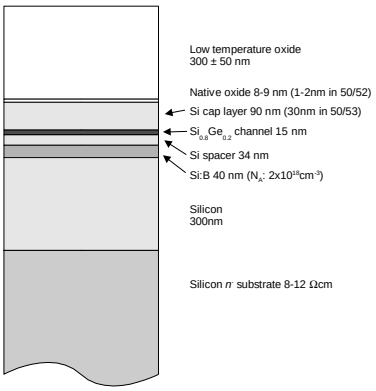

FIG. 7.4: The heterostructure as grown on wafer 50/51. 50/52 and 50/53 differ in the manner shown.

cause some relaxation of the strain in the alloy layer or diffusion of boron dopant atoms. The relative performance of 50/51 and 50/52 should make this clear.

Shallow n^+ contacts were made using low energy (10 keV) implantation of arsenic ions before the native oxide was grown, in a process developed by Dr.C.P.Parry.[153] Deep p^+ contacts were made using 70 keV implantation of BF_2^+ through the native oxide, before the LTO deposition process. Simulations (using the TRIM program of J. F. Zeigler and J. P. Biersack)[162] confirm that the n^+ implantations reach a depth of 20 nm, and the p^+ implantations reach a depth of over 200 nm, well beyond the Boron doping layer. Windows were patterned through the oxide and Ti+Al/Si metal contacts were made.

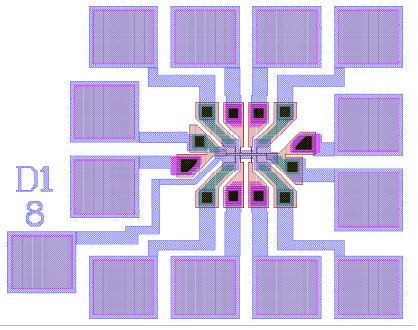

FIG. 7.5: The pattern of a double Hall bar, a typical device. Metal (for contact pads and the gate) is shown in blue. Black areas are windows through the oxide. Red areas show the extent of the heterostructure (Figure 7.4) itself. Shallow n^+ implants ar green, deep p^+ implants are pink. The Hall bar itself is 20 μm wide.

Most of the devices are Hall bars, many with gates and full sets of n^+ and p^+ contacts as shown in Figure 7.5. The aim of the design was to realize the band profile shown in Figure 7.1.[153] Two cap layer thicknesses are available: devices on wafer 50/53 with its smaller cap thickness of 30 nm should be more likely to show interaction effects than devices on wafer 50/51 or 50/52, but are also more likely to suffer from possible problems caused by the shallow n^+ contacts reaching the 2DHG.

Expected Electronic Properties

An estimate of the expected sheet density can be found by considering Figure 3.1 and Equation 3.1. If E_0 and E_A are ignored and the field created by the sheet of carriers is equated with the field required to change the potential by the heterojunction valence band offset ΔE_V over the length d:

$$\frac{p_s q}{\epsilon_0 \epsilon_{Si}} = \frac{\Delta E_V}{d} \tag{7.1}$$

Assuming an offset value of 140 meV for $Si_{0.8}Ge_{0.2}$[14] results in an expected sheet density of 3×10^{11}cm^{-2} as designed by Dr.M.A.Sadeghzadeh using a self-consistent Poisson–Schrödinger calculation.[15,21,153,163]

Applying a positive bias to the gate should produce an inversion layer of electrons at the silicon/silicon dioxide interface and form a band profile as in Figure 7.1. A key feature is that the p^+ implants should exclusively contact the 2DHG in the alloy, and the n^+ implants the 2DEG inversion layer. Each system of n^+ or p^+ contacts and gating should form a MOSFET-like device and this should be readily visible in room-temperature IV characterization.

7.4. Results of IV Characterization

The simplest way of characterizing the structures was to connect suitable gated Hall bar devices in a MOSFET configuration at room temperature. Contacts at either end of the bar were designated (arbitrarily) as source and drain, and voltages were measured relative to the source. The shallow contacts are expected to contact to the inversion layer induced by application of a suitable gate voltage, in exactly the way an n-channel MOSFET operates.

7.4.1. Basic 2-Terminal Characterization at 300 K

Shallow Contact Results

Figure 7.6 shows the drain current characteristics of a gated Hall bar from wafer 50/51. The drain-source voltage, V_{DS}, was applied along the Hall bar via n^+ contacts, so the drain current I_D should be being passed through the device mainly by electrons in an inversion layer at the silicon/silicon-dioxide interface. The characteristics are similar to those of an n-type MOSFET: As the gate voltage is made more positive, the drain current increases. The gradient (that is, the conductivity) of the curves in Figure 7.6 changes when V_{DS} is approximately 0.5 V. However, as V_{DS} continues to increase to over 1.0 V, it can be seen that drain current increases for all gate voltages. This is similar to punch-through, which arises in short-channel MOSFETs when the depletion layer created by large V_{DS} extends all the way from the drain to the source.[157] However, the effect seen here is unlikely to be related directly to punch-through since the Hall bar is 160 μm long, over a hundred times longer than the channel of a "short channe" MOSFET.[157]

Figure 7.7 shows that the drain current has a clear threshold voltage (V_{TS}=15 V) at which $I_D \propto (V_{GS} - V_{TS})^2$. The gradient of the I_D vs. V_{GS} lines (the transconductance) becomes smaller again as V_{GS} increases further and the channel saturates. No significant current flows for gate bias values between zero and -100 V. (These gate voltages are two orders of magnitude greater than those for typical MOSFET device operation since the gate oxide is very thick.)

Deep Contact Results

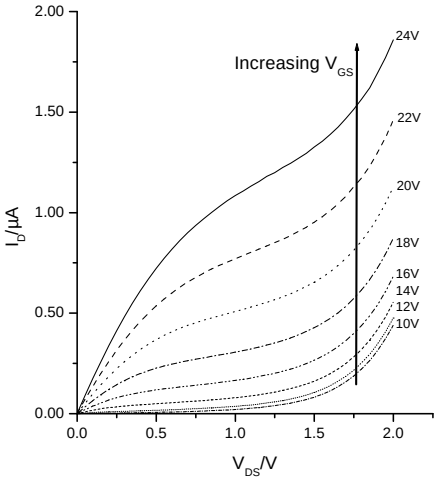

FIG. 7.6: Drain current characteristics of a gated Hall bar on wafer 50/51, as measured using shallow n^+ contacts. Characteristics are similar to those of an n-type MOSFET.

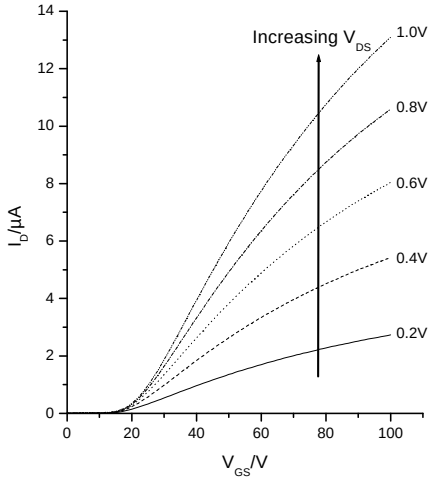

FIG. 7.7: Transfer characteristics of a gated Hall bar, as measured using shallow n^+ contacts. Characteristics are similar to those of an n-type MOSFET.

Figure 7.8 and Figure 7.9 should be compared to Figure 7.6 and Figure 7.7 respectively. Figure 7.8 shows that the (2-terminal) resistance as measured between deep p^+ contacts depends very little on the magnitude of V_{DS} but is influenced by the gate voltage V_{GS}. This can be seen more clearly in Figure 7.9. At positive gate voltages, there is very little variation of I_D with V_{GS}, for a given drain-source voltage but as V_{GS} becomes more negative, I_D increases. Figure 7.7 shows that no current flows between the shallow n^+ contacts for gate voltages smaller than 15 V, provided V_{DS} is below 1.0 V. It does not seem to be possible to prevent conduction between the

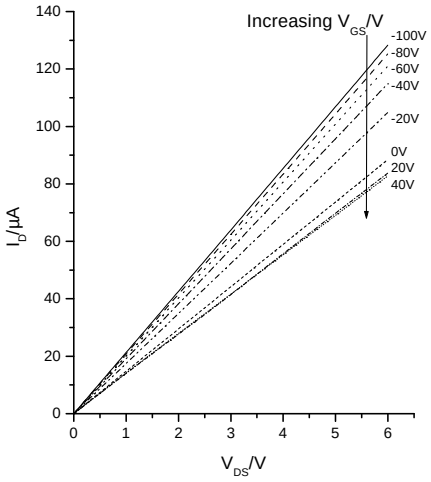

FIG. 7.8: Drain current characteristics of a gated Hall bar, as measured using deep p^+ contacts. Characteristics bear little resemblance to those of a p-type MOSFET.

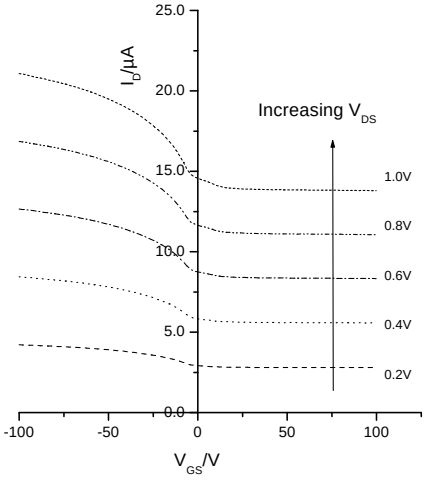

FIG. 7.9: Transfer characteristics of a gated Hall bar, as measured using deep p^+ contacts, following on from Figure 7.8.

deep contacts with suitable gate bias, so either it is impossible to turn the p-channel off or there is conduction through the boron doping layer at all times.

Interpretation of Room Temperature Results

The n-type MOSFET-like behaviour seen in the shallow contact characterization implies that at gate voltages more positive than 15 V an inversion layer forms at the silicon/silicon dioxide interface. If there was some way in which the shallow contacts were providing a current path through the hole gas (which is presumably buried deep in the structure) then there would be no way of preventing drain current flow through the shallow contacts with gate bias: if the deep contacts are contacting the hole gas then Figure 7.9 shows that the hole gas is never depleted to the extent that it does not conduct. This implies that the shallow contacts do indeed only contact to the electron inversion layer for V_{DS} smaller than 1.0 V.

However, as the drain-source voltage across the shallow contacts is increased it can be seen that drain current flows even for sub-threshold gate voltages. Since it is unlikely that there is a direct path between shallow contacts (which are very far apart by usual MOSFET standards) then it is probable that the current is somehow passing through the hole gas. The non-zero V_{DS} needed to invoke this conduction path implies that there is some kind of barrier, probably a depletion zone, between the shallow contacts and the hole gas.

All the above results are from wafer 50/51. By contrast, devices on wafer 50/52 showed very little increase in drain current between p^+ contacts as gate bias was made more negative. This may suggest that the increase seen in Figure 7.9 is due to the formation of a hole gas at the oxide interface; the difference in quality of this interface between 50/51 and 50/52, due to only the former having a native oxide, may account for the increase in conductivity in 50/51 but not in 50/52, for large negative gate bias.

Reduced Separation Between 2DEG and 2DHG

Figure 7.10 shows the drain characteristics of a device on wafer 50/53. This wafer differs from 50/51 (as presented in the structure shown in Figure 7.4) in that the thickness of silicon between the alloy layer and the oxide is only 30 nm in 50/53, where in 50/51 it was 90 nm. Figure 7.10 should be compared to Figure 7.6: quite a high current passes almost regardless of V_{GS} unless V_{DS} is less than half a Volt where the modulation of the inversion layer varies the current.

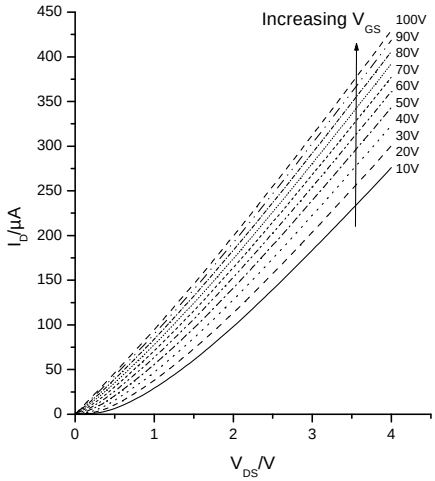

FIG. 7.10: Drain current characteristics of a gated Hall bar device on wafer 50/53, as measured using shallow n^+ contacts.

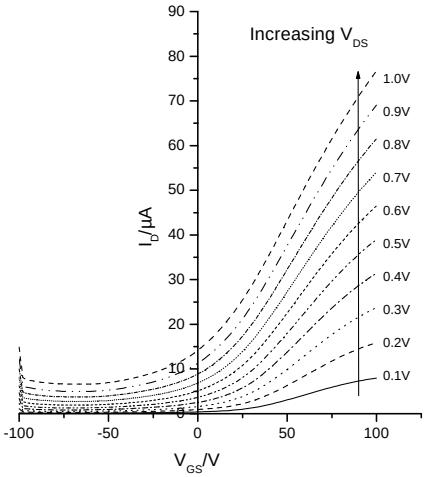

FIG. 7.11: Transfer characteristics of a gated Hall bar on wafer 50/53, as measured using shallow n^+ contacts. The anomaly at V_{GS}=-100 V is caused by the limited compliance of the gate voltage source and the capacitance of the device.

This is not surprising if it is assumed that in 50/53 at room temperature, the shallow n^+ contacts reach close enough to the hole gas in the alloy layer to allow significant conduction at lower V_{DS} values than were needed to cause similar effects in 50/51.

Figure 7.11 shows this same effect more strikingly: in contrast to Figure 7.7, it is almost impossible to prevent current flowing between the shallow n^+ contacts unless a very small V_{DS} is applied.

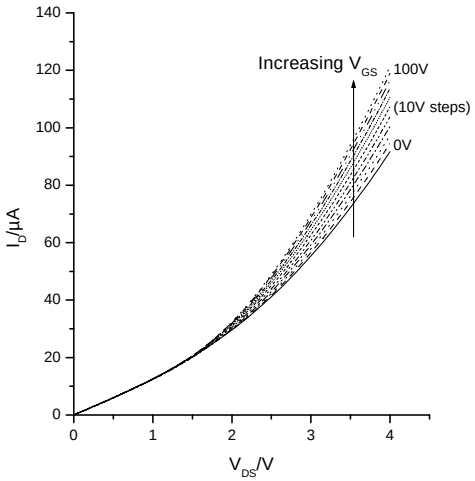

FIG. 7.12: Drain current characteristics of a device on wafer 50/53, as measured using deep p^+ contacts under the influence of a *positive* gate bias.

If it is possible to pass current between the shallow n^+ contacts via holes, it should be expected that it will be possible to pass current between the deep p^+ contacts via electrons and Figure 7.12 demonstrates this: for V_{DS} less than 1.5 V, the current between the deep p^+ contacts is not influenced by the application of a positive gate bias. An inversion layer of electrons is forming at the silicon/silicon-dioxide interface but the resistance between the deep contacts is only visibly changed as V_{DS} increases past 1.5 V.

The resistance between the deep p^+ contacts is de-

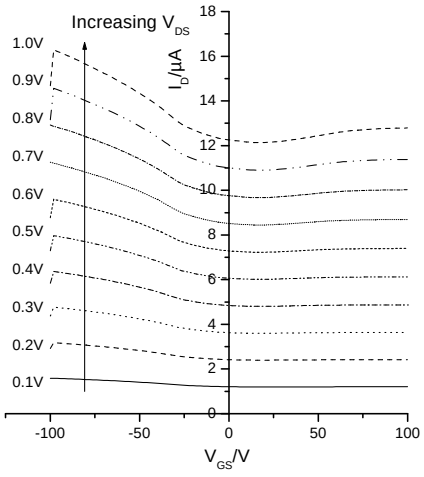

FIG. 7.13: Transfer characteristics of a gated Hall bar on wafer 50/53, as measured using deep p^+ contacts. The anomaly at V_{GS}=-100 V is caused by the limited compliance of the gate voltage source and the capacitance of the device.

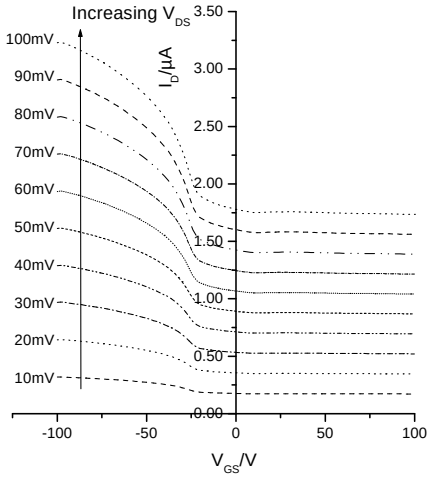

FIG. 7.14: Transfer characteristics of a (different) gated Hall bar on wafer 50/53, as measured using deep p^+ contacts. Smaller V_{DS} has been applied in order to minimize conduction through the electron inversion layer, so that this figure can be compared in form with Figure 7.9.

creased by positive gate bias, for large enough V_{DS}. Since it is unlikely at this temperature that depleting the hole population would lead to an increase in current flow then this effect must be caused by conduction through the electron inversion layer. This can be seen in Figure 7.13 (V_{DS} up to 1 V) but not Figure 7.14 (V_{DS} up to 100 mV): Figure 7.14 is similar in form to Figure 7.9 with roughly constant drain current at all positive gate bias values, but Figure 7.13 shows that drain current increases slightly as V_{GS} becomes more positive.

Room temperature measurement serves two purposes: firstly, it demonstrates the device as if it were a component in an everyday circuit (good performance in this context being the bottom line for efforts in silicon-germanium research) and also it demonstrates that it is worth performing characterization at low temperatures where device parameters can be extracted more easily.

7.4.2. 2-terminal Characterization at Low Temperature

Contact Issues

Making reliable electrical contacts to semiconductor material has always been an issue. In the simplest case of metal on extrinsic non-degenerate silicon, there are two alternatives. If an n-type semiconductor is contacted with a metal and the work function of the metal is greater than of the semiconductor, or if a p-type semiconductor is contacted with a metal such that the work function of the metal is less than of the semiconductor, a depletion layer will form in the semiconductor at the junction, resulting in a rectifying Schottky contact.

However, if an n-type semiconductor is contacted with a metal and the work function of the metal is less than of the semiconductor, or if a p-type semiconductor is contacted with a metal such that the work function of the metal is greater than of the semiconductor, there will be no depletion layer and the contact will be Ohmic. Ohmic contacts are necessary for the passage of current of either polarity through a device, whilst Schottky barrier contacts are useful as gates for biasing purposes.

A device that features both n-type and p-type regions presents a problem as far as this approach goes, because different metals must be used for each type of contact if they are to have Ohmic characteristics. This problem is made worse if a significant density of semiconductor surface states are involved, since their effect is to generate a depletion layer at the junction by 'piniggn' the Fermi surface in the middle of the band gap.

A way of avoiding this problem is to use very heavily doped contact regions, as these devices do. When a very heavily doped semiconductor is contacted with a metal, it may well form a Schottky contact. However, the heavy doping will result in the depletion layer being narrow enough that carriers can tunnel between the semiconductor and the metal to a useful degree.

To create such a heavily doped contact region, the semiconductor must be implanted with a high density of the relevant ions. This high density implantation causes damage to the crystalline structure of the semiconductor which spoils its electrical properties. In order to activate the contacts and restore the crystalline nature of the heavily-doped contact region, the wafer must be annealed. In the case of the wafers under current discussion, the implants were activated with 30 minutes at 530°C, in a nitrogen atmosphere. If the anneal is too long or hot dopants may significantly diffuse or the alloy layer may relax, but this step is insignificant compared to the native oxide growth, which required one hour at 720°C.

A contact created by implantation and annealing will not be truly Ohmic: the resistance of the contact will vary slightly with the voltage across the contact. However, this is not an issue if a 4-terminal method is used for measuring quantitative characteristics of the device provided that the resistance is never as high as, for example, the input impedance of the voltage measuring equipment (typically more than 10 MΩ). However, it is possible that even if a contact is (approximately) Ohmic at room temperature, at temperatures lower than around 100 K the contact will 'freeze o"t and be useless. The IV curve of such a contact at 10K is shown in Figure 7.15. (This was a two-terminal measurement, and only one of the contacts used was frozen.)

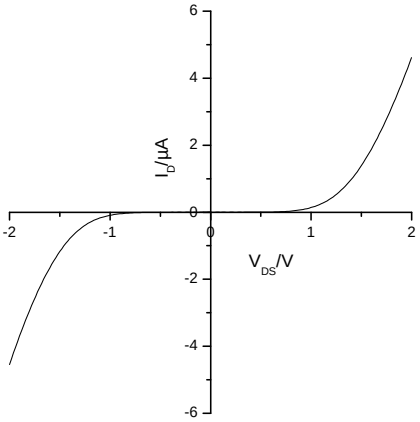

FIG. 7.15: IV curve at 10 K showing a deep, p^+ contact 'frozen out.' The resistance at 0 V V_{DS} is almost 100 MΩ, making the contact useless.

For a Hall bar to be measurable, both of the current contacts and at least three out of the four voltage probe contacts must remain Ohmic. Unfortunately, it was found that only a small fraction of the contacts on each wafer remained Ohmic at low temperature. This is why no Hall mobilities will be presented for devices on wafer 50/53, and why many of the other structures fabricated on the wafers were not characterized.

Hysteresis

Figure 7.16 demonstrates how, as temperature decreases, hysteresis develops in the transconductance characteristics of the p-channel. When the gate voltage is increased the drain current decreases, and as temperature drops the drain current at positive gate voltages clearly saturates at lower and lower values. However, when the gate voltage is decreased there appears a peak in drain current. This hysteresis effect is present no matter how quickly the gate voltage is swept, but the high mobility peak does not persist for more than one hour at 10 K if the gate voltage is held constant.

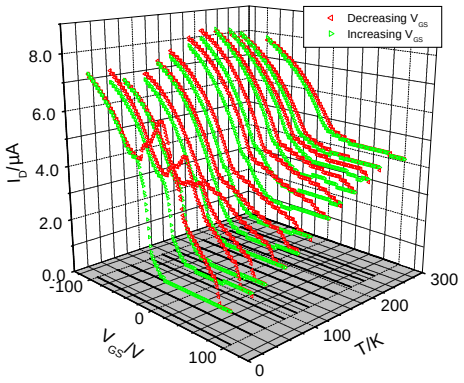

FIG. 7.16: Temperature dependence of transconductance of a device on wafer 50/51 as measured using deep p^+ contacts, showing the development of the hysteresis. The drain voltage is 100 mV.

Figure 7.17 shows transconductance characteristics of the n- and p-channels of a (different) device on wafer 50/51 at 10 K. Hysteresis is visible in the n-channel but is not as striking as in the p-channel. Also, the conductance of the n-channel is generally inferior to the p-channel. Hopefully, Hall effect measurements (section 7.5) will help begin to explain the hysteresis phenomenon.

Conduction Between Deep and Shallow Contacts

It would be fascinating to compare these results with similar results for a device with a much smaller separation between the electron and hole systems. However, no devices existed on wafer 50/53 which had a sufficient number of Ohmic contacts at 10 K, so systematic studies and Hall effect measurements (as will be presented in the following section) were out of the question.

However, Figure 7.18 shows how the vertical field within the device can induce a small current between the deep and shallow contacts: with the drain connected to the deep p+ contacts and the source connected to the shallow n^+ contacts, V_{DS} was held at 0 V whilst V_{GS}

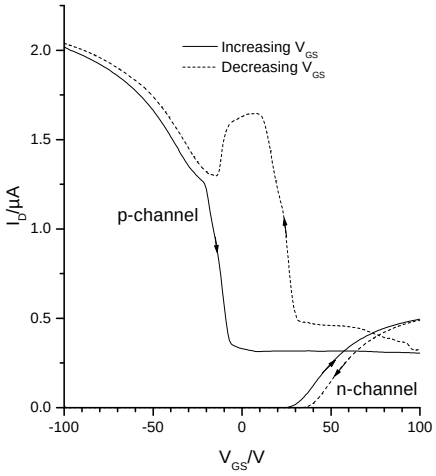

FIG. 7.17: The transconductance characteristics of the n- and p-channels of a device on wafer 50/51 at 10 K, showing hysteresis. The drain voltage is 100 mV, the n-channel threshold voltage is around 30 V, greater if the gate voltage is decreasing rather than increasing.

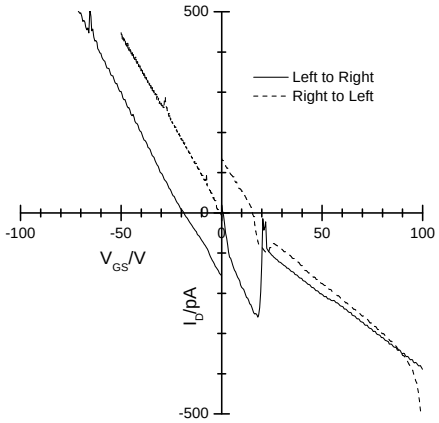

FIG. 7.18: With the drain connected to the deep p^+ contacts and the source connected to the shallow n^+ contacts, V_{DS} was held at 0 V whilst V_{GS} was swept. The bending of the bands in reponse to the gate potential induces a small current between the deep and shallow contacts. A feature can be seen at +20 V which corresponds to the formation of the electron inversion layer.

was swept. The bending of the bands in response to the gate potential induces a small current between the deep and shallow contacts. A feature can be seen at +20 V which corresponds to the formation of the electron inversion layer. A similar spike was seen at V_{GS}=40 V when measuring a device on wafer 50/51 in the same way. As the gate voltage is decreased from 100 V to 0 V, a dip can

be seen is this area. A small spike can be seen at around -15 V as V_{GS} is decreased further, which may correspond to the formation of a hole gas at the oxide interface. The data is discontinuous at 0 V due to hysteresis.

The IV characteristics of such a "tunnelling" arrangement (with current passing vertically through the device between deep and shallow contacts) are interesting but inconclusive. It is clear, though, that there is more to the behaviour than the interaction of the shallow contacts themselves and the hole gas in the alloy layer because there is a dependency on the gate bias, and therefore the electron density in the inversion layer. Such behaviour is probably worthy of detailed study at very low temperatures, in a device with a full set of Ohmic contacts.

7.5. Hall Effect Results

Data in this section was mainly obtained using a d.c. method employing the HP parameter analyzer (described in section 4.2.1) and a differential amplifier developed by R. J. P. Lander (although some results were previously obtained using an a.c. lock-in amplifier method). This meant that potentials between contacts of 0.1 μV could be resolved, increasing the sensitivity by a factor of 1000 over the HP alone.

The Hall effect shows that the system enters a high resistivity phase at a Hall sheet density (Equation 4.7) of $8\times10^{11}\mathrm{cm}^{-2}$. This is seen as the gate voltage increases past -10 V, and also as the gate voltage decreases past 70 V. The Hall effect calculation ceases to be applicable: it yields zero mobility and undefined sheet density indicating non-metallic, 'hopping' transport.[164,165] These points have been removed from Figure 7.19 for the sake of clarity. This could be interpreted as a metal-insulator transition; the longitudinal resistivity at this transition is roughly $\frac{h}{2e^2}$.

Figure 7.19 shows fairly conventional behaviour for gate voltages in the range -100 V to -10 V. Sheet density increases as the gate voltage is made more negative, and the mobility correspondingly decreases. (Mobility versus sheet density is shown in Figure 7.20). However, in the gate bias region between -10 V and +100 V the density and mobility are subject to the same hysteresis as the transconductance: in fact, there is a mobility peak at the (decreasing) V_{GS} of 25 V of 2,100 $\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ (corresponding to a sheet density of $1\times10^{12}\mathrm{cm}^{-2}$; a similar sheet density is seen at a V_{GS} of -25 V, but here the mobility is only 1,500 $\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$) which suggests that the conducting channels are of different natures. This suggests that single-carrier Hall calculations will be misleading. In fact, since the heterostructure was designed to produce a 2DHG with a sheet density of $3\times10^{11}\mathrm{cm}^{-2}$, and consideration of the discussion in Chapter 3 suggests that a sheet density as high as that measured in the alloy layer is physically impossible, either the Hall scattering factor is less than unity even at 10 K, additional conduction channels are present, or the structure itself deviates from specifications.

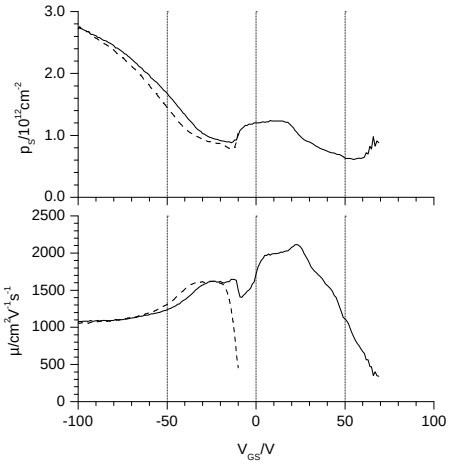

FIG. 7.19: Single carrier Hall calculations (from Chapter 4) applied to ρ_{XX} and ρ_{XY} data from the deep p^+ contacts of a Hall bar device on wafer 50/51 at 10 K. The solid line is data taken as the gate voltage decreases from +100 V to -100 V, the dotted line is data taken as the gate voltage increases. This should be compared with the low temperature data in Figure 7.16.

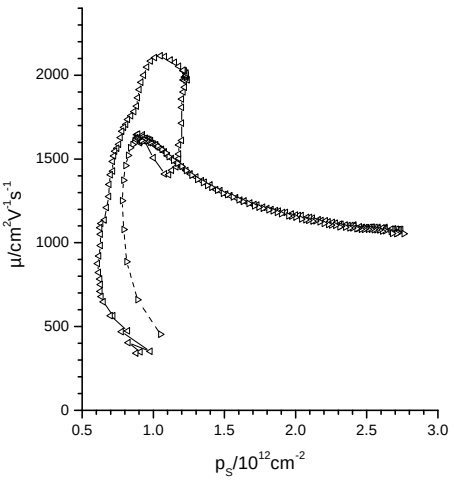

FIG. 7.20: The same data as in Figure 7.19 but plotted as (Hall) mobility against sheet density. The data representing decreasing gate voltage (leftward-pointing triangles, solid line) stand out as a high mobility peak at a sheet density just over $1\times10^{12}\mathrm{cm}^{-2}$.

As Chapters 4 and 6 make clear, this is the sort of problem that magnetoresistance and mobility spectrum analysis would be well-equipped to address. However, in this case there is the additional complication that the state of interest (induced by bringing the gate voltage down from 100 V to 25 V) decays significantly in the

time taken to acquire magnetoresistance data. A solution, involving a gate bias which is cycled between 100 V and 25 V for every measurement (of which there are typically 512 or 1024) would add considerably to the time and complexity of the experimental procedure.

Interpretations of Hall Effect Results

As Figure 7.16 shows, at large positive gate biases (particularly for the case where the gate voltage is increasing) the conductivity through the device drops with temperature. This, along with the disappearance of the Hall effect mentioned above, is consistent with the behaviour of conduction through boron-doped silicon in which transport is freezing out.[165]

At the negative bias limit, it assumed that hole gases exist both within the alloy and at the oxide interface, as is usual in gated heterostructures.[90] If free charge exists within the cap layer, the gate and the alloy layer are screened from each other.

Simulations of the band profile which solve the Poisson equation in one dimension (or even solve the Poisson and Schrödinger equations self-consistently) are not strictly applicable to this hysteresis effect at low temperatures, since it is evidence of a system out of equilibrium. Some results of simulation at higher temperatures will be presented in the following section.

Possible Causes of Hysteresis

Hysteresis effects are seen in MOS systems where traps exist in, or ions are moving through, the gate oxide in response to the gate bias.[10,166] Consider that mobile positive ions (for example, Na^+) exist in the oxide of a device with a certain threshold voltage for electrons. Upon application of a strong positive gate bias, these charges will be forced away from the gate, towards the oxide interface. If the positive charges remain at the interface when the bias is removed then effectively a fraction of the positive bias is stored at the interface. Therefore, a slightly more negative voltage is now required (compared to before the charge was moved by the strong bias) to counteract this stored charge. The threshold voltage for inversion is therefore moved to a more negative gate bias due to the application of a strong positive bias. This is the *opposite* of the observed behaviour in Figure 7.17, so cannot be the correct explanation.

The hysteresis in Figure 7.17 is also in the opposite sense to that seen in n-channel silicon-on-insulator MOS transistors.[167] Hysteresis has been observed in the I_D versus V_{DS} behaviour of silicon MOS transistors at 4.2 K[168,169] where it is discussed in terms of self-heating (not thought to be important and not relevant in this system where the hysteresis is in the I_D versus V_{GS} behaviour) and field-induced dopant ionization.[170]

The fact that the hysteresis effect grows in significance on the same temperature scale that transport in the boron doping slab freezes out (around 60 K, related to the energy of the acceptor level in boron) would suggest that ionization of the dopants induced by changes in bias conditions may be relevant. The transients seen in Figure 7.18 which show charge moving vertically through the device back this up.

7.6. Analysis

Carrier Gas Formation

Simulations were performed using the FISH1d software package, running on the public-access Purdue University Network Computing Hub, PUNCH. This software numerically solves the one-dimensional equilibrium Poisson equation in a semiconductor heterostructure at a given temperature, under specified bias conditions.[171,172]

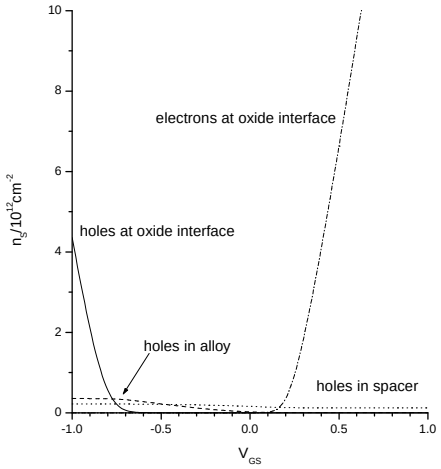

FIG. 7.21: Results of simulations of 50/51 which solve the one-dimensional Poisson equation. The temperature is 300 K. The x-axis scale is arbitrary, since it does not take into account the oxide thickness or interface charge. The two important features are that the electron and hole gases never coexist, and that the density of holes in the alloy reaches a relatively small value before the cap layer begins to populate.

Figure 7.21, featuring results from simulating the structure of wafer 50/51 at 300 K, shows that the electron gas at the oxide interface does not co-exist with a hole gas in the alloy of roughly equal density at any gate bias, and therefore that the band profile in Figure 7.1 is never realized in this heterostructure: the band profile is closer to the lower panel of Figure 7.3. The x-axis scale is arbitrary since the thickness of the oxide and the possibility of interface charge is ignored. (The positions of the electron and hole thresholds, which can be compared to Figure 7.7 and Figure 7.9 respectively, suggests that experimental zero bias corresponds to a simulation bias of around -0.3 V.)

The density of free holes in the dopant layer, not shown in Figure 7.21, varies from $4.7\times10^{12}\text{cm}^{-2}$ at the

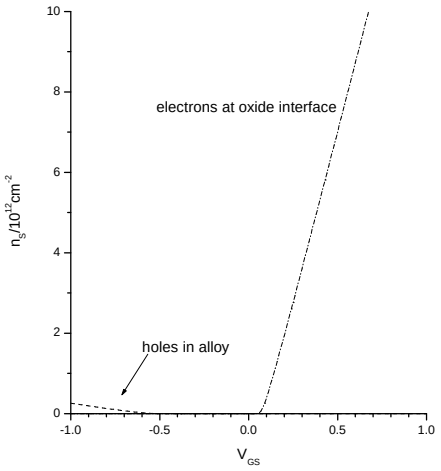

FIG. 7.22: Results of simulations of 50/51 similar to Figure 7.21 but with a temperature of 77K. Again, the electron and hole gases never coexist.

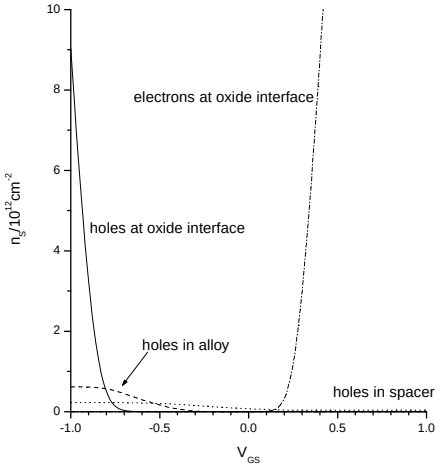

FIG. 7.23: Results of simulations of 50/53 which solve the one-dimensional Poisson equation. The temperature is 300 K. The alloy population reaches a slightly larger value than in Figure 7.21 before the cap layer begins to populate, but otherwise the important features are the same.

most negative bias to $4.6\times10^{12}\text{cm}^{-2}$ at the most positive. Since the Fermi level is pinned close to the valence band by the dopant, a small number of carriers are present in the region un-doped silicon spacer layer.

Figure 7.21 should be compared with Figure 7.9: drain current varies very little for gate voltages larger than 20 V, since current flow is almost entirely through the dopant layer. The slight increase in current flow as V_{GS} decreases from 20 V to -5 V corresponds to the region in Figure 7.21 where the population of holes in the alloy and spacer layer is increasing; the large increase in current flow from -5 V to -100 V corresponds to the formation of a hole gas at the oxide interface. This means that most of the variation in conductance between p-type contacts is due to the gate bias modulating the density of a hole gas formed at the oxide interface.

Figure 7.22 shows results of simulations of wafer 50/51 at 77 K: the gate bias does not go negative enough to induce a hole gas at the oxide interface, and the problem of the electron and holes gases not co-existing is worse than at room temperature.

Simulations of wafer 50/53 at 300 K are shown in Figure 7.23. The thin cap (30 nm rather than 90 nm as in 50/51 or 50/52) essentially increases the capacitance of the device (increases the number of holes in the cap layer) but does not change the fact that significant numbers of electrons and holes never co-exist.

Issues with the Original Design

The above results of simulation make it clear that Figure 7.1 is inaccurate; in fact, it is fundamentally unphysical. In one dimension, Poisson's equation can be written as

$$\frac{\mathrm{d}^2V}{\mathrm{d}z^2} = \frac{-en(z)}{\epsilon_r\epsilon_0} \tag{7.2}$$

where $n(z)$ represents the number density of free charges at a depth z within a semiconductor of relative permittivity ϵ_r.[10,157] Integration of Equation 7.2 with a constant n leads to:

$$V(z) = \frac{-en}{2\epsilon_r\epsilon_0}z^2 + Az + B \tag{7.3}$$

The constants A and B describe, respectively, the external electric field and the external potential. If the system is treated as closed, with only changes in potential through the device being relevant, then only the first term in Equation 7.3 is important.

The implications of Equation 7.3 are that the gradient of the bands (that is, the electric field) within a semiconductor changes where, and only where, free charges are present. Further, the gradient of the bands is proportional to the difference between the total amount of free charge to be found in either direction. It is assumed that the background (unintentional) impurity level within typical heterostructures is at most 10^{15}cm^{-3}, four orders of magnitude lower than the level of intentional doping,[21,153] so the scale on which ionization of unintentional dopants leads to significantly curved energy bands is at least 100 times longer than the scale on which the energy bands curve in regions of ionized intentional doping or the width of the wavefunction of a 2-dimensional carrier gas, which is the order of 10 nm.[21,22,157]

This band bending can be seen in Figure 3.1: the gradient of the bands changes only where ionized dopant or the carrier gas itself exists. In the setback region, the presence of ionized dopant atoms (negative free charges)

below and the carrier gas (positive free charges) above results in an electric field, and therefore a sloping energy band. Since the carrier gas and ionized dopant densities exactly cancel, there is no electric field above the alloy layer and so the bands are flat.

Thus Figure 7.1, reproduced exactly from Reference 153, is misleading since the slope of the energy bands implies that the electric field between the 2DEG and the 2DHG is exactly the same as the field between the 2DHG and the dopant layer. This is only possible if the 2DHG density is actually zero.

There are examples in the literature of hand-drawn band profiles used to justify and illustrate heterostructure designs[2,159,161] which clearly violate Poisson's equation. In some cases, it is possible that a proper consideration of the band profile would show the structure to be of little use for its intended purpose. Whilst a full Poisson-Schrdinger solution is overkill when a device is being sketched out within a proposal or review, it is hoped that consideration of the following discussions of device design will lead to a greater intuitive sense of the band profile in a semiconductor heterostructure.

Calculation of the Expected Sheet Density

The following is assumed to be valid for "zeo" temperature, whichin this context means that $T \ll T_F$ where T_F is given by Equation 3.5. The effective mass m^* is taken to be 0.3 m_e (based on the results in Chapter 5) giving a T_F of 30 K (from Equation 3.5) if the simple calculation given in Equation 7.1 is assumed to be valid.

Firstly it is assumed that the Fermi level is pinned in the bulk of the Boron doping slab, at a level of E_A ∼30 meV relative to the band edge, and that the width of the depletion region at the edge of the doped region is much less than the setback d, which is reasonable considering the high Boron concentration. It is then assumed that the confining potential in the alloy is triangular, with energy levels:[21]

$$E_n = \left(\frac{\hbar^2}{2m^*}\right)^{1/3} \left(\frac{3}{2}\pi q F_0\right)^{2/3} \left(n + \frac{3}{4}\right)^{2/3} \qquad (7.4)$$

$$\text{where} \quad F_0 = \frac{p_s q}{\epsilon_0 \epsilon_{Si}} \text{ and } n = 0, 1, 2 \ldots \qquad (7.5)$$

Only the $n = 0$ level is significantly occupied when $T \ll T_F$ so the picture in Figure 3.2, which shows the Fermi level at $\frac{\hbar^2 k_F^2}{2m^*}$ from the E_0 subband, holds. E_0 calculated with the sheet density value obtained from Equation 7.1 is 32 meV, and the Fermi energy E_F from Equation 3.5 is 2 meV. Equation 7.1 can now be modified to take these into account, assuming flat bands everywhere else in the structure and that the dopant depletion region is narrow:[21]

$$\frac{p_s q}{\epsilon_0 \epsilon_{Si}} = \frac{\Delta E_V - E_0 - E_A - E_F}{d} \qquad (7.6)$$

The sheet density consistent with this field in the setback is $1.4{\times}10^{11}\text{cm}^{-2}$, around half that of the value found from Equation 7.1. Further iteration around the system of Equations 3.5, 7.4, 7.5 and 7.6 leads to convergence on a sheet density of $1.7{\times}10^{11}\text{cm}^{-2}$. This analysis ignores background doping of the supposedly intrinsic layers of the structure and charged interface impurities.* Typical values for these parameters are of the order of magnitude of the calculated sheet density but essentially unpredictable, so their incorporation into Equation 7.5 (see Equation 3.1) would introduce a significant degree of arbitrariness.[21]

However, as made clear by Figures 7.21, 7.22 and 7.23, any application of positive gate bias in order to form an inversion layer of electrons will first deplete the hole gas in the alloy layer, leading to a band profile similar to Figure 7.3. The designed-in hole gas density must be much greater than that which is intended to be present when the electron gas is formed, if they are to co-exist.

7.7. Specifications For New Structures

7.7.1. Vertically Integrated MOSFET

It is generally true that for hetero-MOSFETs to operate usefully, the cap layer must be thin enough that the gate bias modulates the hole density in the alloy layer without forming an inversion layer at the oxide interface.[90] Figures 7.21 and 7.23 show that this is not the case when the cap layer is 90 nm or even 30 nm thick. In fact, cap thicknesses of less than 10 nm are under consideration for current HMOS research devices.[106]

Devices on 50/53 showed serious leakage between the shallow (25 nm) n^+ contacts and the alloy channel (at a depth of 30 nm) so in a thin-cap structure, the n^+ contacts must be very shallow indeed.

Another issue with these devices is that, at room temperature, there is always conduction between the deep p^+ contacts through the boron doping layer. This doping layer is important to ensure that a hole gas forms in the alloy layer under the application of negative gate bias, well before the oxide interface becomes significantly populated with holes, although this may be accomplished by much weaker doping than is used here. However, if the deep contact implantation is performed at such an energy that the alloy is contacted but the boron doping slab is not, then the depletion region in the setback should ensure that conduction is only possible through the hole gas in the alloy layer.

---

* It is possible that a setback of 34 nm was chosen to allow for segregation of boron dopant atoms. This would mean that the true setback would be much smaller. However, to achieve a carrier concentration in the alloy of 10^{12}cm^{-2}, the setback would need to be less than 5 nm.

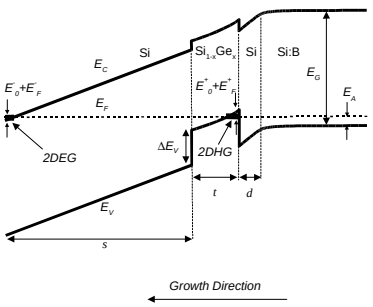

FIG. 7.24: Proposed (low-temperature) band profile of a device which will feature co-existing electron and hole 2-dimensional gases.

The calculation presented above can be adapted to consider co-existing electron and hole gases. A proposed band profile is shown in Figure 7.24 which can be compared to Figure 7.1. Assuming that a consistent solution exists, that background doping ionization and interface impurities are negligible, and that the dopant depletion width is small (since the dopant dose is very high) the electric field in the setback between the dopant layer and the alloy is:

$$F_d = \frac{(p_M - n_s + p_s + n_A^-)q}{2\epsilon_r\epsilon_0} = \frac{n_A^- q}{\epsilon_r\epsilon_0} \qquad (7.7)$$

where n_A^- is the sheet density of ionized dopant atoms, n_s is the electron carrier concentration in the 2DEG, p_s is the hole carrier concentration in the 2DHG and p_M is the concentration of positive charges on the metal gate. The electric field in the cap layer is:

$$F_s = \frac{(p_M - n_s - p_s + n_A^-)}{2\epsilon_r\epsilon_0} = \frac{(n_A^- - p_s)q}{\epsilon_r\epsilon_0} \qquad (7.8)$$

Equations 7.7 and 7.8 explain that the presence of these charges create sloping bands, and make use of overall charge neutrality: $p_M = n_s - p_s + n_A^-$. However, in order for the charges to be present, the fields must be such that the bands cross the Fermi level as shown in Figure 7.24 and this condition can be summarized such that

$$qF_d d = -E_A + \Delta E_V - E_F^+ - E_0^+ \qquad (7.9)$$

$$qF_s(s+t) = E_G - \Delta E_V + E_F^+ + E_0^+ + E_F^- + E_0^- \qquad (7.10)$$

where E_A is the acceptor level in the dopant layer, E_G is the silicon band gap, ΔE_V is the valence band offset, E_F^+ is the Fermi level and E_0^+ is the subband ground state energy for holes in the 2DHG (see Equations 3.5 and 7.4, and Figure 3.2) and E_F^- and E_0^- are the corresponding quantities for the 2DEG. The relative permittivity of germanium is higher than that of silicon, but in considering a thin layer of an alloy with a low germanium content this is ignored.

From charge neutrality it can be seen that if $n_s = p_s$ then $p_M = n_A^-$. Substitution of Equations 7.7 and 7.8 into 7.9 and 7.10 gives:

$$p_s = n_s = \frac{\epsilon_r\epsilon_0}{q^2}\left[\frac{1}{d}(-E_A + \Delta E_V - E_F^+ - E_0^+) - \frac{1}{s+t}(E_G - \Delta E_V + E_F^+ + E_0^+ + E_F^- + E_0^-)\right] \qquad (7.11)$$

If the two largest energies in Equation 7.11 are ΔE_V and E_G (for a 20% germanium alloy ΔE_V=140 meV whereas $E_A \sim E_0 \sim$30 meV and $E_F \sim$3 meV)[14] then

$$p_s = n_s \simeq \frac{\epsilon_r\epsilon_0}{q^2}\left[\frac{\Delta E_V}{d} - \frac{E_G - \Delta E_V}{s+t}\right] \qquad (7.12)$$

A setback d of 34 nm and a gas separation $s+t$ of 105 nm leads to an unphysical negative result for p_s and n_s. For a positive solution to Equation 7.12 to exist in a 20% germanium alloy system, the spacing between the gases must be more than 7 times the setback. If the setback were 10 nm, carrier gases may co-exist according to this approximation. However, once the zero-point energy of each triangular quantum well is taken into account, it becomes non-trivial to find a solution. A self-consistent solution may not exist at all unless very germanium-rich alloys (which are hard to grow pseudomorphically)[9] are used. For a fully-strained pure germanium channel on pure silicon, with a 10 nm setback and 100 nm between carrier gases, Equation 7.12 suggests coexisting gases, each with a density of more than 4×10^{12}cm^{-2}.

From the starting points described above, a true Poisson-Schrödinger solution method may be used to find the electron and hole gas sheet densities. However, it is not guaranteed that there will be a solution featuring matched and coupled-channels for a given carrier gas spacing or hetero-offset, even if Equation 7.12 is fulfilled; the device specifications themselves must be part of the iterative process. A method of genetic algorithms may be suited to finding a viable device structure.

It is also possible that more exotic structures (involv-

ing n and p-type modulation doping, a virtual substrate or a strained-silicon electron channel) would lead to co-existing electron and hole gases which are closer together and each have a high enough sheet density to be metallic.

8. CONCLUSIONS

Aspects of transport in both n- and p-type strained silicon-germanium heterostructures have been investigated, with a view to device applications as well as fundamentals of physics.

In a p-type device with a relatively high mobility, fundamentals of quantum transport have been investigated at millikelvin temperatures and it is shown that conventional theories do not account for the rate at which the quantum phase breaks. New theories are considered and the work is related to theories of the metal-insulator-transition in 2-dimensional systems. It is suggested that there at high sheet densities where the behaviour of the 2-dimensional hole gas is metallic the dephasing rate may saturate at a finite value as the temperature drops to zero. (Traditional theory predicts an infinite dephasing rate at zero temperature).

At a filling factor of $\sim$1.6, a transition from the QHE state into an insulating phase is seen. The resistivity of this phase increases dramatically as the hole gas is depopulated. This transition has been observed in silicon MOSFETs and p-type silicon-germanium, but not in gated heterostructures. Again, it is related to the metal-insulator transition in 2-dimensional systems.

A contrasting p-type device has also been investigated. Here, the work stresses the issues commensurate with such small-scale devices where stray voltages can quickly destroy a whole batch. The device features a pseudomorphic $Si_{0.5}Ge_{0.5}$ alloy channel: there is a possibility that the channel has relaxed to some degree since it was grown, and if this were the case then the implications for electronics incorporating such structures is serious. The possibilities of lateral relaxation, oxide degradation or contamination are also considered.

A room temperature Hall scattering factor of 0.68 ± 0.04 has been extracted by comparing the measured and calculated mobility at 282 K, based on parameters found by fitting the mobility at 25 K.

Traditional theories of the resistivity as a function of temperature in the $\sim$1 K regime are based on weak localization, interactions and screening. The temperature dependence of this device, however, is not satisfactorily explained in this manner, and a new functional form for the resistivity as a function of temperature has been proposed, but now needs theoretical justification.

A new method of mobility spectrum analysis has been applied to magnetoresistance data from an n-type device, at temperatures between 350 mK and 294 K. The results highlight both the strengths of mobility spectrum analysis and its shortcomings: in systems where transport takes place in both a quantum well and a dopant layer, two peaks are resolved. However, the peak shape (which should contain information regarding the energy dependence of the scattering time) owes more to the quality of the data than to scattering mechanisms.

It is also possible that the results in this particular device are invalidated at temperatures above 100 K, due to conduction through the substrate.

Other mobility spectrum analysis methods are briefly reviewed. The common "mirror" peak artefact is investigated and an explanation is suggested. It is believed that whilst the program code used here could be refined slightly, significant improvement to the analysis is unlikely. It may be possible to apply the maximum entropy inversion method to find the energy dependence of the scattering time from the mobility as a function of sheet density or temperature.

Lastly, devices where both n-type and p-type conduction can be induced are investigated in terms of the physics of coupled carrier gases, and the possibilities of vertical CMOS integration. It is shown that carrier gases do not co-exist in this structure, and it is shown that the design of a silicon-germanium heterostructures which may support co-existing electron and hole gases is non-trivial and not necessarily possible.

These ideas are important not only to the growth of the next generation of silicon-germanium devices, but also to the ways in which they are characterized.

---

* Electronic address: daniel@chrastina.net

[1] T. E. Whall and E. H. C. Parker, Thin Solid Films **367**, 250 (2000).

[2] D. J. Paul, Thin Solid Films **321**, 172 (1998).

[3] F. Schäffler, Semicond. Sci. Technol. **12**, 1515 (1997).

[4] E. Basaran, R. A. Kubiak, T. E. Whall, and E. H. C. Parker, Appl. Phys. Lett. **64**, 3470 (1994).

[5] G. Abstreiter, Physica Scripta **T68**, 61 (1996).

[6] K. Ismail, M. Arafa, K. L. Saenger, J. O. Chu, and B. S. Meyerson, Appl. Phys. Lett. **66**, 1077 (1995).

[7] K. Ismail, J. O. Chu, and B. S. Meyerson, Appl. Phys. Lett. **64**, 3124 (1994).

[8] P. W. Li and E. S. Yang, Appl. Phys. Lett. **63**, 2938 (1993).

[9] D. C. Houghton, J. Appl. Phys. **70**, 2136 (1991).

[10] S. M. Sze, *Semiconductor Devices: Physics and Technology* (Wiley, 1985).

[11] B. Laikhtman and R. A. Kiehl, Phys. Rev. B **47**, 10515 (1993).

[12] T. E. Whall, A. D. Plews, N. L. Mattey, P. J. Phillips, and U. Ekenberg, Appl. Phys. Lett. **66**, 2724 (1995).

[13] T. E. Whall, A. D. Plews, N. L. Mattey, and E. H. C. Parker, Appl. Phys. Lett. **65**, 14276 (1994).

[14] M. M. Rieger and P. Vogl, Phys. Rev. B **48**, 14276 (1993).

[15] M. A. Sadeghzadeh, C. P. Parry, P. J. Phillips, E. H. C. Parker, and T. E. Whall, Appl. Phys. Lett. **74**, 579 (1999).
[16] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[17] J. S. Blakemore, *Solid State Physics Second Edition* (Cambridge University Press, 1985).
[18] R. J. P. Lander, M. J. Kearney, A. I. Horrell, E. H. C. Parker, P. J. Phillips, and T. E. Whall, Semicond. Sci. Technol. **12**, 1064 (1997).
[19] A. D. Plews, N. L. Mattey, P. J. Phillips, E. H. C. Parker, and T. E. Whall, Semicond. Sci. Technol. **12**, 1231 (1997).
[20] M. J. Kearney and A. I. Horrell, Semicond. Sci. Technol. **14**, 211 (1999).
[21] C. J. Emeleus, T. E. Whall, D. W. Smith, R. A. Kubiak, E. H. C. Parker, and M. J. Kearney, J. Appl. Phys. **73**, 3852 (1993).
[22] T. Tezuka, T. Hatakeyama, S. Imai, N. Sugiyama, and A. Kurobe, Semicond. Sci. Technol. **13**, 1477 (1998).
[23] A. Yutani and Y. Shiraki, Semicond. Sci. Technol. **11**, 1009 (1996).
[24] M. J. Kearney and A. I. Horrell, Semicond. Sci. Technol. **13**, 174 (1998).
[25] T. J. Grasby, C. P. Parry, P. J. Phillips, B. M. M. McGregor, R. J. H. Morris, G. Braithwaite, T. E. Whall, E. H. C. Parker, and R. Hammond, Appl. Phys. Lett. **74**, 1848 (1999).
[26] J. Volger, Phys. Rev. **79**, 1023 (1950).
[27] R. A. Stradling and P. C. Klipstein, *Hall, Magnetoresistance and Infrared Conductivity Measurements, Growth and Characterisation of Semiconductors 165–185* (IOP Publishing, 1990, 1991).
[28] W. A. Beck and J. R. Anderson, J. Appl. Phys. **62**, 541 (1987).
[29] S. Kiatgamolchai, Ph.D. thesis, University of Warwick, U.K. (2000).
[30] J. H. Davies, *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, 1998).
[31] P. J. Mohr and B. N. Taylor, Rev. Mod. Phys. **72**, 351 (2000).
[32] D. K. Wilson and G. Feher, Phys. Rev. **124**, 1068 (1961).
[33] E. Glaser, J. M. Trombetta, T. A. Kennedy, S. M. Prokes, O. J. Glembocki, K. L. Wang, and C. H. Chern, Phys. Rev. Lett. **65**, 1247 (1990).
[34] G. Hendorfer and J. Schneider, Semicond. Sci. Technol. **6**, 595 (1991).
[35] P. L. Jeune, D. Robart, X. Marie, T. Amand, M. Brousseau, J. Barrau, V. Kalevich, and D. Rodichev, Semicond. Sci. Technol. **12**, 380 (1997).
[36] A. Malinowski and R. T. Harley, Phys. Rev. B **62**, 2051 (2000).
[37] F. F. Fang and P. J. Stiles, Phys. Rev. **174**, 823 (1968).
[38] S. J. Koester, K. Ismail, K. Y. Lee, and J. O. Chu, Appl. Phys. Lett. **70**, 2422 (1997).
[39] T. Englert, D. C. Tsui, A. C. Gossard, and C. Uihlein, Surf. Sci. **113**, 295 (1982).
[40] P. T. Coleridge, R. Stoner, and R. Fletcher, Phys. Rev. B **39**, 1120 (1989).
[41] A. D. Plews, Ph.D. thesis, University of Warwick, U.K. (1996).
[42] J. P. Harrang, R. J. Higgins, R. K. Goodall, P. R. Jay, M. Laviron, and P. Delescluse, Phys. Rev. B **32**, 8126 (1985).
[43] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[44] S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).
[45] B. L. Altshuler, D. Khmelńitzkii, A. I. Larkin, and P. A. Lee, Phys. Rev. B **22**, 5142 (1980).
[46] M. J. Kearney, Semicond. Sci. Technol. **7**, 804 (1992).
[47] B. L. Altshuler, A. G. Aronov, and P. A. Lee, Phys. Rev. Lett. **44**, 1288 (1980).
[48] P. W. Anderson, E. Abrahams, and T. V. Ramakrishnan, Phys. Rev. Lett. **43**, 718 (1979).
[49] M. S. Burdis and C. C. Dean, Phys. Rev. B **38**, 3269 (1988).
[50] S. M. Sze, ed., *VLSI Technology* (McGraw–Hill, 1985).
[51] J. W. Matthews and A. E. Blakeslee, Journal of Crystal Growth **27**, 118 (1974).
[52] J. W. Matthews and A. E. Blakeslee, Journal of Crystal Growth **29**, 273 (1975).
[53] J. W. Matthews and A. E. Blakeslee, Journal of Crystal Growth **32**, 265 (1974).
[54] A. Fischer, H. Kühne, and H. Richter, Phys. Rev. Lett. **73**, 2712 (1994).
[55] G. Braithwaite, N. L. Mattey, E. H. C. Parker, T. E. Whall, G. Bruthaler, and G. Bauer, J. Appl. Phys. **81**, 6853 (1997).
[56] T. Knight and A. Miller, *http://www.ai.mit.edu/people/tks/silicon-dioxide.html* (MIT, 1997).
[57] G. Dorda, Phys. Scr. **T45**, 297 (1992).
[58] W. Hansch, V. R. Rao, C. Fink, K. Kaesen, and I. Eisele, Thin Solid Films **321**, 206 (1998).
[59] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).
[60] B. L. Altshuler, A. G. Aronov, and D. E. Khernelnitskii, J. Phys. C. Solid State Phys. **15**, 7367 (1982).
[61] B. L. Altshuler and A. G. Aronov, Solid State Commun. **38**, 11 (1981).
[62] I. L. Aleiner, B. L. Altshuler, and M. E. Gerhenson, Waves in Random Media **9**, 201 (1999).
[63] P. Mohanty, E. M. Q. Jariwala, and R. A. Webb, Phys. Rev. Lett. **78**, 3366 (1997).
[64] J. Shi and X. C. Xie, Phys. Rev. B **63**, 045123 (2001).
[65] G. Braitwaite, Ph.D. thesis, University of Warwick, U.K. (1999).
[66] H. Fukuyama and E. Abrahams, Phys. Rev. B **27**, 5976 (1983).
[67] S. Chakravarty and A. Schmid, Physics Reports **140**, 193 (1986).
[68] P. Mohanty and R. A. Webb, Phys. Rev. B **55**, R13452 (1997).
[69] D. S. Golubev and A. D. Zaikin, Phys. Rev. Lett. **81**, 1074 (1998).
[70] I. L. Aleiner, B. L. Altshuler, and M. E. Gerhenson, Phys. Rev. Lett. **82**, 3190 (1999).
[71] D. S. Golubev and A. D. Zaikin, Phys. Rev. Lett. **82**, 3191 (1999).
[72] B. L. Altshuler, M. E. Gershenson, and I. L. Aleiner, Physica E **3**, 58 (1998).
[73] D. S. Golubev and A. D. Zaikin, Phys. Rev. B **62**, 14061 (2000).
[74] Z. Wilamowski, N. Sandersfield, W. Jansch, D. Többen, and F. Schäffler, Phys. Rev. Lett. **87**, 026401 (2001).
[75] S. I. Dorozhkin, C. J. Emeleus, T. E. Whall, and G. Landwehr, Phys. Rev. B **52**, R11638 (1995).
[76] S. I. Dorozhkin, C. J. Emeleus, O. A. Mironov, T. E. Whall, and G. Landwehr, Surf. Sci. **361/362**, 933 (1996).

[77] P. T. Coleridge, A. S. Sachrajda, P. Zawadzki, R. L. Williams, and H. Lafontaine, Solid State Commun. **102**, 755 (1997).
[78] P. T. Coleridge and P. Zawadzki, Solid State Commun. **112**, 241 (1999).
[79] P. T. Coleridge, P. Zawadzki, A. S. Sachrajda, Y. Feng, and R. L. Williams, cond-mat/9909292 (1997).
[80] R. B. Dunford, R. Newbury, V. A. Stadnik, F. F. Fang, R. G. Clark, R. H. McKenzie, R. P. Starrett, E. E. Mitchell, P. J. Wang, J. O. Chu, et al., Surf. Sci. **361/362**, 550 (1996).
[81] S. Kivelson, D.-H. Lee, and S.-C. Zhang, Phys. Rev. B **46**, 2223 (1992).
[82] V. J. Goldman, M. Shayegan, and D. C. Tsui, Phys. Rev. Lett. **61**, 881 (1988).
[83] V. J. Goldman, J. K. Wang, B. Su, and M. Shayegan, Phys. Rev. Lett. **70**, 647 (1993).
[84] T. Sajoto, Y. P. Li, L. W. Engel, D. C. Tsui, and M. Shayegan, Phys. Rev. Lett. **70**, 2321 (1993).
[85] H. W. Jiang, C. E. Johnson, K. L. Wang, and S. T. Hannahs, Phys. Rev. Lett. **71**, 1439 (1993).
[86] A. A. Shaskin, V. T. Dolgopolov, and G. V. Kravchenko, Phys. Rev. B **49**, 14486 (1994).
[87] S. V. Kravchenko, W. E. Mason, G. E. Bowker, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **51**, 7038 (1995).
[88] R. M. Feenstra and M. A. Lutz, J. Appl. Phys. **78**, 6091 (1995).
[89] O. A. Mironov, I. G. Gerleman, P. J. Phillips, E. H. C. Parker, M. Tsaousidou, P. N. Butcher, and T. E. Whall, Thin Solid Films **294**, 182 (1997).
[90] R. J. P. Lander, C. J. Emeleus, B. M. McGregor, E. H. C. Parker, T. E. Whall, A. G. R. Evans, and G. P. Kennedy, J. Appl. Phys. **82**, 5210 (1997).
[91] B. M. M. McGregor, R. J. P. Lander, P. J. Phillips, E. H. C. Parker, and T. E. Whall, Appl. Phys. Lett. **74**, 1245 (1999).
[92] R. J. P. Lander, Y. V. Ponornarev, J. G. M. van Berkum, W. B. de Boer, R. Loo, and M. Caymax, J. Appl. Phys. **88**, 2016 (2000).
[93] T. Ueno, T. Irisawa, and Y. Shiraki, Physica E **7**, 790 (2000).
[94] S. D. Sarma, Phys. Rev. B **33**, 5401 (1986).
[95] A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986).
[96] G. J. Dolan and D. D. Osheroff, Phys. Rev. Lett. **43**, 721 (1979).
[97] P. T. Coleridge, P. Zawadzki, A. S. Sachrajda, R. L. Williams, and Y. Feng, Physica E **6**, 268 (2000).
[98] P. T. Coleridge, R. L. Williams, Y. Feng, and P. Zawadzki, Phys. Rev. B **56**, R12764 (1997).
[99] R. Berkovits, J. W. Kantelhardt, Y. Avishai, S. Havlin, and A. Bunde, Phys. Rev. B **63**, 085102 (2001).
[100] A. R. Hamilton, M. Y. Simmons, M. Pepper, E. H. Linfield, P. D. Rose, and D. A. Ritchie, Phys. Rev. Lett. **82**, 1542 (1999).
[101] M. Y. Simmons, A. R. Hamilton, M. Pepper, E. H. Linfield, P. D. Rose, and D. A. Ritchie, Phys. Rev. Lett. **84**, 2489 (2000).
[102] A. Sahnoune, J. O. Ström-Olsen, and H. E. Fischer, Phys. Rev. B **46**, 10035 (1992).
[103] P. Mohanty, Physica B **280**, 446 (2000).
[104] H. E. Fischer, Private Communication (1998).
[105] P. J. Briggs, A. B. Walker, and D. C. Herbert, Semicond. Sci. Technol. **13**, 680 (1998).
[106] M. J. Palmer, G. Braithwaite, T. J. Grasby, P. J. Phillips, M. J. Prest, E. H. C. Parker, T. E. Whall, C. P. Parry, A. M. Waite, A. G. R. Evans, et al., Appl. Phys. Lett. **78**, 1424 (2001).
[107] M. Nafria, J. Suñe, D. Yelamos, and X. Aymerich, IEEE Trans. Electron Devices **43**, 2215 (1996).
[108] P. M. Lenahan, J. F. C. Jr, and B. D. Wallace, J. Appl. Phys. **81**, 6822 (1997).
[109] S. Lombardo, A. L. Magna, C. Gerardi, M. Alessandri, and F. Crupi, Appl. Phys. Lett. **75**, 1161 (1999).
[110] A. Zaslavsky, K. R. Milkove, Y. H. Lee, B. Ferland, and T. O. Sedgewick, Appl. Phys. Lett. **67**, 3921 (1995).
[111] Z. Dziuba, T. Przeslawski, K. Dybko, M. Górska, J. Marczewski, and K. Regiński, J. Appl. Phys. **85**, 6619 (1999).
[112] S. A. Studenikin, A. V. Chaplik, I. A. Panaev, G. Salis, K. Ensslin, K. Maranowski, and A. C. Gossard, Semicond. Sci. Technol. **14**, 606 (1999).
[113] J. S. Kim, D. G. Seiler, and W. F. Tseng, J. Appl. Phys. **73**, 8324 (1993).
[114] K. Regiński, J. Marczewski, Z. Dziuba, and E. Grodzicka, J. Appl. Phys. **82**, 6102 (1997).
[115] J. R. Meyer, C. A. Hoffman, J. Antoszewski, and L. Faraone, J. Appl. Phys. **81**, 709 (1997).
[116] I. Vurgaftman, J. R. Meyer, C. A. Hoffman, D. Redfern, J. Antoszewski, L. Faraone, and J. R. Lindemuth, J. Appl. Phys. **84**, 4966 (1998).
[117] D. S. Sivia, Los Alamos Science **19**, 180 (1990).
[118] M. Jarrell and J. E. Gubernatis, Physics Reports **269**, 133 (1996).
[119] J. P. Hague, Ph.D. thesis, University of Warwick, U.K. (2001).
[120] E. T. Jaynes, Phys. Rev. **104**, 620 (1957).
[121] J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, Phys. Rev. B **44**, 6011 (1991).
[122] E. Gallicchio and B. J. Berne, J. Chem. Phys. **105**, 7064 (1996).
[123] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, 1992).
[124] M. Levinshtein, S. Rumyantsev, and M. Shur, *Semiconductor Parameters, Volume 1* (World Scientific, 1996).
[125] J. Achard, C. Varenne-Guillot, F. Barbarin, and M. Dugay, Appl. Surf. Sci. **158**, 345 (2000).
[126] J. S. Kim, J. Appl. Phys. **86**, 3187 (1999).
[127] S. Madhavi and V. Venkataraman, Thin Solid Films **369**, 333 (2000).
[128] E. H. Putley, *The Hall Effect and Related Phenomena* (Butterworth & Co., 1960).
[129] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, Appl. Phys. Lett. **51**, 1934 (1987).
[130] D. J. Paul, A. Ahmed, M. Pepper, A. C. Churchill, D. J. Robbins, D. J. Wallis, and A. J. Pidduck, J. Vac. Sci. Technol. B **16**, 1644 (1998).
[131] G. Stöger, G. Brunthaler, G. Bauer, K. Ismail, B. S. Meyerson, J. Lutz, and F. Kuchar, Semicond. Sci. Technol. **9**, 765 (1994).
[132] S. Agan, O. A. Mironov, E. H. C. Parker, T. E. Whall, C. P. Parry, V. Y. Kashirin, Y. F. Komnik, V. B. Krasovitsky, and C. J. Emeleus, Phys. Rev. B **63**, 075402 (2001).
[133] M. J. Kearney, A. I. Horrell, and V. M. Dwyer, Semicond. Sci. Technol. **15**, 24 (2000).
[134] F. Stern and S. E. Laux, Appl. Phys. Lett. **61**, 1110 (1992).

[135] J. J. Harris, K. J. Lee, T. Wang, S. Sakai, Z. B. I. Moerman, E. J. Thrush, J. B. Webb, H. Tang, T. Martin, D. K. Maude, et al., Semicond. Sci. Technol. **16**, 402 (2001).
[136] L. B. Rigal, D. K. Maude, M. Potemski, J. C. Portal, L. Eaves, Z. R. Wasilewski, G. Hill, and M. A. Pate, Phys. Rev. Lett. **82**, 1249 (1999).
[137] J. J. Harris, K. J. Lee, D. K. Maude, J.-C. Portal, T. Wang, and S. Sakai, J. Phys. Cond. Matt. **13**, L175 (2001).
[138] G. Nachtwei, Physica E **4**, 79 (1996).
[139] S. Komiyama and Y. Kawaguchi, Phys. Rev. B **61**, 2014 (2000).
[140] W. Desrat, D. K. Maude, L. B. Rigal, M. Poemski, J. C. Portal, L. Eaves, M. Henini, Z. R. Wasilewski, A. Toropov, G. Hill, et al., Phys. Rev. B **62**, 12990 (2000).
[141] H. Iizuka, S. Kawaji, and T. Okamoto, Physica E **6**, 132 (2000).
[142] L. Eaves, Physica B **298**, 1 (2001).
[143] Y. Morita and Y. Hatsugai, Physica B **298**, 24 (2001).
[144] K. Takashina, R. J. Nicholas, B. Kardynal, N. J. Mason, D. K. Maude, and J. C. Portal, Physica B **298**, 8 (2001).
[145] G. Baccarani and P. Ostoja, Solid State Eletron. **18**, 579 (1975).
[146] Y. E. Lozovik and V. I. Yudson, Solid State Commun. **19**, 391 (1976).
[147] H. C. Tso, P. Vasilopoulos, and F. M. Peeters, Phys. Rev. Lett. **68**, 2516 (1992).
[148] X. Zhu, P. B. Littlewood, M. S. Hybertsen, and T. M. Rice, Phys. Rev. Lett. **74**, 1633 (1995).
[149] P. B. Littlewood and X. Zhu, Phys. Scr. **T68**, 56 (1996).
[150] J. S. Thakur, D. Neilson, and M. P. Das, Phys. Rev. B **57**, 1801 (1998).
[151] T. Fukuzawa, E. E. Mendez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990).
[152] J. A. Kash, M. Zachau, E. E. Mendez, J. M. Hong, and T. Fukuzawa, Phys. Rev. Lett. **66**, 2247 (1991).
[153] T. E. Whall, *Novel Transport Phenomena in Advanced Silicon Structures* (EPSRC Final Report, 1996).
[154] U. Sivan, P. M. Soloman, and H. Shtrikman, Phys. Rev. Lett. **68**, 1196 (1992).
[155] S. Shapira, E. H. Linfield, and M. Pepper, Appl. Phys. Lett. **74**, 1603 (1999).
[156] L. J. Cooper, N. K. Patel, V. Drouot, E. H. Linfield, D. A. Ritchie, and M. Pepper, Phys. Rev. B **57**, 11915 (1998).
[157] G. A. C. Jones, *Microelectronics and VLSI* (Lecture Course Notes, Cambridge University, 1997).
[158] R. M. Sidek, U. N. Straube, A. M. Waite, A. G. R. Evans, C. Parry, P. Phillips, T. E. Whall, and E. H. C. Parker, Semicond. Sci. Technol. **15**, 135 (2000).
[159] T. E. Whall and E. H. C. Parker, J. Mat. Sci.: Material in Electronics **6**, 249 (1995).
[160] T. E. Whall and E. H. C. Parker, J. Phys. D **31**, 1397 (1998).
[161] K. L. Wang, S. G. Thomas, and M. O. Tanner, J. Mat. Sci.: Material in Electronics **6**, 311 (1995).
[162] J. F. Zeigler, J. P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids* (edited by J. F. Zeigler, Pergamon, 1985).
[163] M. A. Sadeghzadeh, *Private communication* (1998).
[164] C. J. Emeleus, M. A. Sadeghzadeh, P. J. Phillips, E. H. C. Parker, T. E. Whall, M. Pepper, and A. G. Evans, Appl. Phys. Lett. **70**, 1870 (1997).
[165] J. G. Massey and M. Lee, Phys. Rev. Lett. **79**, 3986 (1997).
[166] Y.-C. King, IEEE Trans. Electron Devices **31**, 446 (2001).
[167] P. S. Liu and G. P. Li, Appl. Phys. Lett. **60**, 1845 (1992).
[168] E. Simoen and C. C, J. Appl. Phys. **73**, 3068 (1993).
[169] E. Simoen and C. C, J. Appl. Phys. **73**, 3074 (1993).
[170] E. Rosencher, V. Mosser, and G. Vincent, Phys. Rev. B **29**, 1135 (1984).
[171] J. L. Gray and M. S. Lundstrom, IEEE Trans. Electron Devices **ED-32**, 2102 (1985).
[172] M. S. Lundstrom, M. A. Stettler, T. T. Herman, and P. E. Dodd, *Fish1d 2.3 User's Manual* (Purdue University, http://punch.ecn.purdue.edu/, 1995).