National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre reference

Our file Notre reference

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

REENTRANT METAL-INSULATOR TRANSITIONS IN SI-MOSFETS

By

John W. M. Campbell

A thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Department of Physics

University of Ottawa
Ottawa, Ontario
April 25, 1995

ISBN 0-612-07841-8

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Abstract

This thesis describes a study of reentrant metal-insulator transitions observed in the inversion layer of extremely high mobility Si-MOSFETs. Magneto-transport measurements were carried out in the temperature range 20 mK - 4.2 K in a ^{3}He/^{4}He dilution refrigerator which was surrounded by a 15 Tesla superconducting magnet. Below a melting temperature ($T_M \sim 500$ mK) and a critical electron density ($n_s \sim 9 \times 10^{10}$ cm^{-2}), the Shubnikov-de Haas oscillations in the diagonal resistivity displayed enormous maximum values at the half filled Landau levels while maintaining deep minima corresponding to the quantum Hall effect at filled Landau levels. At even lower electron densities the insulating regions began to spread and eventually a metal-insulator transition could be induced at zero magnetic field.

The measurement of extremely large resistances in the milliKelvin temperature range required the use of very low currents (typically in the 10^{-12} A range) and in certain measurements minimizing the noise was also a consideration. The improvements achieved in these areas through the use of shielding, optical decouplers and battery operated instruments are described.

The transport signatures of the insulating state are considered in terms of two basic mechanisms: single particle localization with transport by variable range hopping and the formation of a collective state such as a pinned Wigner crystal or electron solid with transport through the motion of bound dislocation pairs. The experimental data is best described by the latter model. Thus the two dimensional electron system in these high mobility Si-MOSFETs provides the first and only experimental demonstration to date of the formation of an electron solid at zero and low magnetic fields in the quantum limit where the Coulomb interaction energy dominates over the zero point oscillation energy.

The role of disorder in favouring either single particle localization or the formation of a Wigner crystal is explored by considering a variety of samples with a wide range of mobilities and by varying the ratio of the carrier density (controlled by the applied gate voltage) to the impurity density (fixed during sample growth). A phase diagram showing the boundaries between the two dimensional electron gas, the Wigner solid, and the single particle localization induced insulator is established in terms of carrier density and sample mobility.

Statement of Originality

The work carried out during the tenure of the author's Ph.D. research project is, to the best of his knowledge, original and has focused on transport measurements using the highest mobility Si-MOSFETs available anywhere. The specific work undertaken has included the following:

Developing an integrated computer control system to combine different types of battery and line operated instruments in various experimental configurations related to electronic transport measurements

Working on noise reduction in electronic transport measurements on very highly resistive samples at milliKelvin temperatures which have allowed the first ever narrow band noise measurements to be made in the quantum Wigner regime

Characterizing metal-insulator transitions observed in high mobility Si-MOSFETs using electronic transport techniques over a range of temperatures, magnetic fields, and electron densities

Analysing these experimental results in terms of different theoretical models and concluding that these measurements represent the first experimental evidence for the existence of a Wigner crystal in Si-MOSFETs

The unique scientific significance of this work is that it involves a detailed study of a new phenomenon (ie. the reentrant insulating states observed in Si-MOSFETs at low electron densities and temperatures) and that it provides the first evidence for the existence of a quantum Wigner solid in any system.

Publications in Refereed Journals

J.W. Campbell, M. D'Iorio and V.M. Pudalov, *Activated Transport in the 2D Electron Solid in SI*, Physica B **194-196**, 1241 (1994)

V.M. Pudalov, M. D'Iorio and J.W. Campbell, *Termination of the Integer QHE by an Electron Solid in a Dilute 2D Electron system at mK-Temperatures*, Physica B **194-196**, 1287 (1994)

V.M. Pudalov, M. D'Iorio and J.W. Campbell, *Electron Solid at Zero Magnetic Field in a Dilute 2D Electron System in Si*, Physica B **194-196**, 1289 (1994)

V.M. Pudalov, M. D'Iorio and J.W. Campbell, *Nonlinear Transport and Generated Narrowband Noise in the 2D Electron Solid in Si*, Physica B **194-196**, 1291 (1994)

V.M. Pudalov, M. D'Iorio and J.W. Campbell, *Termination of the Integer Quantum Hall Effect by an Electron Solid*, Surf. Sci. **305**, 107 (1994)

M. D'Iorio, V.M. Pudalov, S.V. Kravchenko, and J.W. Campbell, *Collective Insulating State at Zero Magnetic Field in a Dilute 2D electron system*, Surf. Sci. **305**, 115 (1994)

V.M. Pudalov, M. D'Iorio, and J.W. Campbell, *Hall resistance and quantized Hall effect to insulator transitions in a 2D electron system*, JETP Lett. **57**, 608 (1993)

V.M. Pudalov, M. D'Iorio, S.V. Kravchenko, and J.W. Campbell, *Zero-Magnetic-Field Collective Insulator Phase in a Dilute 2D Electron System*, Phys. Rev. Lett. **70**, 1866 (1993)

M. D'Iorio, J.W. Campbell, V.M. Pudalov, and S.G. Semenchinsky, *Transport Properties of a 2D electron solid in Si-MOSFETs*, Surf. Sci. **263**, 49, (1992)

J.W.M. Campbell, F. Guillon, M. D'Iorio, M. Buchanan and R.J. Stoner, *Characterization of the Surface Acoustic Wave Induced Potential in a GaAs/AlGaAs Heterostructure*, Solid State Comm. **84**, 735 (1992)

S.V. Kravchenko, V.M. Pudalov, J. Campbell, and M. D'Iorio, *Electron solid at the Si Surface in Zero Magnetic Field*, JETP Lett. **54**, 528 (1991)

Acknowledgements

I would like to thank all those who have helped me in this long endeavour of producing a thesis, both at the University of Ottawa and at the National Research Council of Canada. Particular gratitude is due to Marie D'Iorio, my supervisor at NRC, for allowing me to use the facilities of her lab, for all that she taught me about dilution refrigerators, magnets and electronic transport measurements and for her continual editing of my almost unending stream of rewrites. Accolades are also due to Francis Guillon, my supervisor at University of Ottawa, for letting me get my hands dirty with his cryostat and for his critical comments on my thesis.

To Vladimir Pudalov, I owe a great debt of gratitude, since he provided the samples which were used in all the experiments and also taught me the secrets of reducing noise in the measurements and of how to work with delicate Si-MOSFETs without destroying them.

I must also thank Don Brown for his tireless technical work in keeping the lab at NRC running smoothly and for his enjoyable company in the early mornings. His willingness to listen to my stories about Little League baseball is deeply appreciated.

I also wish to thank all those other professors, graduate students, researchers and summer students who have formed part of the two communities that I have interacted with during the past four years at the university and at NRC.

A special acknowledgement is due to the National Sciences and Engineering Research Council and to the University of Ottawa for their financial support.

Finally, I would like to thank my parents, who made it possible for me to get to university in the first place, and my wonderful wife, who has put up with me during this final difficult period of writing a thesis.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

In 1934 Eugene Wigner calculated the correlation energy due to the mutual repulsion between free electrons in an electron gas [Wig34]. He noted that this energy is minimized if the electrons are organized in a close packed lattice and theorized that under certain conditions the Coulomb correlation energy could dominate the electron energetics and force the electrons to form such an ordered lattice. Later it was realized that the same phenomenon could also occur in a two dimensional electron system (2DES), such as the inversion layer of a silicon based metal-oxide-semiconductor field effect transistor (Si-MOSFET). In time, this electron lattice came to be known as a Wigner crystal or electron solid. Since then numerous detailed theoretical predictions have been made about the Wigner crystal. Its phonon spectrum, transport characteristics, and the range of temperatures and electron concentrations over which it can exist have all been determined using a variety of different assumptions about the underlying physics. Not until 1979, however, when C. C. Grimes and G. Adams discovered a crystalline structure in electrons suspended above the surface of liquid ^{4}He [Gri79], was there any direct experimental evidence for the existence of a two dimensional Wigner crystal, and convincing evidence of a Wigner solid in a semiconductor based 2DES was not detected until 1988 [And88].

This thesis discusses experiments which were carried out on extremely high quality Si-MOSFETs and which, for the first time in the "quantum regime", show a series of reentrant metal-insulator transitions with many of the characteristics of a Wigner crystal. Although reentrant metal-insulator transitions have been previously observed in GaAs/AlGaAs heterostructures near filling factor $\nu = 1/5$ and attributed to the freezing and melting of a Wigner crystal with changing magnetic field [Wil88, Gol90, Jia90b], these experiments took place in the "extreme quantum regime", where the cyclotron energy dominates the zero point motion energy of the electrons. This work also describes a metal-insulator transition at zero magnetic field which also shows the characteristic signatures of a Wigner crystal. This is the first experimental system to show evidence for a zero field electron solid in the quantum limit as was originally proposed by Wigner.

The Si-MOSFETs used in this work have remarkably high peak electronic mobilities (up to 75000 cm^2/Vs) and were fabricated in Russia in the mid

1980's. The magneto-transport measurements were performed in a $^3He/^4He$ dilution refrigerator (base temperature $\approx 20\,mK$) with a 15 Tesla superconducting magnet. Some of the results have been reproduced in low temperature laboratories at Nijmegen, Oklahoma, Oxford and Chernogolovka.

The material presented in this thesis is in six chapters. The basic physics of two dimensional electron systems in the quantum limit is summarized in Chapter 2. Localization mechanisms are first discussed for non-interacting particles, followed by a consideration of collective phenomena such as the charge density wave and the pinned Wigner crystal. The discussion of Wigner crystallization reviews the traditionally accepted predictions for the ideal infinite system at zero temperature and with zero disorder. Special attention is then given to the predictions of recent models which make more realistic assumptions about the samples in which the Wigner crystal is expected to form.

The Si-MOSFET samples used in this work were of remarkably high quality and their high mobility was crucial to the experiments. Chapter 3 gives explicit information on how these MOSFETs were fabricated, and discusses the experimental procedures used to characterize them. The experimental techniques required to study the strongly insulating state at low temperatures demanded that very low currents be used. Noise could also be amplified by the high sample resistance, and in the dilute electron regime contact resistances became appreciable. Thus, Chapter 3 gives details of some of the steps taken in the design of the experiments to have small, stable currents, to reduce the noise in the measurements and to demonstrate that contact effects and Joule heating were not significant factors in the experimental results.

Chapter 4 is the first chapter of experimental measurements and analysis. Since the Wigner crystal was initially observed as a series of metal-insulator transitions which were reentrant with magnetic field, this chapter focuses on the data obtained when the samples were subject to a perpendicular magnetic field. In addition to examining the raw data, several derived quantities such as a thermal activation energy and a critical melting density are calculated by considering the temperature and electron density dependences of the measurements. The results are initially analyzed in terms of various single particle localization models, but none of these is compatible with all the data. The model of a collective insulator such as a pinned charge density wave or Wigner crystal is then considered and found to be capable of explaining all of the results.

In Chapter 5, the results obtained at zero magnetic field are presented. Much of the data is similar to that in Chapter 4, but some unique measurements are presented. The broad band ($1-300\,kHz$) voltage noise background

and the detailed voltage noise spectrum in the 4−12 kHz regime are examined to look for evidence of a moving periodic potential. A more detailed analysis of the length scales appropriate for the measurements is also undertaken. This length scale analysis provides further evidence that the insulating state is not caused by single particle localization. Finally, a more specific analysis of the data is undertaken with respect to the predictions of two specific collective models; that of a pinned charge density wave with charge transport through phase slip and that of a pinned Wigner crystal with charge transport through the motion of bound dislocation pairs. These two models are known to be compatible, and both of them make predictions which are borne out by the data. The zero field insulator is best described as a pinned Wigner crystal.

Chapter 6 contain a brief set of conclusions and some conjecture on the direction of possible future research. The most important conclusion is that a two dimensional Wigner crystal forms in the inversion layer of the high mobility Si-MOSFETs studied in this thesis when the electron density and temperature are sufficiently low, even in the zero magnetic field "quantum regime" where no Wigner crystal has previously been observed.

Chapter 2. Theoretical Background

2.1 Two-Dimensional Electron Systems

2.1.1 Introducing Si-MOSFETs

The subject of this thesis is the study of the electron solid or silicon metal-oxide-semiconductor field effect transistors (Si-MOSFETs), which are one example of a quasi two dimensional electron system (2DES). There are actually two other physical systems which are commonly used to generate a quasi two dimensional electron (or hole) system. One is a thin layer of superfluid ^{4}He in a vacuum, where the electrons are restricted to a narrow region just above the surface of the liquid helium. The other is a semiconductor heterostructure where the electrons (or holes) are found at the interface between two semiconductors with different band gaps such as silicon and silicon germanium ($Si_{1-x}Ge_x$) or gallium arsenide (GaAs) and aluminum gallium arsenide ($Al_xGa_{1-x}As$). Evidence for the existence of the two-dimensional Wigner crystal has now been found in all three of these systems, but this thesis will focus on Si-MOSFETs.

Figure 2.1 shows a cross-section of a typical Si-MOSFET. The base is a semiconducting silicon wafer which has been lightly p-doped by diffusion of electron acceptor impurities such as boron. Then an insulating layer of silicon dioxide is grown on top of the semiconductor. In those regions where electrical contact with the quasi-two dimensional electron system is required, the oxide is removed and a donor impurity such as phosphorous is diffused into the semiconductor to create a small n-type contact region. The final part of the structure is a layer of metal which is evaporated onto the oxide layer to form the gate and on top of the n-doped contacts to create contact pads which simplify the task of connecting external wires to the sample. Generally the plane of the Si/SiO_2 interface is labelled the x-y plane and the z-axis is the one which passes through the gate, the oxide and then the silicon itself (i.e. from top to bottom in the figure).

In Figure 2.1 a positive potential known as the gate voltage, V_g, is applied between the metal gate and one of the contact pads. Initially this potential is zero and the p-doped silicon is a very poor conductor because the band gap between the valence and conduction band edges at $T = 0\,\mathrm{K}$ is $1.17\,\mathrm{eV}$, which

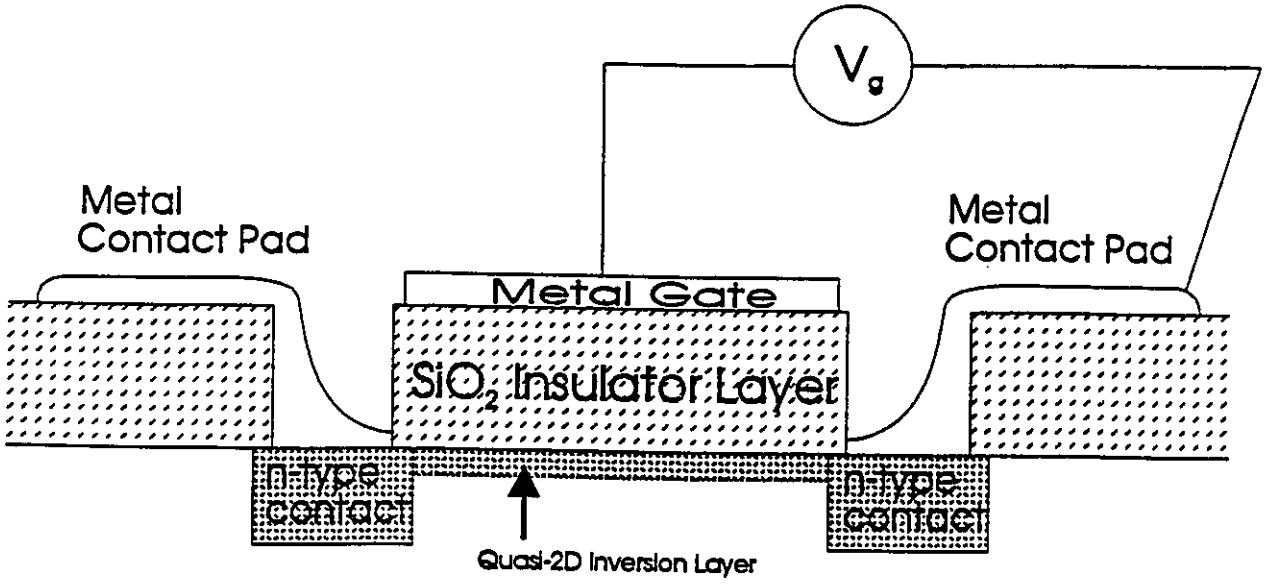

Figure 2.1: This is a cross-section of a typical Si-MOSFET showing the metal gate, silicon dioxide insulator, p-doped semiconductor, and two n-doped contacts with metal contact pads. The quasi-two dimensional electron layer (the inversion layer) forms at the interface between the semiconductor and the oxide.

is far greater than the thermal energies of 0.1 meV or less which electrons have in the temperature regime below $T = 1$ K. Thus the initial band picture for a MOSFET at low temperatures is the one shown in the upper half of Figure 2.2. The Fermi energy, E_F, lies well below the conduction band edge, E_C, and slightly above the acceptor energy band, E_A. Conduction by holes released from the acceptor band is more likely than conduction by electrons in the conduction band, but the ionization energies of typical doping impurities (i.e. 45 meV for boron) are still two orders of magnitude greater than the thermal energies available in this temperature range, so the MOSFET will act as an insulator in this state. If V_g is gradually increased, the gate will become increasingly positive with respect to the silicon and an electric field will develop in the z-direction. This electric field influences the potential of the bands in the silicon, causing them to bend downwards to lower energies. At a specific value of V_g, known as the threshold potential (V_t), the conduction band edge crosses the Fermi energy and electrons become available for conduction. Increasing V_g beyond this point moves the crossover point further into the silicon layer and thus increases the number of electrons in the conduction band. Eventually, the situation shown in the lower part of Figure 2.2 develops, where electrons exist in a triangular well of width Z_i, bounded on one side by the extremely high potential barrier of the oxide and on the other by the conduction band edge which rises linearly as it moves away from the oxide. Since this triangular region has electrons as the majority carrier rather than the holes which are expected for p-doped silicon, it is called an n-type inversion layer. In a typical experimental situation Z_i will be on the order of 50$\AA$, which is why the inversion layer is a quasi two dimensional system. The electrons are free to move in the x-y plane, but are fixed in the z-dimension.

2.1.2 Electronic States in the Inversion Layer

The gate potential perturbs the Hamiltonian for a free electron only in the z-direction, and even in that direction the electron wavefunction will normally vary faster than the gate potential, so the standard effective mass approximation techniques developed in three dimensional systems can be used. This gives a Hamiltonian of the form:

$$\left[\frac{\vec{p}^2}{2m^*} + V(z)\right]\Psi(x,y,z) = E\Psi(x,y,z) \tag{2.1}$$

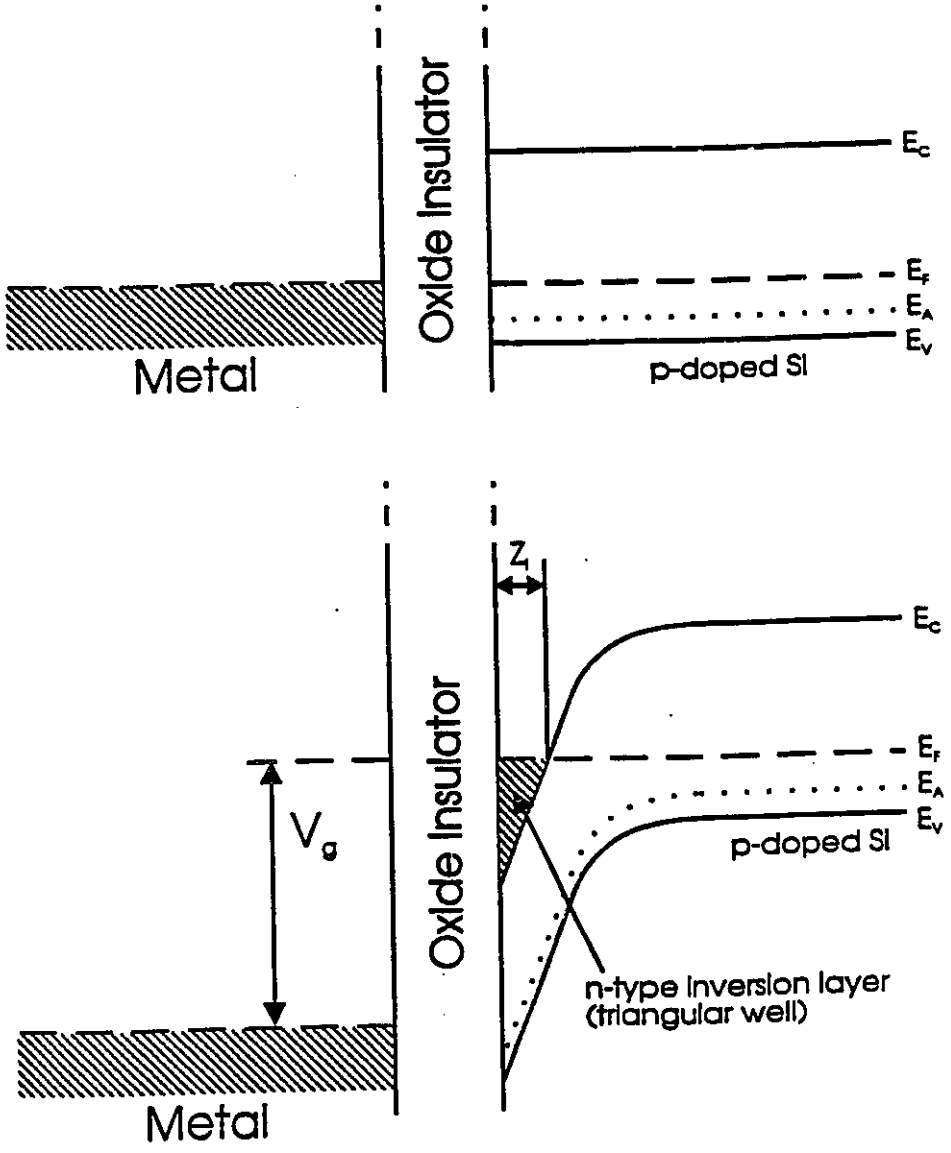

Figure 2.2: The upper diagram shows the band structure of a Si-MOSFET with zero applied gate voltage. The lower diagram shows how the gate potential creates an electric field which decreases the conduction band edge until an inversion layer is created. The electrons in this inversion layer in turn act to screen the electric field, so that the energy levels are unchanged far from the oxide-semiconductor interface.

For a parabolic band the variables are separable and for a perfectly isotropic effective mass equation, this becomes:

$$\left[\frac{p_z^2}{2m^*} + V(z) + \frac{p_{xy}^2}{2m^*}\right] \Psi(x,y,z) = E\Psi(x,y,z) \tag{2.2}$$

Here p_z is the component of electron momentum perpendicular to the interface and p_{xy} is the momentum component parallel to the interface and thus in the inversion layer. The periodic potential of the crystal is accounted for by the use of the effective mass approximation. The simplest way to solve this system is to use the envelope function approximation (See [And82], [Wei91] or [Ste67a] for details), in which:

$$\Psi(x,y,z) = C\,\Phi_l(z)\,\exp(i(k_x x + k_y y)) \tag{2.3}$$

Here C is a normalization constant and $\Phi_l(z)$ is the eigenfunction in the z-direction which is associated with the energy level of the electron in the potential well which defines the inversion layer. The x and y portion of this wavefunction is identical to that of free electrons in the three dimensional effective mass approximation, so that the electrons have unrestricted movement within the inversion layer. The equation which must then be solved in order to obtain $\Phi_l(z)$ is simply:

$$\left[\frac{p_z^2}{2m^*} + V(z)\right] \Phi_l(z) = E_l\Phi_l(z) \tag{2.4}$$

The energy of an electron which is in level l of the inversion layer potential well can now be written as:

$$E = \frac{\hbar^2 k_{xy}^2}{2m^*} + E_l \tag{2.5}$$

If the potential well is treated as a perfectly triangular well then the equation defining $\Phi_l(z)$ can be made more explicit:

$$-\frac{\hbar^2}{2m^*}\frac{d^2\Phi_l(z)}{dz^2} + eFz\Phi_l(z) = E_l\Phi_l(z) \tag{2.6}$$

where m^* is the effective mass of electrons along the z-axis and F is the electric field created by the potential difference between the gate and the semiconductor. It is assumed that the potential barrier of the oxide is infinitely high and that the semiconductor-oxide interface is located at $z = 0$, so that $\Phi_l(z) = 0$ for $z \leq 0$. The other boundary condition is that $\Phi_l(z)$

should approach zero as z approaches infinity. The solutions for this equation must then take the form of Airy functions (Ai) and for this particular differential equation the solutions turn out to be:

$$\Phi_l(z) = Ai\left(\left(\frac{2m^*}{\hbar^2 e^2 F^2}\right)^{\frac{1}{3}}(eFz - E_l)\right) \tag{2.7}$$

The energy levels in the well are given by:

$$E_l = -\left(\frac{e^2F^2\hbar^2}{2m^*}\right)^{\frac{1}{3}} a_l \tag{2.8}$$

where a_l is the l^{th} zero of Ai(a). For small l, this can be approximated by:

$$E_l \approx \left(\frac{\hbar^2}{2m^*}\right)^{\frac{1}{3}}\left(\frac{3\pi eF}{2}\left(l+\frac{3}{4}\right)\right)^{\frac{2}{3}} \tag{2.9}$$

Thus the electron energies are fully quantized in the z-direction and so only k_x and k_y can contribute to the density of states.

The triangular well approximation fits the Si-MOSFET particularly well, but any quasi 2DES will exhibit this quantization of energy levels in one direction and will thus have a two dimensional density of states (DOS) given by:

$$D_{2D}(E) = D_{2D}(k_{||})\frac{dk_{||}}{dE} = g_s g_v \frac{1}{(2\pi)^2}(2\pi k_{||})\frac{dk_{||}}{dE} \tag{2.10}$$

Here $k_{||}$ is the wavevector in the plane parallel to the plane of the 2DES (for a Si-MOSFET this is the bulk silicon cleavage plane on which the oxide was grown), g_s is the spin degeneracy factor and g_v the valley degeneracy factor. For motion in the plane of the 2DES, we assume a simple parabolic band structure and get:

$$E = \frac{\hbar^2 k_{||}^2}{2m^*} \tag{2.11}$$

Thus the density of states in two dimensions for a single quantum state is given by:

$$D_{2D}(E) = \frac{g_s g_v m^*}{2\pi\hbar^2} \tag{2.12}$$

This means that the DOS in two dimensions for a given value of E does not depend on E but simply on the number of states for the quantized z-axis which have energy less than E. It also leads to the very important conclusion that the density of states for the lowest level in a two-dimensional system

remains finite down to the very bottom of that level, whereas the density of states for a three dimensional system tends towards zero as the energy approaches the bottom of the level. Thus, quantities which depend on the DOS, such as optical absorption or scattering, do not approach zero in the limit of very few electrons in the conduction band. This difference in the density of states is illustrated by Figure 2.3, which shows the two dimensional DOS for a 2DES with $g_s = 2$ and $g_v = 1$ contrasted with the three dimensional DOS for an equivalent layer of the three dimensional semiconductor.

For a parabolic band structure, this gives:

$$k_F = \left(\frac{4\pi n_s}{g_s g_v}\right)^{\frac{1}{2}} \tag{2.13}$$

where n_s is the total number of electrons per unit area, so that the Fermi surface reduces to a circle of radius k_F. Perhaps the most important effect of the reduced dimensionality, however, is its effect on the screening capability of the electrons. In a three dimensional system, a point charge of magnitude Ze (assumed sufficiently weak to allow a linear response model to be used) will have a screened Coulomb potential of the form:

$$\phi = \left(\frac{Ze}{\kappa R}\right)\exp(-Q_s R) \tag{2.14}$$

where κ is the dielectric constant of the medium, R is the distance from the point charge, and Q_s is the three dimensional screening parameter for the semiconductor. For a two dimensional system in which the electrons are assumed to be in a perfectly 2D layer in the plane $z = 0$, and in which the external charge is located at a point $r = 0$, $z = z_0 \leq 0$, where r is a distance in the $z = 0$ plane, the asymptotic form of the screened Coulomb potential [Ste67b] is:

$$\bar{\phi}(r) \approx \frac{Ze(1 + \bar{q}_s z_0)}{\bar{\kappa}\bar{q}_s^2 r^3} \tag{2.15}$$

Here $\bar{\kappa}$ is the effective dielectric constant for the two dimensional layer (usually just the average of the dielectric constant in the two media on either side of the interface which defines the 2DES) and $\bar{q}_s$ is the two dimensional screening parameter given by:

$$\bar{q}_s = \frac{2\pi e^2}{\bar{\kappa}}\frac{dn_s}{dE_F} \tag{2.16}$$

Comparing Equations 2.14 and 2.15 we see that the potential decays as an inverse cube in the 2DES and as an exponential in the three dimensional

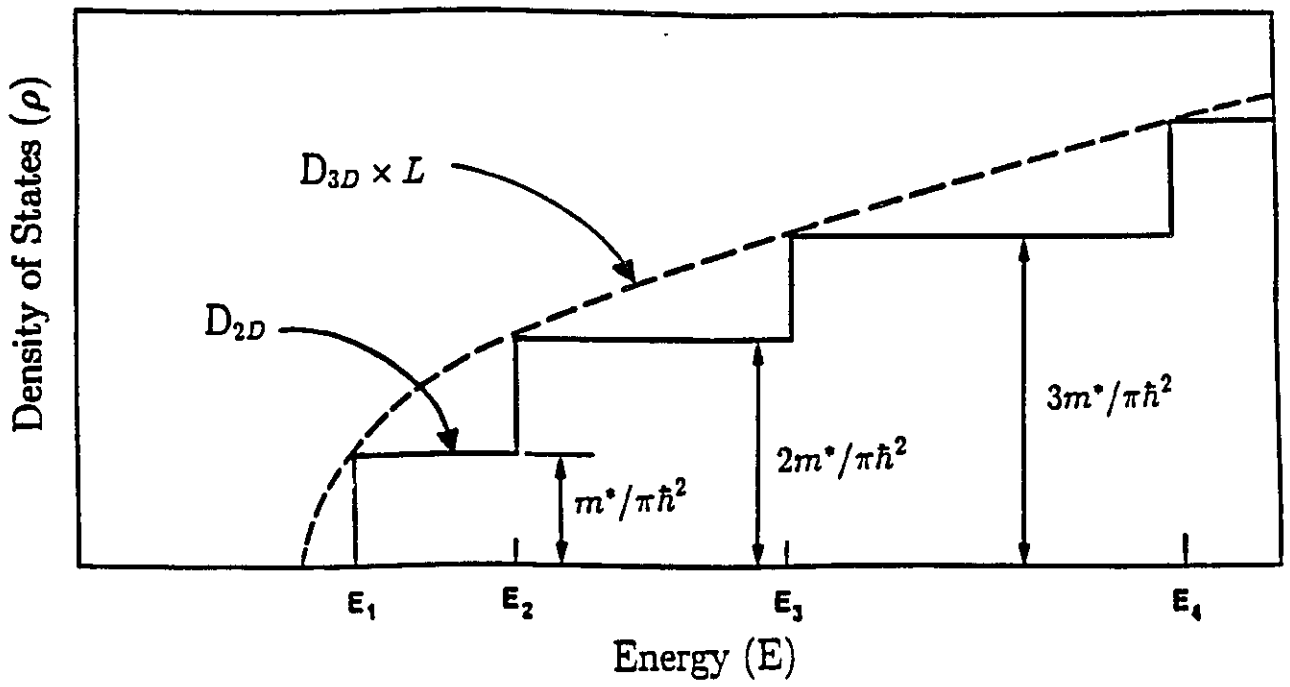

Figure 2.3: The solid line is the DOS for a 2DES with $g_s = 2$ and $g_v = 1$, whereas the dotted line indicates the DOS for a layer of a three dimensional system with equivalent thickness, L.

system, which means that, in general, potentials have much longer range in a two dimensional system. This increases the effects of the Coulomb interaction between electrons and also between charged impurities and electrons. Specific details of these calculations, and of how to generalize the screening equations to a finite width 2DES or to non-parabolic bands can be found in references [Ste67a, And82, Bas88].

2.1.3 Different 2DES Systems

It was stated in the beginning of this section that a two dimensional electron system can form at the interface between two different semiconductors, between a semiconductor and an insulator, or at a vacuum interface. The differences between these systems have important effects on the characteristics of the electrons they contain. Table 2.1 lists some of the most basic differences. The effective masses, m^*, are given in terms of the free electron mass, m_e, and the dielectric constants, κ, are relative to vacuum, which is why the electrons floating in a vacuum above the surface of liquid helium have a value of one in both columns. The two semiconductor systems have quite different effective masses, dielectric constants and valley degeneracies.

Figure 2.4 shows the Brillouin zone of a silicon crystal. It has six equivalent energy minima located along the (100) directions. Each of these minima is called a valley and has its own electrons with effective masses which must be calculated relative to the long axis of the valley, as was first studied by J. Luttinger and W. Kohn [Lut55, Koh55]. The effective mass for motion perpendicular to the long axis of each valley is $m_t = 0.19m_e$ and the effective mass for motion parallel to this axis is $m_l = 0.916m_e$. This six valley system should produce a sixfold degenerate ground state for the electrons. Experiments have revealed that this is not the case, however, because scattering can occur between electrons in different valleys [Bal70]. The result is that the electronic ground state in silicon is actually split into three separate levels with degeneracies of 1, 2, and 3 respectively. The degeneracy is further reduced in the two dimensional case when the silicon crystal is cleaved so that a MOSFET structure can be grown. It is usually cleaved perpendicular to one of the (100) directions, and since the electrons in the inversion layer are confined to move parallel to this cleavage plane, only two of the valleys will contain electrons which move in the inversion layer with the lower effective mass, m_t, and have the higher effective mass, m_l, for motion perpendicular to the inversion layer. Equation 2.8 showed that the confinement energy of electrons in the inversion layer is inversely proportional to the effective mass in the z-direction and so these two valleys will have lower electron energy

2-D System	m^*	κ	g_v
Electrons on Liquid Helium	1.0	1.0	1
Si-MOSFET	0.19	7.7	2
GaAs/GaAlAs Heterostructure	0.067	13	1

Table 2.1: Three Different Two-Dimensional Electron Systems

states and will be filled with electrons first. This reduces the valley degeneracy for electrons in a Si-MOSFET to $g_v = 2$. The value of m^* quoted in the table is the effective mass, m_t, in these remaining two valleys.

Another difference between the two semiconductor systems is the band structure. Figure 2.5 shows the three dimensional energy levels along several different axes for both silicon and gallium arsenide. The specific labelled points correspond to certain symmetry points which can be seen in the Brillouin zone of Figure 2.4 (which applies equally to both semiconductors even though only silicon has the six degenerate valleys). The energy scale is in electron-Volts and has arbitrarily set the zero value at the top of the valence band. Since effective mass is inversely proportional to the curvature of the energy bands, it is easy to see that there are energy bands with different effective masses even at the same point in the Brillouin zone. The effective masses which were used in Table 2.1 are those for the first conduction sub-band (i.e. the band with the lowest energy above $E = 0\,\text{eV}$), and for a typical Si-MOSFET this is the only band which should be significant until the electron concentration exceeds $2 \times 10^{12}\,\text{cm}^{-2}$ [And82]. The figure also clearly shows that gallium arsenide is a direct gap semiconductor (i.e. the maximum energy position in the valence band is located at the Γ point, the same point in the Brillouin zone as the minimum energy position in the conduction band), whereas silicon is an indirect gap semiconductor in which the minimum energy gap occurs for an electron excited from the Γ point to the minimum of the first conduction sub-band which occurs along the $\Gamma - \Delta - X$ Axis. This is important because it means that exciting an electron to the conduction band requires only energy in gallium arsenide, but both energy and momentum in silicon. These differences must be kept in mind when analyzing experiments performed on the different two dimensional systems.

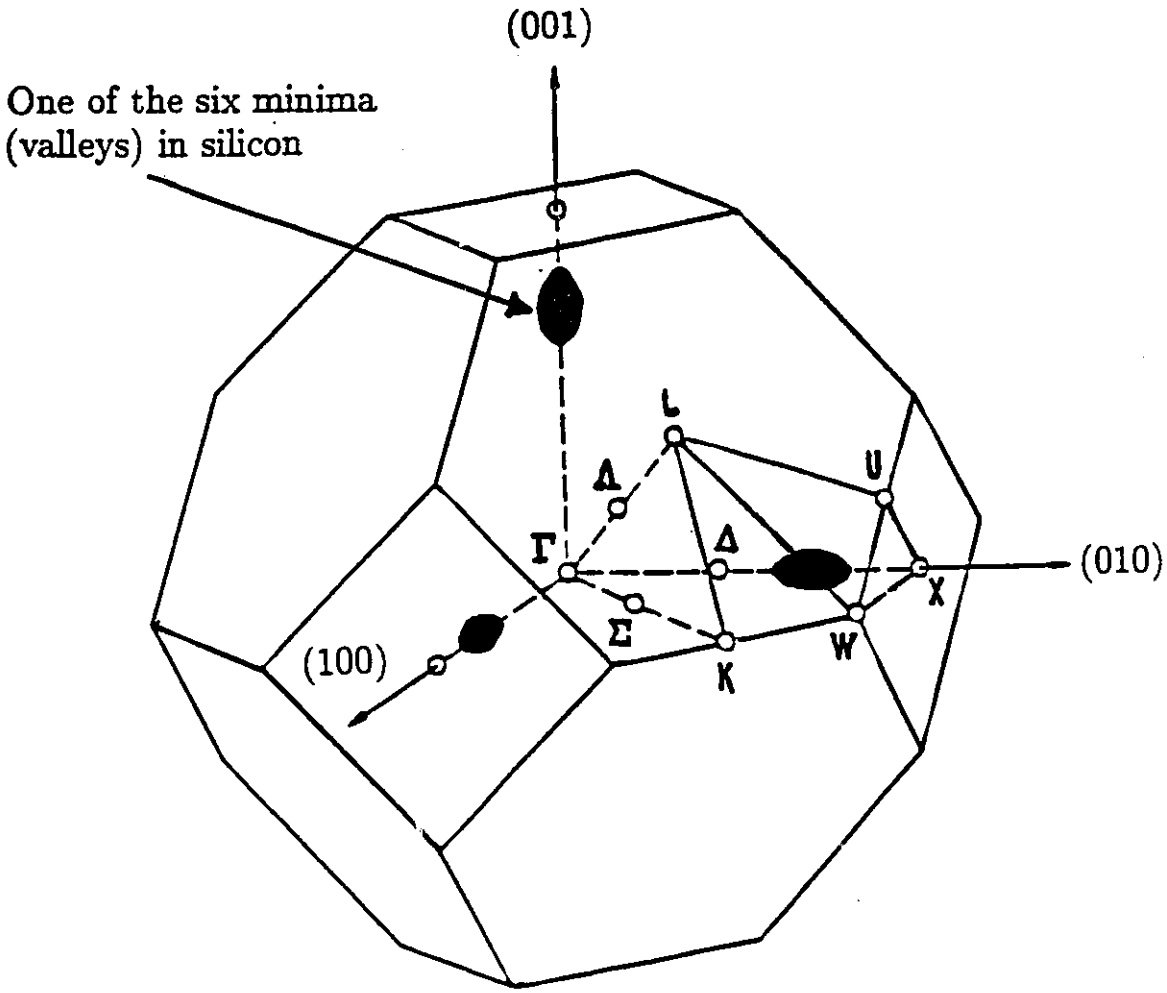

Figure 2.4: This figure shows the first Brillouin zone of silicon, revealing its six equivalent energy minima along the (100) directions.

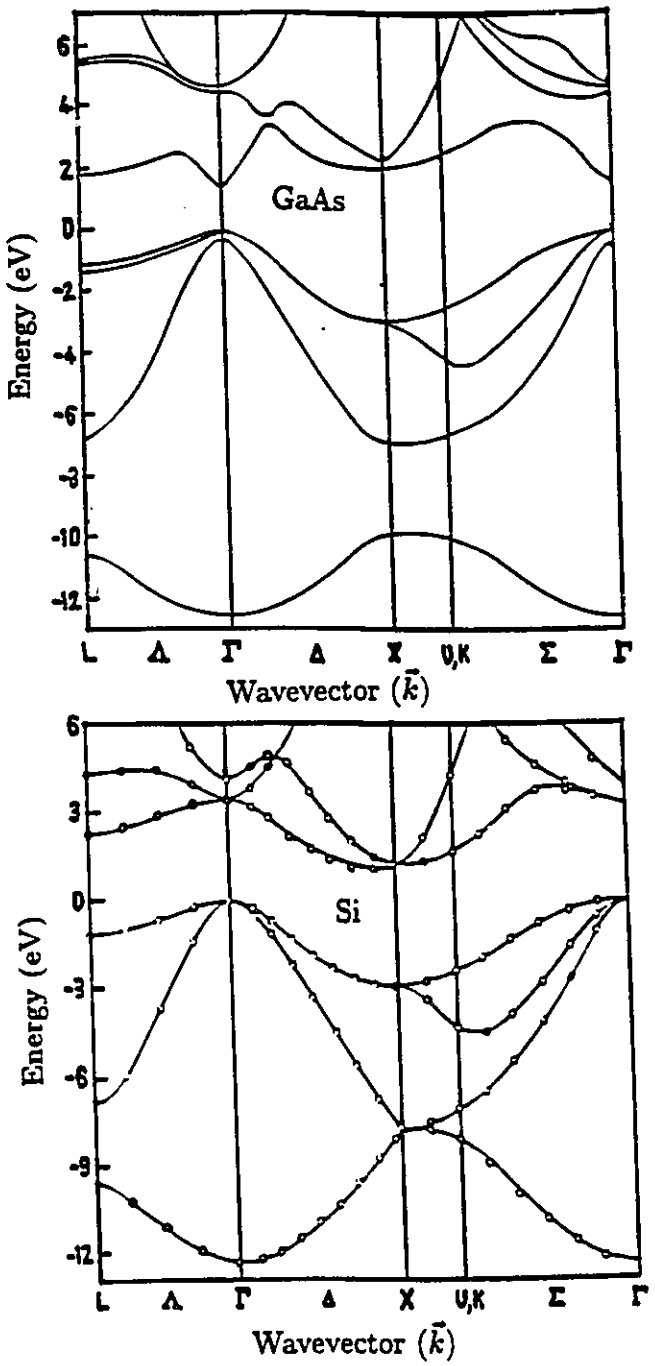

Figure 2.5: The upper diagram shows the band structure of gallium arsenide and the lower shows the band structure of silicon. The energy scale in each case is relative to the top of the valence band.

2.2 Perpendicular Magnetic Fields

2.2.1 Landau Levels

It is a common technique in electronic transport measurements to apply a magnetic field in the z-direction. This field exerts a Lorentz force on the electrons moving in the plane of the 2DES, and quantizes their energy states into Landau levels. This can be expressed mathematically by defining a vector potential, $\vec{A} = (0, Bx, 0)$, which corresponds to the magnetic field $\vec{B} = (0, 0, B)$. The Schrodinger wave equation for the x-y plane component of the electron wave function, Ψ, is given by:

$$H\Psi(x,y) = \frac{1}{2m^*}(\vec{p} - e\vec{A})^2\Psi(x,y) = E\Psi(x,y) \tag{2.17}$$

where $\vec{p} = (p_x, p_y, 0)$ is the momentum of the electron in the 2DES. Substituting in the trial wave function $\Psi = \exp(ik_y y)\Phi(x)$, and using the standard convention in which $p_x = -\imath\hbar\frac{\partial}{\partial x}$ and $p_y = -\imath\hbar\frac{\partial}{\partial y}$ leads to an equation of the form:

$$\frac{1}{2m^*}\left[\frac{-\hbar^2\partial^2}{\partial x^2} + \hbar^2k_y^2 - 2e\hbar k_y Bx + e^2B^2x^2\right]\Phi(x) = E\Phi(x) \tag{2.18}$$

It is now convenient to define the terms, $\omega_c = eB/m^*$, as the cyclotron frequency, and $l = \sqrt{\hbar/eB}$, as the magnetic length. These define the fundamental scales of what can now be expressed as a shifted harmonic oscillator problem.

$$\left[\frac{p_x^2}{2m^*} + \frac{m^*w_c^2}{2}(x - k_y l^2)^2\right]\Phi(x) = E\Phi(x) \tag{2.19}$$

The solutions of Equation 2.19 are quantized and have energies given by:

$$E_n = \hbar\omega_c(n + \frac{1}{2}) \tag{2.20}$$

Here n is called the Landau level index and when periodic boundary conditions are applied it becomes apparent that each Landau level can contain a limited number of electron states per unit area:

$$n_B = g_s g_v \frac{1}{2\pi l^2} = \frac{g_s g_v eB}{h} \tag{2.21}$$

The filling factor, ν is then defined as the number of Landau levels which are filled at a given electron density (i.e. $\nu = n_s/n_B$). This analysis reveals that

the electronic states in a 2DES subject to a magnetic field are quantized in energy so that each energy level for the confinement potential in the z-plane is further broken down into Landau levels which are spaced further apart as the magnetic field increases. All of the experiments discussed in this thesis occur in the low electron density regime where all of the conduction electrons are in the ground state of the confinement potential. Thus the Landau levels form the main ladder of energy levels. There can also be additional energy splittings due to spin and valley interactions, however, and these create further sub-levels around each main Landau level. This modifies Equation 2.20 to become:

$$E_n = \hbar\omega_c(n + \frac{1}{2}) \pm \frac{1}{2}g^*\mu_B B + \Delta_v(B) \tag{2.22}$$

Here g^* is the effective g-factor which reflects the enhancement of the exchange energy due to electron-electron interactions, $\mu_B = e\hbar/2m_e$ is the Bohr magneton, and $\Delta_v(B)$ is the field dependent valley splitting energy (in silicon, $\Delta_v(B) = 2.4 + 0.6B$ measured in units of Kelvin). This situation is illustrated in Figure 2.6, where the density of states per unit energy is shown for several Landau levels, first with no additional interactions, then with spin splitting, and finally with valley splitting.

2.2.2 Quantum Hall Effect

Figure 2.7 shows a top view of a typical Hall bar configuration for a 2DES, which has a longitudinal current, I_{SD}, flowing from source to drain and three pairs of contacts which can be used to measure both the longitudinal voltage, V_L, or the Hall voltage, V_H. A magnetic field applied in the z-direction (orthogonal to the paper) will subject the electrons to a Lorentz force in the y-direction of $-ev_dB$, where v_d is the electrons' drift velocity and B is the magnetic field in Tesla. The classical Hall effect describes how this Lorentz force will cause electrons to accumulate on one side of the Hall bar until the electric field created by repulsion from the charge is sufficient to counter the Lorentz force and the y-component of the electron drift is once again zero. This creates a classical Hall voltage, $V_H = I_{SD}B/en_s$. In a two dimensional system, where the resistivity tensor has a very simple form, it is common to convert V_H and V_L to resistivities.

$$\rho_{xx} = \frac{V_L}{I_{SD}\frac{2l}{W}}, \quad \rho_{xy} = \frac{V_H}{I_{SD}} = \frac{B}{en_s} \tag{2.23}$$

Thus the initial expectation is that ρ_{xy} should be directly proportional to B and ρ_{xx} should be constant. If ρ_{xx} and ρ_{xy} are measured for a good quality

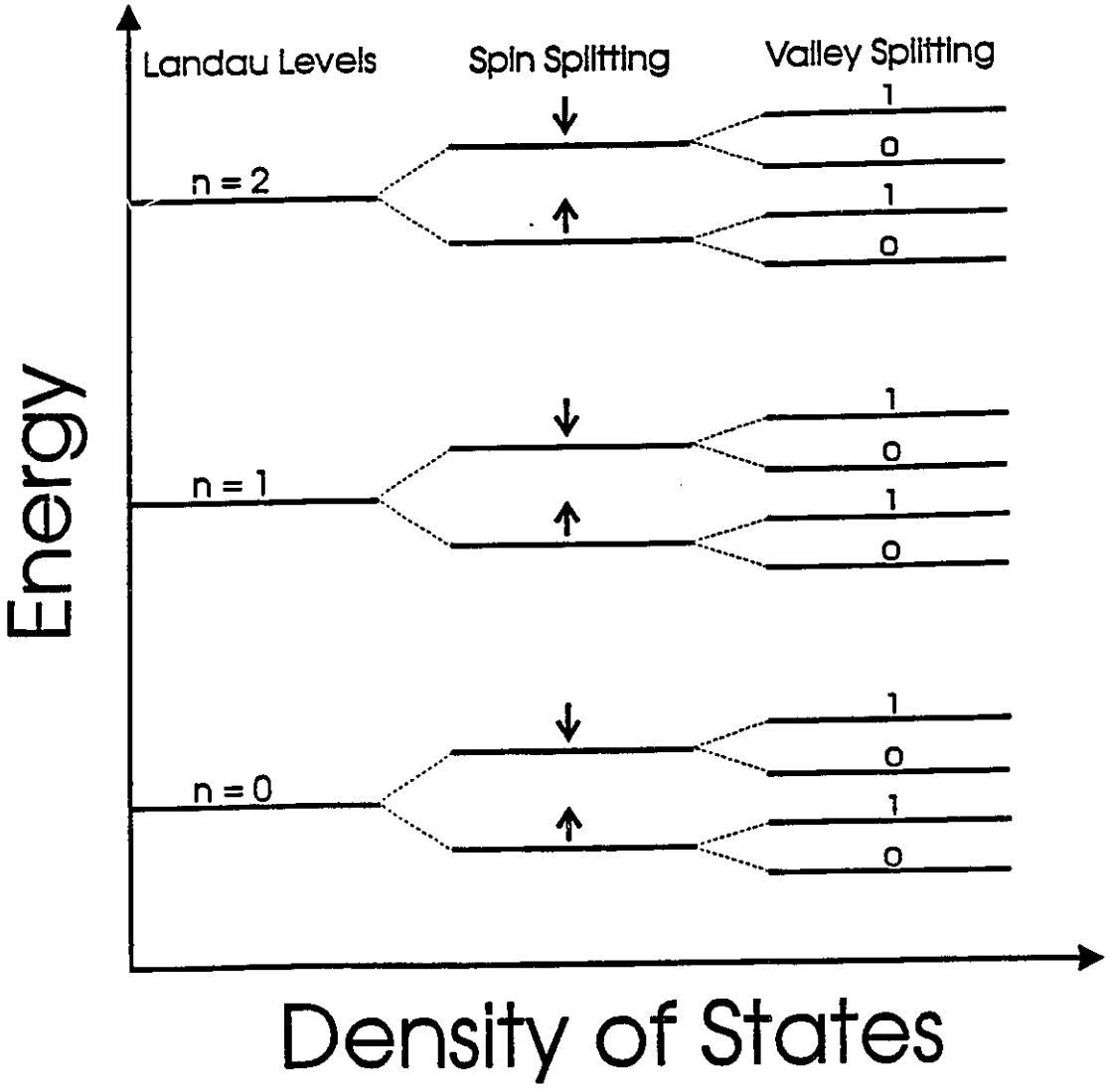

Figure 2.6: This figure shows the density of states per unit energy for a 2DES in which a magnetic field has created Landau levels. The effects of spin and valley splitting are also shown. Note that the density of states is a delta function with non-zero values only at specific energy values.

sample at low temperature, however, the results are much more complex, as shown in Figure 2.8.

At low magnetic fields the Hall resistivity follows the expected classical line of $\rho_{xy} = B/en_s$, but there are oscillations in the value of ρ_{xx} from maxima to minima as the magnetic field increases. These oscillations, which are called the Shubnikov-de Haas effect, are caused by the Landau level quantization.

For this type of metallic conduction, only the electrons which are within an energy interval of k_BT of the Fermi surface can take part in the diffusion process. The resistivity is proportional to the probability of an electron being scattered, which is itself proportional to the density of states available for the electron to scatter into. This density of states increases dramatically, and thus ρ_{xx} has a maximum, every time a Landau level passes through the Fermi level such that:

$$E_F = (n + \frac{1}{2})\hbar\omega_n = (n + \frac{1}{2})\frac{\hbar eB_n}{m^*} \tag{2.24}$$

where n is the Landau level index and spin and valley effects have been ignored for convenience. For small values of the magnetic field (corresponding to large values of n) the dependence of E_F on B can be neglected and the Shubnikov-de Haas oscillations are periodic in $1/B$, with a period $\Delta(1/B)$ which depends on the Fermi energy and thus on the electron density, $n_s = D_{2D}(E)E_F$, since we have assumed that only the first sub-band in the inversion layer is filled. This leads to the expression:

$$\Delta(1/B) = \frac{\hbar e}{m^*E_F} = \frac{\hbar eD_{2D}(E)}{m^*n_s} = \frac{e}{2\pi\hbar n_s} \tag{2.25}$$

This phenomenon can be used to accurately determine the electron density, n_s, in a 2DES. The maxima can be labelled from high to low magnetic field and a plot of maximum number versus inverse magnetic field should then result in a straight line with slope $2n_sh/e$. The Shubnikov-de Haas effect can also be used to determine the quantum lifetime, τ_q, (the mean time between scattering events) in the 2DES using the temperature dependence of the magnitude of the resistance peaks [Col91]:

$$\Delta R = 4R_0\frac{2\pi^2k_BT/\hbar\omega_c}{\sinh(2\pi^2k_BT/\hbar\omega_c)}\exp\left(\frac{-\pi}{\omega_c\tau_q}\right) \tag{2.26}$$

Here, ΔR is the change in resistance at a maximum or minimum from the zero magnetic field value, R_0. There are some important constraints, however, which limit the regime in which the Shubnikov-de Haas effect is observed.

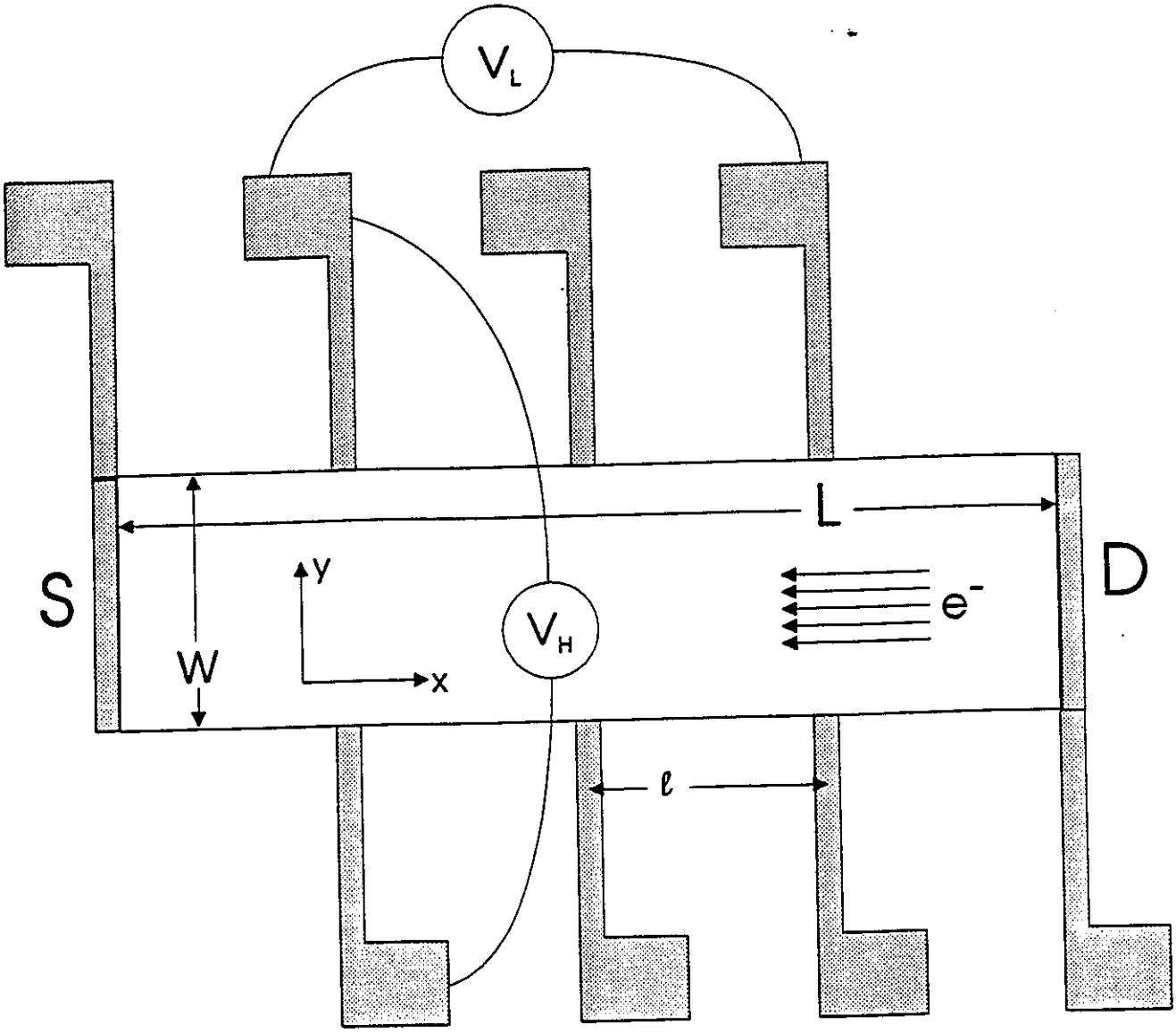

Figure 2.7: This is a top view of a commonly used Hall bar geometry in which the 2DES is in the plane of the paper and the magnetic field is applied perpendicular to it. A constant current flows from Source to Drain, and V_L and V_H are simultaneously measured.

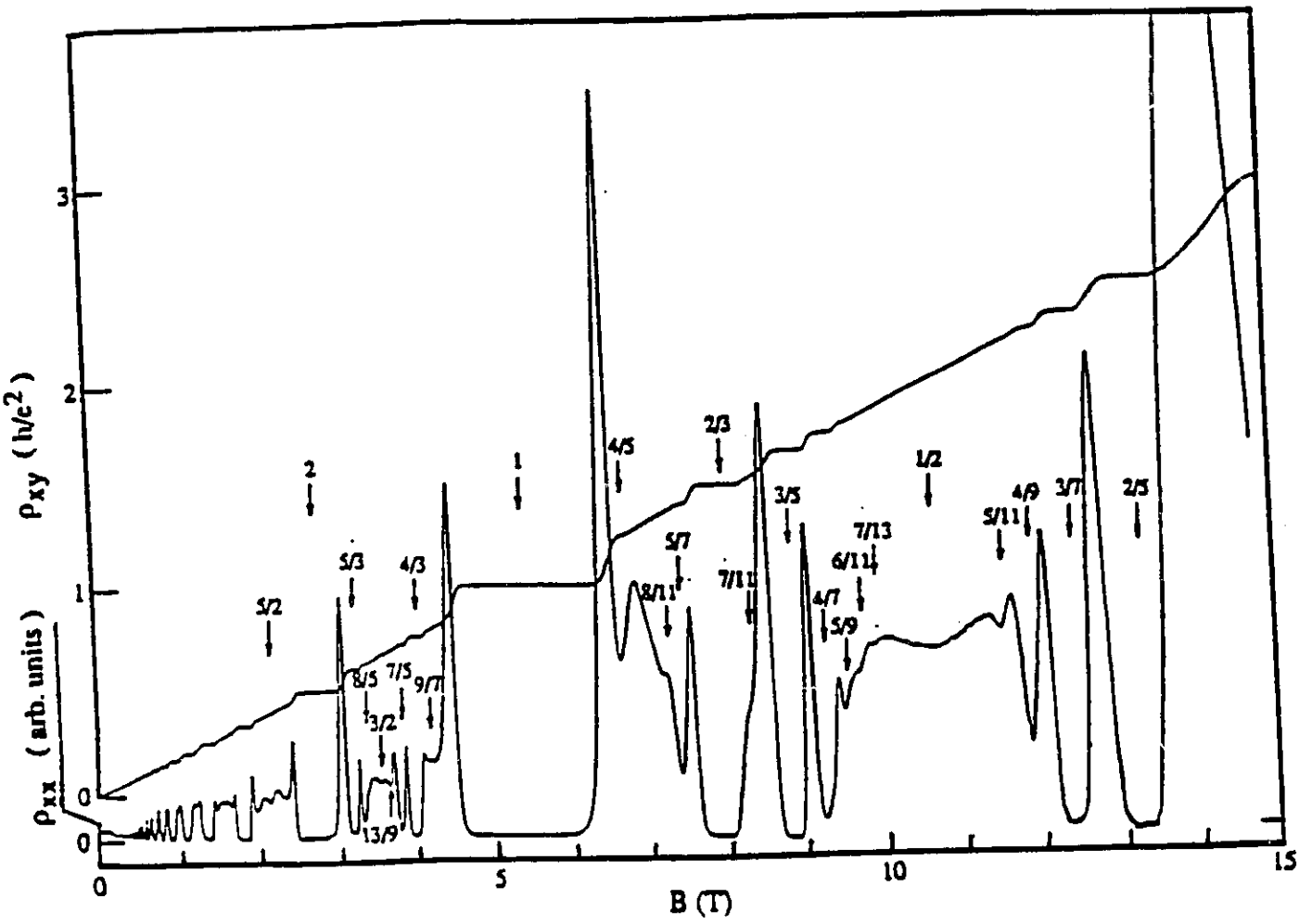

Figure 2.8: The graph of ρ_{xx} in arbitrary units and ρ_{xy} in units of h/e^2 versus B shows how the longitudinal resistance approaches zero and the Hall resistance is quantized at integer and specific fractional Landau level filling factors. This figure is reproduced from [Saj90].

The first of these is that the Fermi distribution function has a finite width of k_BT, so that if the spacing between adjacent Landau levels is less than k_BT then it is not possible to distinguish between the Fermi level being coincident with a Landau level or lying between adjacent Landau levels. The Shubnikov-de Haas effect is therefore only observed when B is large enough that:

$$\hbar\omega_c > k_BT \tag{2.27}$$

The second constraint is caused by the quantum uncertainty in the energy of electrons, $\Delta E = \hbar/\tau_q$, where τ_q is the quantum lifetime of an electron. This uncertainty must also be less than the energy gap between Landau levels and this leads to the equation $\hbar/\tau_q < \hbar\omega_c$ or alternatively:

$$\omega_c\tau_q > 1 \tag{2.28}$$

The electronic transport mobility, which is commonly used as an indicator of the sample quality, is defined in terms of the average time between scattering events, $\mu_t = e\tau_t/m^*$, but it is common to assume that $\tau_t = \tau_q$ so that Equation 2.28 can be expressed as:

$$\omega_c\tau_q = \frac{eB}{m^*}\frac{m^*\mu_t}{e} = \mu_t B > 1 \tag{2.29}$$

It is also important to realize that spin and valley effects also become important at some finite magnetic field, and these must then be taken into account in calculating n_s. The general principle, however, is always the same.

The Shubnikov-de Haas effect is responsible for the oscillations in ρ_{xx} at low magnetic fields where ρ_{xy} is still following the classical line, but at slightly higher magnetic fields the minima in ρ_{xx} become wider and deeper and occur simultaneously with plateaus in ρ_{xy} which have quantized values of the form $h/(je^2)$, where j is an integer. This is called the integer quantum Hall effect (IQHE). Very clear examples of the plateaus and corresponding minima are shown in Figure 2.8 for $j = 1$ and $j = 2$. At even higher magnetic fields, the fractional quantum Hall effect (FQHE) can occur. This manifests itself in the same way as the IQHE, except that j can be one of a special hierarchy of fractions, many of which are shown in Figure 2.8. The discovery of the IQHE was initially made by K. von Klitzing, G. Dorda, and M. Pepper using Si-MOSFET samples and was published in 1980 [Kli80]. The FQHE was not discovered until 1982 [Tsu82], when high quality GaAs/AlGaAs samples were available. In subsequent years, however, the FQHE was also observed in Si-MOSFETs [Pud84, Fur86], but the majority of experimental work on the FQHE has utilized GaAs/AlGaAs heterostructures. The theory which

underlies both the integer and fractional QHE has been well developed since 1982, and some excellent reference texts are available [Cha88, Pra90, Lan92]. It has become apparent that both phenomena are related to the interaction of electrons with magnetic flux quanta. For the IQHE, this occurs in a regime where electron-electron interactions through the Coulomb force are small so that the charge per electron in the highest Landau level is the relevant quantity (thus j is an integer) whereas for the FQHE it occurs in a higher field regime where electron-electron interactions must be included because the rigid binding of the electrons into cyclotron orbits reduces their kinetic energy. In this regime the excitations which are measured are fluctuations in the multi-electron groundstate called quasiparticles which have fractional charge (allowing j to be a fraction).

Understanding the IQHE and FQHE requires one further addition to the picture of Landau levels which was shown in Figure 2.6. In any real sample the Landau levels are broadened by impurity scattering [Har85], so that the density of states follows a Gaussian or Lorentzian form rather than a delta function at the centre of each Landau level. An example of this more realistic situation is shown in Figure 2.9. The density of states between Landau levels is small enough that most of the electron states are localized by impurity potentials, whereas the larger density of states near the centres of the Landau levels allows some electron states to remain extended and able to conduct. If the Fermi energy lies in the region of localized states between Landau levels j and $j+1$ (so that $\nu = j$) and if $\hbar\omega_c$ is large enough that there is no overlap between the conducting regions near each Landau level centre, then the concentration of conduction electrons, n_s, which enters into the equation defining ρ_{xy} will remain at jn_B, the number of states contained in j completely filled Landau levels, until the magnetic field increases sufficiently that the Fermi energy reaches Landau level j and some of the electron states start to depopulate. Combining Equations 2.21 and 2.23 (and ignoring spin and valley splittings) results in a Hall resistivity of:

$$\rho_{xy} = \frac{V_H}{I_{SD}} = \frac{B}{en_s} = \frac{B}{ej\left(\frac{eB}{h}\right)} = \frac{h}{je^2} \tag{2.30}$$

Thus the Hall resistance will be quantized over a finite range of the magnetic field, just as is observed in the IQHE. The vanishingly small longitudinal resistance can also be simply explained by noting that there are no unoccupied extended states for the conducting electrons in the lower Landau levels to scatter into when the Fermi energy lies between Landau levels. Thus the scattering probability becomes extremely low and so does ρ_{xx}. This provides a simple explanation of the IQHE.

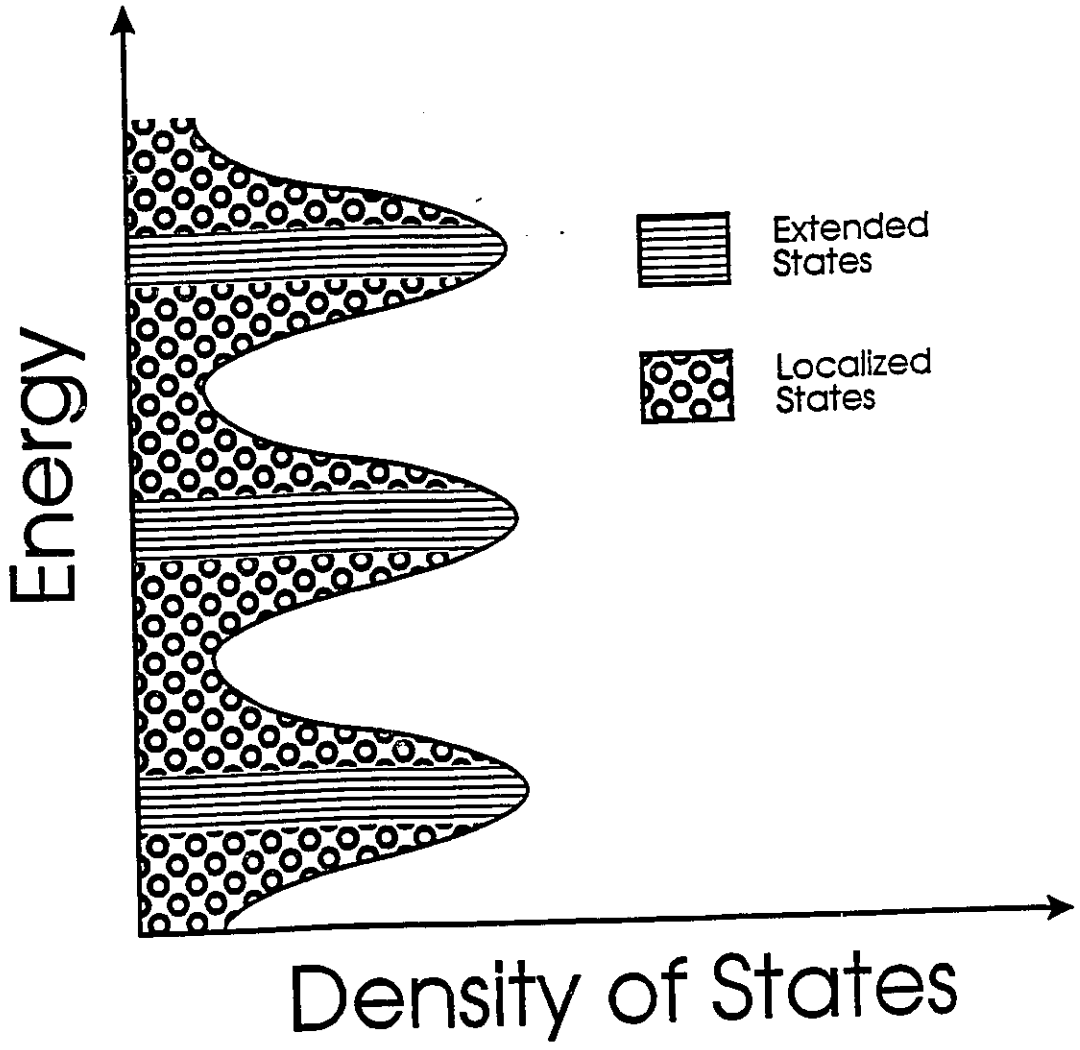

Figure 2.9: The two dimensional density of states per unit energy under a perpendicular magnetic field with impurity broadening of the Landau levels is shown. There are cores of extended states at the centre of each Landau level surrounded by a Lorentzian distribution of localized states.

In the FQHE regime, electrons can not be treated independently and it is necessary to solve for the many electron wave function. This can not be done exactly for a real system because there are too many electrons. It is possible, however, to use variational methods to approach the solution for an ideal system, a method which was first tried succesfully by R. B. Laughlin [Lau83a, Lau83b]. He proposed a type of Jastrow function solution:

$$\Psi_m = \prod_{j<k}^{N_e} (z_j - z_k)^m \prod_{j=1}^{N_e} e^{-|z_j|^2/4l^2} \tag{2.31}$$

Here $z_j = x_j - iy_j$ is the position of the j^{th} electron (which is confined to a circle of radius l by the magnetic field) expressed as a complex number, N_e is the total number of electrons in the system, and $m = 1/\nu$ is the inverse filling factor. For m any odd integer, this wave function satisfies Fermi statistics and is a possible solution for the multi-electron wave function. When the probability distribution of the electrons, $|\Psi_m|^2$, is worked out, it has the same form as the probability distribution function of a classical liquid with particle density $\rho_m = 1/(2\pi l^2 m)$. The energy of the system can also be worked out numerically for this ground state as a function of ν and it turns out to have cusps at the odd integer values of m. This means that any deviation from these filling factors will require a finite amount of energy, which is why the ground state of the FQHE is often referred to as an incompressible electron liquid. The finite energy which is required to change the density by a single particle in the classical liquid analogy must give rise to some type of excitation in the multi-electron wave function. When Laughlin calculated the wave function for this elementary excitation, it became apparent that it carried a charge of $\pm e/m$. He called the excitation with positive charge a quasihole and the excitation with negative charge a quasielectron. Physically, they can be thought of as an area of approximately πl^2 in which the local electron density is slightly different from the average background density of the liquid so that $1/m$ of an electron is either accumulated or expelled.

The quasielectrons and quasiholes behave very similarly to normal electrons or holes in the lowest Landau level. They can exist in extended states and conduct charge or they can be localized by impurity potentials. They also exist in bands, although in this case the bands are not centred at completely filled Landau levels but at the fractional filling factors where the FQHE ground state is stable. This means that in a manner exactly analogous to that described above for the IQHE, the quasiparticles can give rise to a quantized ρ_{xy} and a vanishing ρ_{xx} when the magnetic field is between two fractional filling factors and the small number of quasiparticles which

exist so far from the stable filling factor are all in localized states. Going one step further, it is even possible for the quasiparticles to form their own incompressible quantum fluid. The ground states for these quantum fluids will then have their own excitations which are quasielectrons and quasiholes with different fractional charge. This series of different quantum fluids which arises out of each of the initial $\nu = 1/m$ states is called the hierarchy of states. It explains most of the fractions in Figure 2.8 which do not fit the initial $\nu = 1/m$ (quasielectrons) or $\nu = 1 - 1/m$ (quasiholes) model. The first few fractions which develop from the $\nu = 1/3$ state, for instance, are 2/5, 3/7 and 4/9.

The only phenomena shown in Figure 2.8 which are not explained by either the Shubnikov-de Haas effect, the IQHE or the standard FQHE models are those which occur at the half integer filling factors, $\nu = 1/2, 3/2$ and 5/2. The Hall resistivity does not deviate from the classical line at these filling factors in Figure 2.8, but there are unusual features in ρ_{xx}. Other researchers have observed a quantized Hall plateau at $\nu = 5/2$ in a single layer 2DES [Wil87] and at a total filling factor of 1/2 in a double layer system [Sue92, Eis92]. A new theoretical approach has been developed which has had some success in explaining this phenomenon for $\nu = 1/2$ or $\nu = 1/4$ by treating each electron as a fermion which carries an integer number of flux quanta of a Chern-Simons gauge field [Hal94]. This produces an effective magnetic field which changes the effective filling factor for the transformed electrons.

Another fascinating property of the QHE is that it exhibits some universal scaling phenomena. The half widths of the ρ_{xx} maxima in between the IQHE minima and also for some transitions between certain FQHE minima have been measured to vary with temperature as $\Delta B \propto T^{\kappa}$ where $\kappa = 0.42 \pm 0.04$. At the same time, the maxima of $d\rho_{xy}/dB$ (which indicate the sharpness of the transition from one plateau to the next) have been observed to vanish with increasing temperature as $(d\rho_{xy}/dB)_{max} \propto T^{-\kappa}$. This value of κ appears to be a universal scaling exponent for many of the measured transitions [Wei88, Jai90] in GaAs/AlGaAs heterostructures. In the IQHE regime, another fundamental behaviour has been clearly observed. If the magnetic field is held constant and the temperature lowered towards zero, then the measured values of $(\sigma_{xx}, \sigma_{xy})$ will follow flow lines towards the fixed points $(\sigma_{xx}, \sigma_{xy}) = (0, n)$, where n is an integer [Wei85]. This is in agreement with a theoretical two-parameter scaling theory developed by A. M. M. Pruisken [Pru85]. The original diagram from Pruisken's paper is repeated in Figure 2.10 to provide a more visual illustration of the flow and the fixed points. This type of universal behaviour has led to the suggestion [Jai90] that all

plateaus in both the IQHE and FQHE regimes can be assigned to one of several universality classes of minima which occur at filling factors:

$$\nu = \frac{n}{n(m-1)+1} \tag{2.32}$$

In this theory all minima with the same value of m would correspond to the same universality class and transitions between them should show the same scaling behaviour and universal exponent, but for transitions between minima in different universality classes the scaling behaviour would not be observed.

2.3 Impurities and Localization

2.3.1 Introduction to Impurities

All samples are affected by the presence of impurities, since there are always impurities incorporated into the sample during the growth process. This was mentioned in the previous section in relation to the broadening of the density of states about the Landau level centres and the reduction in the number of extended electron states due to the presence of impurities. The impurities have two basic effects. The first is that they act as scattering centers so that the mean time between scattering events, τ, is decreased as the concentration of impurities, n_i increases. This smoothes out any discontinuous changes in the density of states such as those which occur at the centres of Landau levels or at the conduction band edge so that they take place over an energy range, $\Gamma = \hbar/2\tau$. Both charged (Coulombic) impurities and neutral impurities (such as those caused by interface defects at the silicon-oxide interface in a Si-MOSFET) can contribute to this broadening. The second effect is that the Coulombic impurities can localize electrons within their potential wells, destroying the extended states which are required for metallic conduction. It is simpler to understand the effects of impurities and localization in the absence of a magnetic field, so that is the first case which will be considered.

If the density of states (DOS) around the Fermi energy in a 2DES at zero magnetic field is sufficiently high, the electrons will be able to screen the impurity potentials and extended states will exist. The 2DES will then have the transport characteristics of a metal, in that the diagonal conductivity, σ_{xx}, will decrease with decreasing temperature, and at the low temperatures ($T < 1\,\mathrm{K}$) relevant to this thesis the diagonal resistivity, $\rho_{xx} < h/e^2 \approx 25\,\mathrm{k\Omega}$. When the DOS is low and the electrons are localized the 2DES becomes insulating. In this case, σ_{xx} will decrease with increasing temperature and

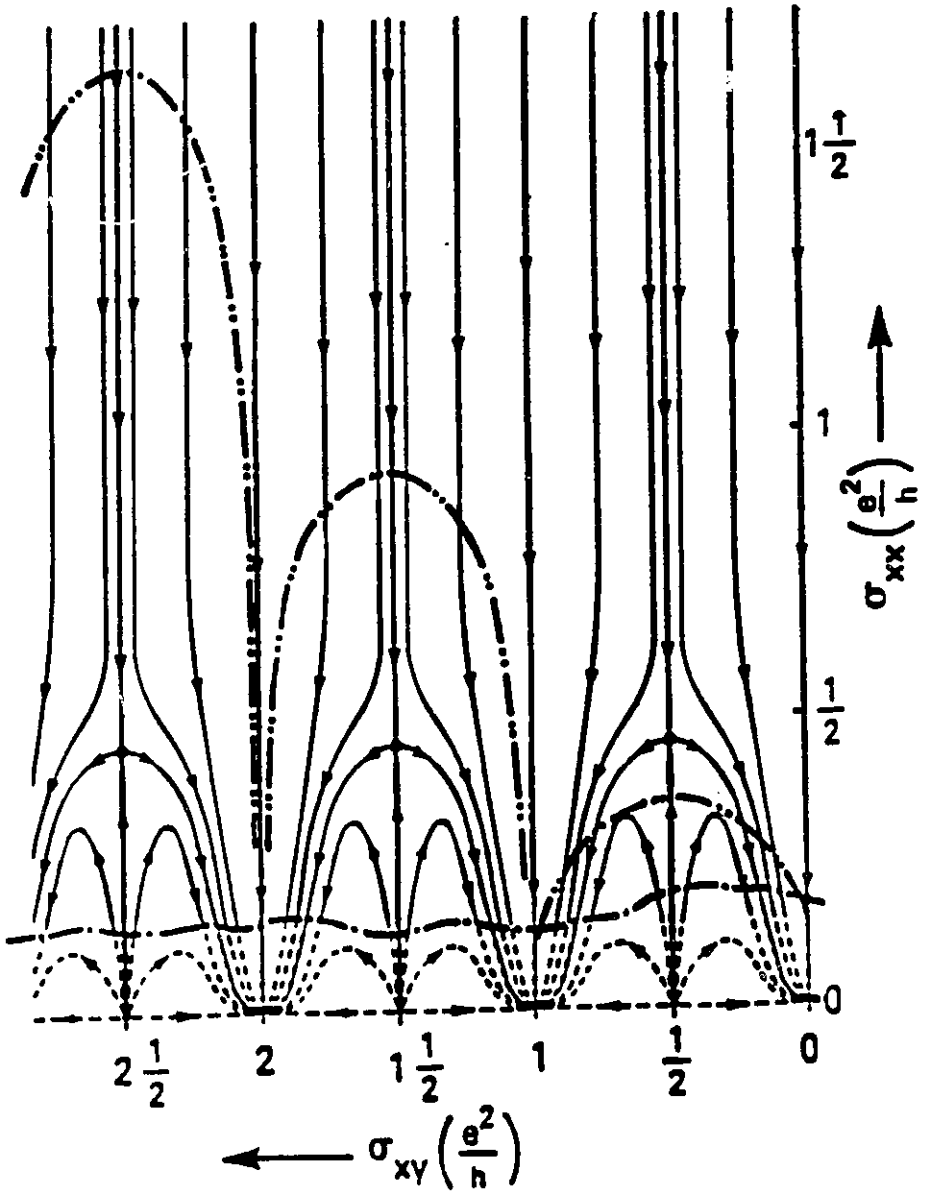

Figure 2.10: This figure shows the flow lines and the stationary points for the quantum Hall effect as a diagram of σ_{xx} versus σ_{xy} from the scaling theory of A. M. M. Pruisken [Pru85].

$\rho_{xx} >> h/e^2$. The strict theoretical definitions are that for a metal, σ_{xx} is finite as $T \rightarrow 0$ and for an insulator, $\sigma_{xx} \rightarrow 0$ as $T \rightarrow 0$, but since experiments can not be performed at $T = 0$, the working definitions given above are more useful. Since the DOS in a semiconductor at zero magnetic field (See Figure 2.3) increases with increasing energy, there will always be some energy, E_m, above which the density of states is sufficiently large that there will be extended states. E_m is called a mobility edge, because for $E_F > E_m$ the electrons are in mobile extended states, whereas for $E_F < E_m$ the electrons are localized. A transition in which the Fermi energy goes through E_m (as the concentration of electrons is decreased by reducing the gate voltage in a Si-MOSFET for instance) is called an Anderson transition. It is a specific example of the more general class of metal-insulator transitions.

There are two types of localization. These are strong localization, in which the electrons are spatially localized at impurity sites, and weak localization, in which the electrons are free to move, but their coherence is destroyed by the impurity potentials. The former is characterized by an exponential dependence of σ_{xx} on temperature, and the latter by a logarithmic dependence of σ_{xx} on temperature. Weak localization occurs at higher values of n_s and requires longer length scales than strong localization. Even a small magnetic field quickly destroys any possibility of weak localization, however, because the magnetic length quickly becomes the limiting length scale. Applied fields of tens or hundreds of Gauss (1 T = 10000 G) are sufficient to reduce the magnetic length below the electron mean free path which prevents weak localization. This criterion can be expressed as:

$$\left(\frac{\hbar}{eB}\right)^{\frac{1}{2}} < v_F \tau \tag{2.33}$$

There are also two commonly recognized methods by which the electrons can become localized. Anderson localization (also known as single particle localization) is a single electron effect in which the potential wells created by each impurity are assumed to have different, uncorrelated depths. When two potential wells are unequal, it is impossible for an electron to form a symmetric wave function which is shared equally between them. Thus the electron is localized near the well created by one of the impurity sites. It is still possible, however, for electrons to be partially shared between wells which are close in energy if they are also sufficiently close in space. The positions of those wells which can share an electron state are called resonant sites, and a group of connected resonant sites is called a cluster. The Anderson transition from an insulator to a metal occurs when one cluster becomes infinite in size

and there is an extended state which includes the entire sample. Since most of the electron states are still localized when this occurs, it is common to use percolation theory to describe the conduction near the transition. This theory portrays the potential wells as a series of valleys surrounded by hills of higher potential with the electrons as lakes of charge in the valleys. At a specific energy, known as the percolation energy, the lakes will connect and there will be a single coastline which extends throughout the sample. This coastline is an extended state and so the percolation energy is equivalent to the mobility edge.

Mott localization, the second major type, includes electron-electron effects but assumes that the impurities are regularly spaced and are all of approximately the same strength, so that they form their own impurity subband. In this case, the electrons can adjust to the periodic impurity potential, just as they adjust to the potential of the underlying silicon lattice. The localization effect occurs because the Coulomb repulsion between any pair of electrons makes it unfavourable for them to come together at the same impurity site. Thus there is an energy gap with very low DOS when the impurity band is half filled (i.e. $n_s = n_i$). If the Fermi energy is in this region the sample will be localized. In most two dimensional electron systems, however, the impurity potentials vary enough that a well defined impurity band with a Coulomb gap does not form. Electron-electron interactions can still play a role in Anderson localization by creating a Coulomb gap in the DOS near the Fermi level, which affects the conductivity even in the localized regime and changes the temperature dependence of σ_{xx}.

It has been suggested [Abr79] that at zero magnetic field an infinite, zero temperature, two dimensional electron system will always be localized. Any experimentally observed metal-insulator transition is then assumed to be a transition between a weakly localized and strongly localized state. Although experiments actually take place at finite temperatures on finite size samples, the experiments discussed in this thesis occur at low enough temperatures ($T < 1\,\mathrm{K}$) and in large enough length samples ($L > 1\,\mathrm{mm}$) that this theory still applies. More recent theoretical calculations for short range scatterers [Azb92] have revealed that a true metallic state is possible even with $B = 0$ and $T = 0$ provided that the Fermi energy lies above a mobility edge. Since most of the low temperature scattering in the inversion layer of a Si-MOSFET is short range scattering, either from impurity ions located close to the inversion layer (for low electron densities), or from surface defects at the semiconductor oxide interface (for higher electron densities), a true metallic state should be possible provided n_s is large enough that E_F lies above the mobility edge.

This brings up the point that there are distinct types of scattering which affect the electronic characteristics of a semiconductor. It is important to understand the differences between the typical scattering mechanisms found in GaAs/AlGaAs heterostructures and those found in Si-MOSFETs (electrons on liquid helium will not be considered in this discussion because this system has so few impurities) in order to determine the impurity concentration in a given sample. The first consideration is how each semiconductor based 2DES is designed. In a p-doped Si-MOSFET, electrons are attracted to the inversion layer at the semiconductor/oxide interface using the gate potential so that electrons from a low concentration of n-type dopants over a volume of the semiconductor can be concentrated in the relatively narrow width of the inversion layer. The main source of impurities is not the intentionally introduced dopants in the bulk of the silicon, but impurity ions which diffused into the MOSFET during its fabrication. These can be as close to the 2DES as the width of the triangular well (ie. $45 - 80\,\mathring{A}$ in a Si-MOSFET) and each one acts as a point potential at that distance, which means that the scattering is mostly short range. The other major source of scattering is through surface defects at the interface between the silicon and the oxide (surface roughness scattering) and since this is also very close to the 2DES, each defect again acts as a point scatterer producing short range scattering. At low temperatures ($T < 1\,\mathrm{K}$), impurity scattering dominates at low electron densities and surface roughness scattering at high electron densities. In GaAs/AlGaAs heterostructures the electrons come from an n-doped layer which is separated from the 2DES by an undoped spacer layer. The most important contribution to the scattering comes from the n-type impurities in this layer. Since the spacer layer is typically $500\,\mathring{A}$, any potential fluctuations on a length scale of less than $500\,\mathring{A}$ will be averaged out before they reach the 2DES. This means that the potential variations in a GaAs/AlGaAs heterostructure occur over longer distances than those in Si-MOSFETs at either high or low electron densities and thus there is more short range scattering in a Si-MOSFET than in a GaAs/AlGaAs heterostructure.

The next consideration is the influence of impurities on the transport mobility, μ_t. In the Drude model (as discussed in [Ash76]), the electrons in the 2DES are treated as non-interacting particles in a gas which travel in straight lines except when they collide. The collision mechanism is not specified, but at low temperatures ($T < 1\,\mathrm{K}$) where thermal phonons are frozen out, the main source of scattering is collisions with impurities or defects. If every scattering event occurs over a wide angle so that it significantly changes the direction of the electron's motion, then the conductivity of the 2DES will be inversely proportional to the number of scattering events and

directly proportional to the mean time between scattering events, τ. Since these scattering events affect the transport characteristics of the 2DES, τ is generally designated using the subscript, t, for transport. Then the mobility is defined as:

$$\mu_t = \frac{e\tau_t}{m^*} = \frac{\sigma_{xx}}{en_s} \tag{2.34}$$

An experimental measure of μ_t can thus be used to calculate τ_t and from this, the density of impurities affecting the 2DES. Detailed calculations for using this technique on a Si-MOSFET have been published by A. Gold [Gol85, Gol86a].

Both scattering due to Coulombic impurities and due to surface roughness are accounted for in Gold's calculations. The former depends principally on the density of Coulombic impurities, n_i, whereas the latter depends on the height of the surface roughness, Δ, and the mean distance between surface defects, Λ. Using experimental estimates of Δ and Λ, Gold obtained an expression for the scattering rate, and thus the mobility, which only contained a single unknown, the impurity density, n_i. Setting this equal to the experimentally determined transport mobility discussed above allowed him to solve for n_i. He noted that the results obtained using this method were reliable for all the Si-MOSFETs he studied.

Some papers [Har85, Sar85, Col91] have suggested, however, that small angle scattering may play an important role, particularly in GaAs/AlGaAs heterostructures. Small angle scattering events have little effect on the electrons' drift velocity because they do not significantly change the direction of electron motion. This means that they do not influence μ_t, but they are important in calculating the collision broadening of the single particle energy levels and thus the many physical quantities which are related to the collision broadened density of states. In order to account for all scattering events, it is necessary to use the quantum lifetime, τ_q, which is similar to τ_t, except that it weights scattering events at any angle equally, whereas τ_t has a weighting factor of $(1 - \cos\theta)$ (where θ is the scattering angle), which reduces the contribution of small angle scattering in the calculation of the scattering probability. The temperature variation of the magnitude of the Shuubnikov-de Haas resistance peaks can be used to calculate τ_q, as discussed in the previous section. Experiments on a variety of samples have revealed that the ratio of τ_q to τ_t is close to unity for Si-MOSFETs, but varies between 5 and 100 for GaAs/AlGaAs heterostructures. This is because the scattering in Si-MOSFETs is mostly short range (and therefore more likely to be wide angle), whereas the scattering in high quality GaAs/AlGaAs heterostructures with spacer layers is mostly long range, small angle scattering. Since

Gold's calculations assume that all the Coulombic scattering is wide angle, this makes them unsuitable for use with GaAs/AlGaAs heterostructures but ideal for Si-MOSFETs.

2.3.2 Magnetic Field Effects

It was previously noted that a magnetic field applied perpendicular to the plane of the 2DES will destroy any weak localization by reducing the magnetic length below the electron mean free path. At higher magnetic fields where the Landau levels are resolved, the density of states varies with magnetic field. This can cause the oscillations in resistivity known as the Shubnikov-de Haas effect, but if the oscillations in the DOS are large enough it may also allow the mobility edge to pass down through the Fermi energy and the magnetic field may then produce a transition from an insulator back to a metallic state. This has been experimentally observed in a low mobility 2DES, which was a strong insulator at zero magnetic field but showed a distinct IQHE ρ_{xx} minimum at $\nu = 2$ [Jia94]. Thus a magnetic field can destroy both weak and strong localization.

A perpendicular magnetic field can also induce a metal-insulator transition and create a localized state. Increasing the magnetic field tends to increase the impurity binding energy and so the effects of any existing impurities in a 2DES will be enhanced as the magnetic field increases and more and more electrons are localized by the impurity potential. This phenomenon is called magnetic freeze out, because the magnetic field essentially freezes the electrons in place by causing them to become localized. After the electrons have been localized, further increases in the magnetic field serve only to increase the impurity binding energies even more and so magnetic freeze out is characterized by a conductivity which starts to decrease at some specific value of the magnetic field and then continues to decrease rapidly as B is increased. It is also true that the higher the impurity density, the lower the magnetic field at which magnetic freeze out will occur. Finally, since increasing the magnetic field continues to increase the binding energy, this is one form of insulator which can not be removed by increasing the field strength.

2.3.3 Characterizing Anderson Localization

Since much of the experimental work discussed in this thesis involves insulating states, it is useful to describe the characteristics which are associated with Anderson localized states. This will help to determine whether or not

a given insulator is caused by single particle localization or by some other effect.

One characteristic of every type of strong localization mentioned so far (including Anderson localization) is the existence of a mobility edge. This is simply an energy gap between the Fermi energy and the energy which is required for an electron to be in an extended state where it can contribute to conduction. Thermal energy causes the electrons to have a thermodynamic distribution of energies and as the temperature is increased, more electrons will move into extended states above the mobility edge. The net result is a temperature activated conductivity. The initial work describing this type of conduction was performed in the 1970's [Pep74, Mot75, Pep75], but more modern reviews are also available [Shk84, Kas87]. The general agreement is that the conductivity will be of the form:

$$\sigma_{xx} \propto \exp\left(\frac{-\Delta}{T}\right) \tag{2.35}$$

Here $\Delta/2$ is the magnitude of the energy gap (measured in Kelvin) between E_F and E_m. It should be noted, however, that any other conduction mechanism in which electrons must overcome an energy gap in order to reach the conducting states will also exhibit this thermally activated behaviour.

At lower temperatures, where very few electrons are above the mobility edge there is a more specific prediction. Electrons can tunnel from one resonant site to another site at a different location within the sample which has a similar energy. A series of tunneling events allows the electron to move from one end of the sample to the other and thus contribute to conduction. Since the distance between the resonant sites may vary from one tunneling event to the next, this process is called variable range hopping. The temperature dependence of this conduction mechanism is called the Mott law and in two dimensions it is given by:

$$\sigma_{xx} \propto \exp\left(\frac{-T_0}{T}\right)^{1/3} \tag{2.36}$$

where T_0 is a constant which varies from sample to sample. This equation was developed by assuming that the density of states is constant near the Fermi level, and the unusual exponent of 1/3 comes from the assumption that only states near the Fermi level can contribute, so the average hopping distance is related to the distance between states near the Fermi energy. From the previous discussion of Mott localization, it is apparent that Coulomb interactions can reduce the DOS and create a Coulomb gap under the right

conditions. This requires a modification of the Mott law [Efr75] so that:

$$\sigma_{xx} \propto \exp\left(\frac{-T_1}{T}\right)^{1/2} \tag{2.37}$$

The precise conditions under which each of these equations apply are hard to quantify, but if either one of these temperature dependences is observed experimentally then it is an indicator that the 2DES is Anderson localized and that conduction is taking place through variable range hopping.

Another key signature of localization is that the electron wavefunctions are, of course, localized. If some experimental measure can be taken of the average extent of a localized electron's wave function, then this localization length, ξ, will increase as the Fermi energy comes closer to the mobility edge and the average cluster size increases. Percolation theory defines this energy dependence as:

$$\xi \propto |E - E_m|^{-\beta} \propto |n_s - n_c|^{-\beta} \tag{2.38}$$

where n_c is the critical electron density at which percolation begins and β is a universal exponent. For a two dimensional system, $\beta = 4/3$ for classical percolation and $\beta = 7/3$ if quantum tunneling is permitted (i.e. in the variable range hopping regime).

The temperature dependence of σ_{xx} and the electron density dependence of ξ are two characteristics of Anderson localization. The fact that an Anderson localized 2DES should be an insulator with $\sigma_{xx} \to 0$ as $T \to 0$ is another. This means that the diagonal resistivity, $\rho_{xx} = \sigma_{xx}/(\sigma_{xx}^2 + \sigma_{xy}^2)$, will become infinite at zero temperature. It has also been generally accepted that under a magnetic field the Hall resistivity, $\rho_{xy} = \sigma_{xy}/(\sigma_{xx}^2 + \sigma_{xy}^2)$, of an Anderson localized state will become infinite at zero temperature. Recently, however, a new type of Anderson localized insulator, called the Hall insulator, has been proposed [Kiv92, Fer93]. It exists between certain quantum Hall states and a global phase diagram has been suggested that determines which transitions from the Hall insulator to a quantum Hall state are allowed. This model is quite different from the Anderson localized states discussed previously in that as $T \to 0$, $\sigma_{xx} \to 0$ and $\sigma_{xy} \to 0$ in such a way that $\sigma_{xy} \propto (\sigma_{xx})^2$. This allows $\rho_{xx} \to \infty$ while ρ_{xy} remains finite. In fact, ρ_{xy} retains its classical value of B/en_s, leading to the novel concept of a 2DES which is localized and shows very large ρ_{xx} except at certain filling factors associated with the QHE while simultaneously exhibiting a ρ_{xy} curve which follows the classical B/en_s line interrupted by QHE plateaus.

2.4 Charge Density Waves

2.4.1 Definitions and Descriptions

A charge density wave (CDW) is a system in which the electronic charge has an ordered spatial variation in density as given by the equation:

$$\rho(\vec{r}) = \rho_0 + \rho_1 \cos(2\vec{k}_F \cdot \vec{r} + \Phi) \tag{2.39}$$

Here $\rho(\vec{r})$ is the charge density as a function of spatial position $\vec{r}$, $\vec{k}_F$ is the wave vector of the CDW, and Φ is its phase. The two dimensional Wigner crystal (the subject of this thesis) consists of electrons which are confined to specific sites by their mutual Coulomb repulsion. Since it involves charged electrons in an ordered lattice, the Wigner crystal can be considered as a specific form of two dimensional charge density wave. Thus, an understanding of charge density waves and their characteristic experimental signatures is useful in interpreting the results of experiments in the Wigner crystal regime.

The phenomenon of charge density waves has been studied mostly in one-dimensional inorganic chain compounds such as $NbSe_3$ [Fle79], but some work has also occurred in two-dimensional systems using inorganic layer compounds or electrons on liquid helium [Jia90a]. An excellent review of the field was given by G. Gruner [Gru88]. In the CDW state, the electron lattice is usually incommensurate with the atomic lattice of the solid which contains the CDW. This means that the positions of maximum electron density are not correlated with the positions of the atoms in the surrounding solid. The CDW is affected by impurity potentials in the atomic lattice, however, which serve to pin it so that the phase is fixed. This leads to a gapless phase excitation mode in the phonon spectrum, in contrast to the amplitude mode which has a gap.

2.4.2 Transport Properties

The charge density wave has several clearly defined electronic transport properties which have been experimentally established [Fle79, Zet82, Gru88, Nad92]. These include a distinct melting temperature, a thermally activated conductivity, non-linear I-V curves with a threshold electric field, and unusual noise characteristics in the region beyond the threshold field. All of these properties can be incorporated into the theoretical models which are used to describe charge transport in a CDW.

The first transport property is a melting temperature, T_m, above which none of the characteristic CDW transport properties can be observed. This

temperature is usually determined by a sharp increase in the exponent from the exponential dependence of ρ_{xx} on T as the temperature decreases below T_M. The fact that the CDW state is characterized by a higher resistivity than the electron gas state is important. It is generally accepted that there are impurities present in every sample which serve to spatially pin the charge density wave. This situation is illustrated in Figure 2.11. A one-dimensional CDW is shown as a graph of negative charge density, $p(x) \propto \langle \Psi(x) \rangle^2$, where $\Psi(x)$ is the electron wavefunction, versus position. The wavefunction tails overlap so that the electron density is always positive, and in the absence of an external electric field the charge density varies sinusoidally (as given by Equation 2.39) over the coherence length of the CDW. A positively charged impurity is shown in the figure. It pins the central portion of the CDW in its electrostatic potential well. If an external field is applied the CDW will deform around the pinning centre, as shown by the dashed curve in Figure 2.11, but this deformation will not produce any current, because all of the electrons are still spatially constrained by the phase coherence of the deformed CDW. Any conduction which occurs in this regime has traditionally been expected to come from those few electrons which remain in the uncondensed electron gas phase [Gru88] which continues to exist even after most of the electrons have undergone the transition to a CDW. Since the density of electrons available for conduction is so low, the resistivity of this pinned CDW state is high.

If the electric field is increased beyond a critical threshold value, E_T, however, the entire CDW will become depinned [Fuk78] and start to slide coherently in the direction opposite to the electric field. Since all the electrons in the CDW are now contributing to the current, the differential resistivity will drop very sharply as this threshold field is exceeded. This is why the $I-V$ curves measured for charge density waves are strongly non-linear, with a high differential resistivity close to the origin which drops to a much lower value as soon as the applied voltage causes the electric field to exceed the critical value. Figure 2.11 only shows a small coherent portion of a one dimensional CDW, pinned by a single impurity. In a real system, there are many separately pinned coherent segments of the CDW, and since these may become depinned at slightly different threshold fields, this smoothes out the I-V non-linearity. As the temperature is lowered, the coherence length decreases and eventually phase coherence is lost [Zet82] so that the non-linearity becomes much more difficult to detect. This is particularly noticeable in samples with many impurities because the impurities contribute to a loss of phase coherence by locking the phase at specific values at the randomly placed impurity centres. A more detailed approach to the conduction mechanism for the CDW in the

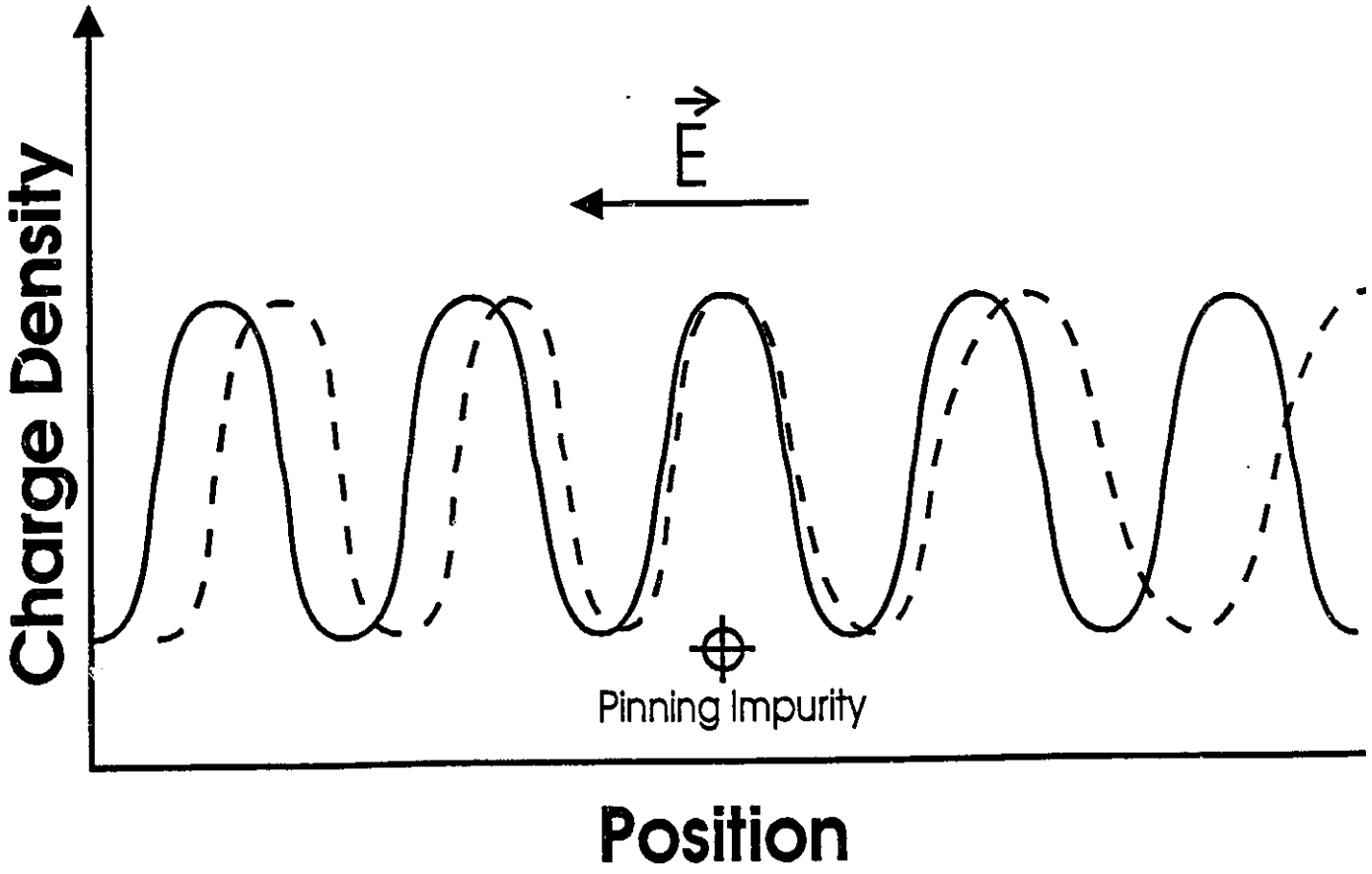

Figure 2.11: The negative charge density (ie. electron density) is plotted versus position for a one-dimensional CDW pinned by a positive impurity both with (dashed curve) and without (solid curve) an external applied field.

above threshold regime will be discussed in Section 2.5.4.

This pinned CDW model also explains why there is thermally activated conduction in the charge density wave state. The mixed state which contains both a CDW and an electron gas phase must have an energy gap between the CDW ground state and the conduction band in the electron gas. Electrons can be thermally excited across this gap to become available for conduction, even though the CDW itself remains pinned. Thus there should be a temperature activated conductivity with an activation energy equal to the width of this gap so that:

$$\sigma = \sigma_0 \exp\left(\frac{-E_g}{k_B T}\right) \tag{2.40}$$

At temperatures well below T_m, it is possible that the electron gas phase will be frozen out and only the CDW will remain. The conduction is still thermally activated in this regime, but the activation energy is now the energy required to change the phase of the local CDW within a coherent segment by 2π, as given by the phase slippage conduction model [Nad92] (see Section 2.5.4).

Another feature of this simple model of a pinned CDW which is strained and then depinned by an external electric field is that it predicts a precise relationship between the current carried by the charge density wave and the frequency spectrum of the voltage noise produced by the periodic electron potential of the CDW sliding past stationary impurity potentials. In the one-dimensional case the current carried by the CDW is simply:

$$j_{CDW} = -en_s v_{CDW} = -\frac{e}{2\pi}\frac{d\phi}{dt} \tag{2.41}$$

where n_s is the concentration of electrons in the CDW, v_{CDW} is the sliding velocity of the CDW, and $\phi(x,t)$ is the phase of the charge density wave. The frequency of the narrow band noise is then:

$$f_0 = \left|\frac{v_{CDW}}{\lambda}\right| = \frac{j_{CDW}}{en_s\lambda} \tag{2.42}$$

where λ is the wavelength of the CDW over which the phase changes by 2π. When the threshold field is exceeded and the CDW begins to slide, a sequence of peaks can be observed in the Fourier transform of the potential measured between two different points on the one dimensional sample. At temperatures where the CDW is coherent, the fundamental frequency of these peaks corresponds exactly to that predicted by Equation 2.42 as the electron concentration and CDW current are varied, but at temperatures which are well below T_M, when the phase coherence length starts to decrease (below 100 K

in $NbSe_3$ for instance [Nad92]), the ratio of f_0 to j_{CDW} increases. There is also a large increase in the background noise level over a wide frequency band (up to MHz frequencies) as the threshold field is exceeded. This broad band noise does not exhibit the frequency structure predicted by Equation 2.42, and is thought to be created by the interference between noise generated in different phase coherent segments of the CDW. Thus, the observed frequency spectra are a superposition of the noise peaks at specific frequencies caused by the periodic nature of the CDW and the broad band noise caused by the phase differences between different microscopic segments of the CDW.

The size of the phase coherent segments is defined by the correlation length, L_D. In the two dimensional case, it is possible to determine L_D for a pinned CDW or Wigner crystal [Nor92] by using an experimental determination of the threshold electric field, E_T, combined with a theoretical estimate of the transverse shear modulus, K_T.

$$L_D^2 \approx \frac{K_T a}{2\pi\rho_c E_T} = \frac{K_T a}{2\pi e n_s E_T} \tag{2.43}$$

Here a is the average interelectron spacing in the 2DES and ρ_c is the average charge density ($\rho_c = n_s e$). There is a numerical prefactor in Equation 2.43, but it is generally close to unity [Nor92], so it can be ignored.

Although most of these transport properties have been discussed in terms of a one dimensional charge density wave, they also apply in the two dimensional case. Thus it is legitimate to expect that a two dimensional Wigner crystal will also show these characteristics.

2.5 The Wigner Crystal

2.5.1 Early Theories

In 1934, Eugene Wigner used perturbation methods to calculate the correlation energy due to the mutual repulsion between free electrons in an electron gas [Wig34]. He noted that this energy could be minimized if the electrons were organized in a close packed body centred lattice with interelectron spacing r. Wigner demonstrated that the correlation energy can be as high as $0.292e^2/r$ if the electrons' kinetic energy can be neglected and he realized that in a system where the electrons' kinetic energy was much smaller than $0.292e^2/r$, the system would settle into the ordered three dimensional lattice he had described. The question was whether or not such conditions could exist in any real system. At zero temperature the zero point oscillation motion of an electron (the only kinetic energy left available at zero temperature)

gives it an energy which varies as r^{-2}. Since the correlation energy varies as r^{-1}, it is obvious that if the electron density is lowered sufficiently, the correlation energy will become the dominant factor and the electrons should freeze into what is now known as a Wigner crystal.

Later the same theory was applied to two dimensional systems of electrons using the specific examples of electrons suspended just above the surface of liquid Helium by a perpendicular electric field [Cra71] and of electrons or holes in a semiconductor inversion layer [Cha72]. The theory of these two dimensional Wigner crystals was extensively developed over the next few years. L. Bonsall and A. Maradudin examined the static ground state energy for the five possible two dimensional Bravais lattices and determined that the minimum energy configuration was that of a hexagonal (or triangular) lattice [Bon76]. Once the Bravais lattice is known, it is possible to explicitly calculate the dispersion relationship for the phonon modes in this lattice. Bonsall and Maradudin carried out this calculation [Bon76, Bon77] and demonstrated that, in contrast to the three dimensional Wigner crystal, the phonon frequencies approached zero as the wave vectors approached zero (i.e. there were no gaps in the dispersion curves). Other groups attempted to determine the exact value of r (and the corresponding electron density) at which the electron gas would freeze into a Wigner crystal. The calculations all used the same basic principle which had been suggested by Wigner, but the detailed calculations of the correlation energy used different methods, such as variational wave function treatments or Monte-Carlo simulations.

There are two fundamental elements to any calculation of this nature. The first is the average potential energy per electron, $< U >$, in a 2DES due to the mutual Coulomb repulsion. This can be expressed in cgs units (which are traditionally used in theoretical discussions of the Wigner crystal and are also consistent with the experimental convention of expressing electron densities in cm^{-2}) as:

$$< U > \approx \frac{e^2}{r_0 \kappa} = \frac{e^2 \sqrt{\pi n_s}}{\kappa} \tag{2.44}$$

where κ is the static dielectric constant in the system and r_0 is the average interelectron distance in the two dimensional plane which can be expressed in terms of the electron density, $n_s = (\pi r_0^2)^{-1}$. The second is the mean value of the electrons' largest kinetic energy component, $< K >$. In order for the crystal to form, $< U >$ must be larger than this competing energy component by some factor Γ_0 so that:

$$\frac{< U >}{< K >} = \Gamma_0 \tag{2.45}$$

is the equation which defines the phase boundary of the crystal. There are three different regimes where it is thought that the Wigner crystal can exist and each one of these is associated with a different competing energy component.

The first and simplest regime to consider is the limit of zero temperature, where the only source of kinetic energy is the zero point motion of the electrons due to the Heisenberg uncertainty principle. This is called the quantum regime. In the Wigner lattice the electrons have a position uncertainty of $\Delta x \approx r_0$, since their confinement to the positions in the triangular lattice means that their wave functions should not overlap. This leads to a momentum uncertainty of $\Delta\rho \approx \hbar/r_0$, where $\hbar = h/2\pi$. The momentum uncertainty means that the electrons have an intrinsic minimum kinetic energy caused by quantum fluctuations of:

$$< K >= \frac{\Delta\rho^2}{2m^*} = \frac{\hbar^2}{2r_0^2 m^*} = \frac{\pi n_s \hbar^2}{2m^*} \tag{2.46}$$

In this case $< U > \propto \sqrt{n_s}$ and $< K > \propto n_s$ so that if n_s is increased above some critical cold melting concentration, n_{CM}, the Wigner crystal will melt even at zero temperature. The specific value of n_{CM} in a given system can be determined by substituting Equations 2.46 and 2.44 into Equation 2.45, which reveals that:

$$n_{CM} = \left(\frac{2e^2 m^*}{\sqrt{\pi}\kappa\hbar^2\Gamma_0}\right)^2 \tag{2.47}$$

Figure 2.12 shows a general first approximation phase diagram for the Wigner crystal in the phase space of magnetic field, electron concentration and temperature. In this diagram, the quantum regime is the segment on the n_s axis from the origin to n_{CM}.

If the temperature is increased above zero then the zero point quantum motion will not be the only source of kinetic energy. In the high temperature limit, known as the classical regime, the largest source of kinetic energy is the thermodynamic kinetic energy:

$$< K >= k_B T \tag{2.48}$$

where k_B is Boltzmann's constant. In this regime the Wigner crystal will melt when the temperature is raised beyond some critical melting temperature, T_M, or when the density is lowered below some critical value, n_M. The relationship between these two numbers is obtained by substituting the specific equations for $< U >$ and $< K >$ into Equation 2.45, which reveals

that:

$$n_M = \frac{(\kappa k_B T_M \Gamma_0)^2}{\pi e^4} \tag{2.49}$$

This equation suggests that if the electron density is sufficiently high, then the Wigner crystal could actually exist at room temperature. This does not occur, however, because if n_s becomes too large then the kinetic energy which comes from the zero point quantum motion of the electrons will become larger than $< U > /\Gamma_0$ and the crystal will melt. In order to account for these circumstances when the temperature is not high enough to be in the simple classical regime defined by Equation 2.48 and not low enough to be in the $T \rightarrow 0$ quantum regime, P. M. Platzman and H. Fukuyama [Pla74] used the Fermi function to evaluate the mean kinetic energy per electron:

$$< K >= \frac{2}{n} \int \frac{d^2\rho}{(2\pi)^2} \frac{\rho^2/2m^*}{e^{\frac{\rho^2/2m^* - \mu}{k_B T}} + 1} \tag{2.50}$$

Here μ is the chemical potential energy (equivalent to the Fermi energy in an electron gas) and n is the number density which must also be defined using the Fermi function:

$$n = 2 \int \frac{d^2\rho}{(2\pi)^2} \frac{1}{e^{\frac{\rho^2/2m^* - \mu}{k_B T}} + 1} \tag{2.51}$$

These two equations can be evaluated for arbitrary temperature and electron density, and this leads to parametric equations for the phase boundary which allow it to be expressed in terms of a dimensionless temperature T/T_0, where $T_0 = 2e^4 m^*/k_B \hbar^2 \kappa^2 \Gamma_0^2$, and dimensionless electron density n_s/n_{CM}. The result is a curve like the one shown in Figure 2.12 in the (T, n_s) plane, which shows that there will always be some temperature, T_{M0}, beyond which the crystal will melt regardless of the value of n_s.

Now the third axis of the phase diagram will be considered. Figure 2.12 indicates that turning on a magnetic field tends to increase the stability of the Wigner crystal both with respect to temperature and electron density. This statement is a simplification of the complex physics which actually takes place as the magnetic field is turned on, but it is possible to easily describe the situation in the high magnetic field limit which is known as the extreme quantum regime. In this regime the temperature is again assumed to be close to zero and the magnetic field is high enough that the magnetic length is lower than r_0 and thus the cyclotron energy is the largest component of the electrons' kinetic energy. For these high magnetic fields, all electrons will be in the lowest Landau level and thus:

$$< K >= \hbar\omega_c(0 + \frac{1}{2}) = \frac{eB\hbar}{2m^*} \tag{2.52}$$

This leads to an expression for the critical magnetic field beyond which the Wigner crystal will solidify regardless of the electron concentration:

$$B_C = \frac{2em^*\sqrt{\pi n_s}}{\hbar\kappa\Gamma_0} \tag{2.53}$$

In real systems, however, the electron gas is not the ground state at very large magnetic fields because the FQHE becomes dominant. Since the energy of the Laughlin ground state has downwards cusps at specific fractional filling factors, the Wigner crystal may freeze and then melt near a fractional filling factor and then freeze again as the magnetic field is increased; a phenomenon known as reentrant behaviour. Alternatively, in lower quality samples, magnetic freeze out (as discussed in Section 2.3.2) may occur before B_C is reached so that neither the FQHE nor the Wigner crystal will be observed.

In all three regimes, the value of Γ_0 is critical to calculating the phase boundary for the Wigner crystal. Different melting mechanisms have been proposed for the Wigner crystal and used to determine Γ_0. Platzman and Fukuyama [Pla74], for instance, assume that the crystal melts through lattice vibrations and use a self-consistent harmonic approximation (SCHA) to calculate the phonon modes. When no self-consistent solution can be found for a transverse mode (it is assumed that only an electron solid can exhibit a transverse mode) the crystal is deemed to have melted. This means that the transition will be first order, in that the state variables which define the system will change discontinuously at the phase transition point, and it leads to a value of $\Gamma_0 = 3$ in the classical regime. In the quantum regime they use the same technique but choose, like most theorists, to express the result in terms of the critical Wigner-Seitz radius, r_{sc}, at which cold melting occurs. The Wigner-Seitz radius expresses the inter-electron distance in terms of the Bohr radius, $r_s = r_0/a_B = m^*e^2/(\hbar^2\kappa\sqrt{\pi n_s})$ and it has the advantage of being dimensionless and leading to a result for r_{sc} which is independent of system specific parameters. Substituting into Equation 2.47, for instance, gives:

$$r_{sc} = \frac{m^*e^2}{\hbar^2\kappa\sqrt{\pi n_{CM}}} = \frac{\Gamma_0}{2} \tag{2.54}$$

Platzman and Fukuyama calculate a value of $r_{sc} \approx 4.5$. This is of the same order of magnitude as $\Gamma_0/2 \approx 1.5$, but the factor of three difference indicates the lack of subtlety in the simple treatment which has been described here.

A wide range of values have been assigned to both r_{sc} (for the quantum regime) and Γ_0 (for the classical regime) by theorists, depending upon the calculation methods they used to solve the many body electron interaction

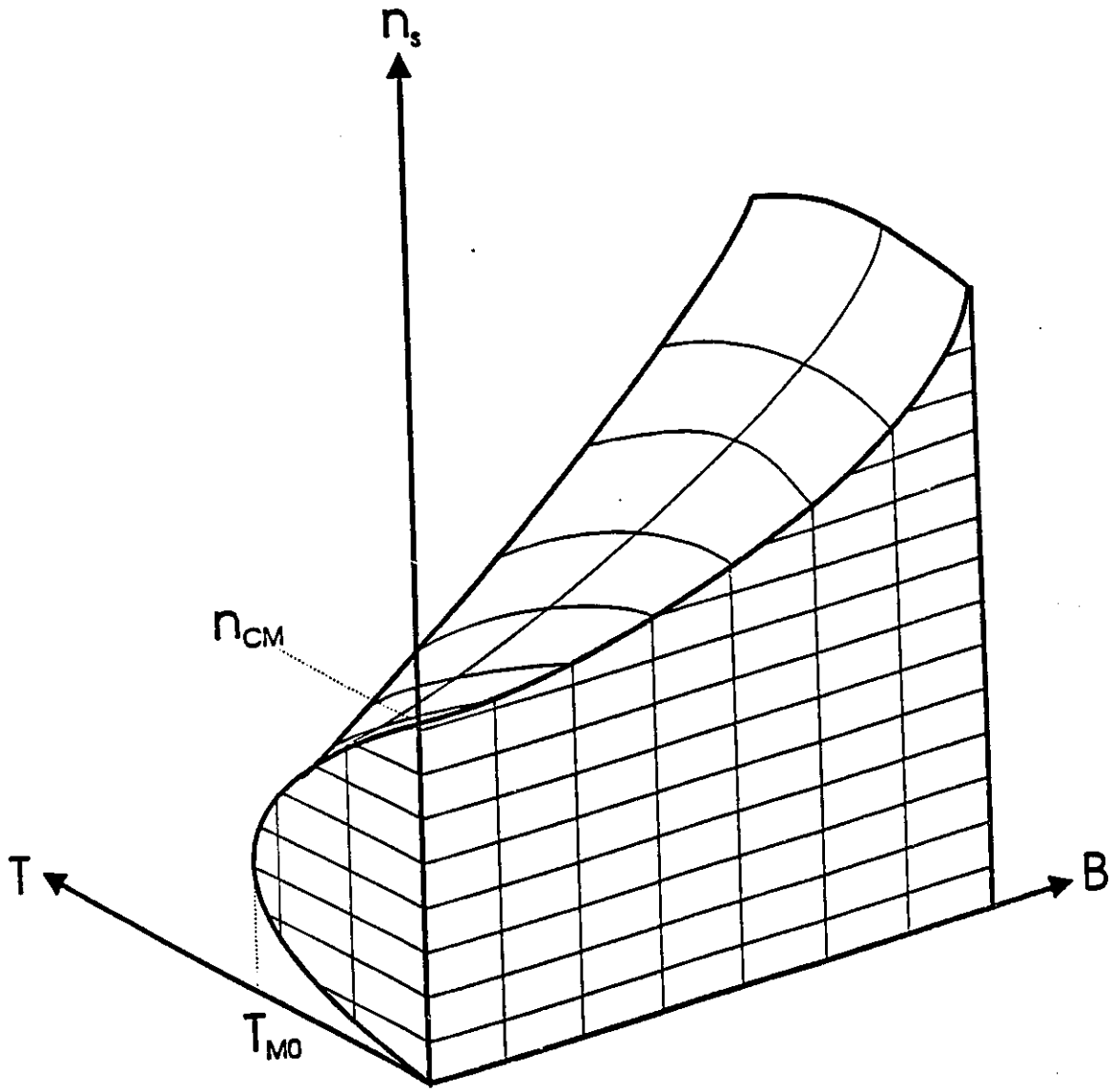

Figure 2.12: This phase diagram shows a plausible phase space in magnetic field, B, temperature, T, and electron density, n_s, in which the Wigner crystal may exist. The values of n_{CM} and T_{M0} indicated on the graph are arbitrary and serve only to indicate what point in the phase diagram these terms refer to.

problem and also upon the assumptions they made about the precise melting mechanism in the crystal. These vary from $r_{sc} = 4.5$ up to $r_{sc} = 33$ [Cep78, Ima84] or $r_{sc} = 37 \pm 5$ [Tan89], and from $\Gamma_0 = 3$ up to $\Gamma_0 = 78$ [Tho78] or $\Gamma_0 = 95$ [Hoc75]. In recent years the most widely accepted value for r_{sc} has been 37 ± 5, given by B. Tanatar and D. M. Ceperley [Tan89] in 1989. Their calculations assume zero temperature and magnetic field and that the sample is free of impurities. They use variational Monte Carlo methods and fixed node Green's function Monte Carlo methods to calculate the ground state energies of the crystal and of the electron liquid over a range of electron densities. This is, in principle, a very accurate method of determining r_s, but it does not account for the important effect which impurities in a real system can have on the local potential and thus it is not necessarily suitable for direct comparison with experiments. In contrast, the most commonly accepted value for Γ_0 comes directly from experimental work. In 1979 C. C. Grimes and G. Adams reported the observation of the sudden appearance with decreasing temperature of a radio frequency (rf) resonance in a sheet of electrons suspended at the surface of a layer of liquid helium [Gri79]. They attributed this resonance to a coupling of plasmon-ripplon modes excited by the rf electric field with the periodicity of the Wigner lattice. The temperature at which this occurred corresponded to $\Gamma_0 = 137 \pm 15$. Even today this number is still considered accurate. The value, $\Gamma_0 = 127 \pm 3$, is a more modern experimental value with a smaller error bar, but it still falls within the range defined by Grimes and Adams.

2.5.2 Experimental Work

The system of electrons at the interface between liquid helium and a vacuum was the first experimental system to exhibit strong evidence for a Wigner crystal. The electrons are trapped at the surface by the combination of their own image potential in the liquid which is a long range attractive force and the short range repulsive barrier which prevents them from entering the liquid. Their wavefunctions are typically confined to $\approx 10^{-6}$ cm in the direction normal to the liquid-vacuum interface, but they are free to move parallel to the interface so they are usually held in place by a metal guard ring which limits them to the small area of the liquid helium which is being studied. Since liquid helium has very few impurities, this is almost an ideal system. The electron density can be controlled using a filament and a parallel plate capacitor to be anywhere between 10^5cm^{-2} and 10^{10}cm^{-2}. This fulfills the low electron density requirement for observing the Wigner crystal in the classical regime but falls far short of the densities required to be in the quantum

regime. Higher electron densities cause a surface instability which allows electrons to escape to either the lower electrode which is placed below the surface of the liquid helium to attract electrons from the filament, or to the guard ring electrode [Wil71, Gor73]. The rf resonances and corresponding absorption peaks observed by Grimes and Adams for electrons above the surface of liquid helium have also been observed for positive ions trapped beneath the surface [Mel90]. Both data sets can be interpreted as showing the periodicity of a triangular lattice, which is convincing evidence for the existence of a Wigner crystal. Other studies have examined shear waves in the Wigner crystal [Dev84] or have used thin films of helium on top of a rough substrate (such as glass) to investigate the effects of impurities on the crystal [Jia89, Jia90a]. Electronic transport characteristics such as resistivity as a function of temperature, current as a function of applied voltage ($I-V$ curves) and voltage noise amplitude as a function of applied DC voltage were measured. These studies have shown that the crystal displays the electronic transport characteristics of a two dimensional charge density wave (CDW) which is pinned by impurities, as discussed in the previous section. Since a Wigner crystal is an ordered array of charged electrons, it can be described in terms of a pinned two dimensional CDW, and thus it is expected to display the characteristic signatures of low conductivity, a distinct melting temperature, non-linear I-V curves with a threshold electric field, temperature activated conduction, and both broad band and narrow band noise in the region above the threshold field where the crystal began to slide coherently. This is exactly what was observed. Figure 2.13, for example, which comes from [Jia90a], gives an example of both temperature melting of the crystal and (in the insert) an R-V curve (similar to an I-V curve) with a threshold electric field.

When high quality semiconductor based 2DESs became available they were also good candidates to show the two dimensional Wigner crystal. Si-MOSFETs were initially the subject of most studies, but during the 1980's the focus shifted to GaAs/AlGaAs heterostructures. In contrast to the case of electrons on liquid helium, both of the semiconductor systems could easily be made with high enough electron densities to exist in the quantum rather than the classical regime. In fact the problem turned out to be that their typical electron densities were too high, and lowering them to the theoretical value of n_{CM} resulted in all the electrons being localized by impurity potentials before the Wigner crystal formed. The differences between the three systems which have been mentioned so far are highlighted in Table 2.2. It lists some physical quantities for each system as well as n_{sc} (calculated using both $r_{sc} = 10$ and $r_{sc} = 37$) and T_M (calculated using n_{st}, a typical electron density at which

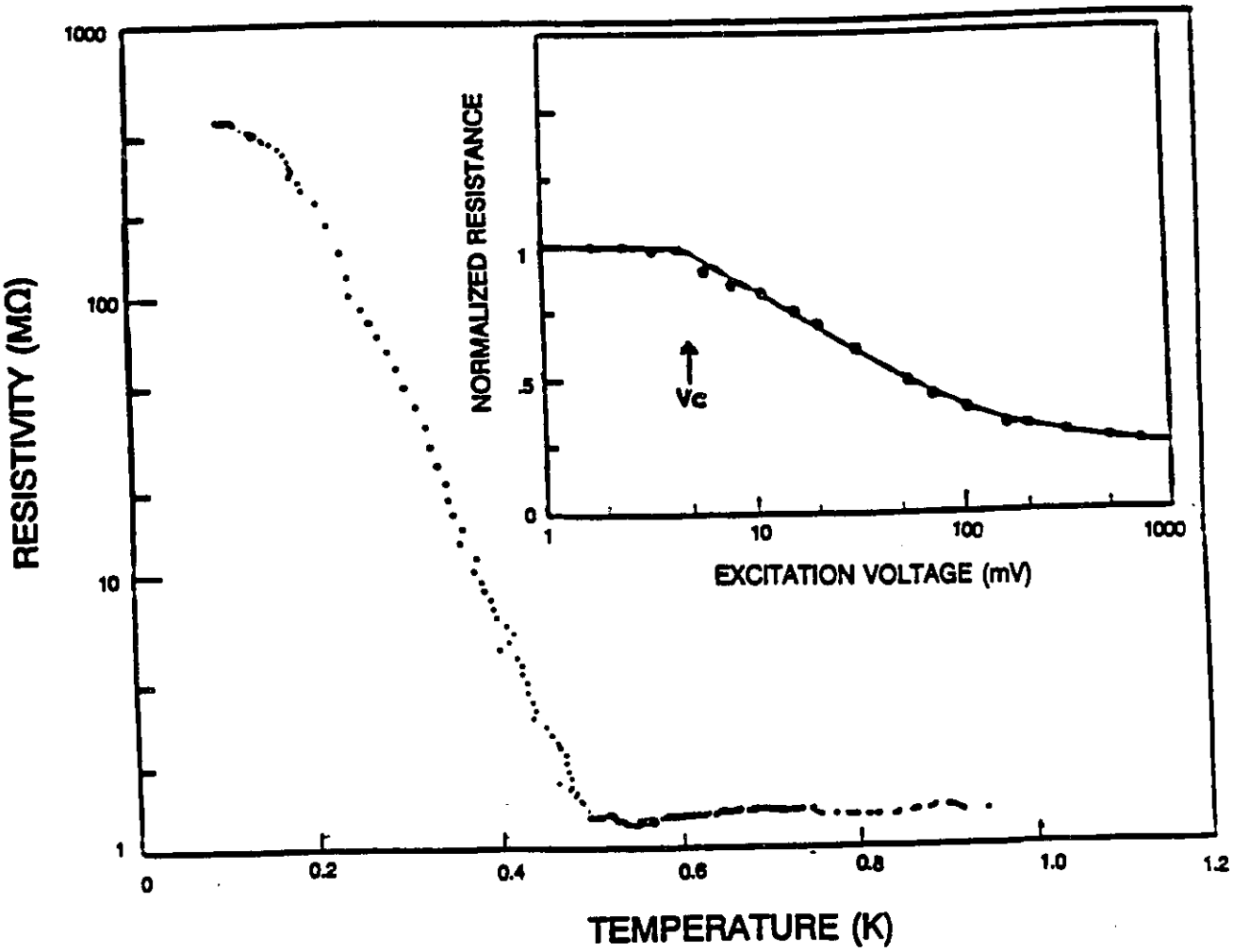

Figure 2.13: This graph of ρ_{xx} versus T shows the temperature melting of the Wigner crystal. The inset shows the critical potential (i.e. threshold field) at which sliding begins. This data is from electrons above a thin film of liquid helium [Jia90a].

Parameter	Si-MOSFET	GaAs/$Al_{.33}Ga_{.67}As$	Electrons on Helium
m^*	0.19	0.067	1.0
κ	7.7	13	1.0
a_B (Å)	21	104	0.53
n_{st} (cm^{-2})	8.0×10^{10}	8.0×10^{10}	4.0×10^{9}
n_{CM} (cm^{-2}) ($r_{sc} = 37$)	5.1×10^{9}	2.1×10^{8}	8.3×10^{12}
n_{CM} (cm^{-2}) ($r_{sc} = 10$)	6.9×10^{10}	3.0×10^{9}	1.1×10^{14}
T_M (K) ($\Gamma = 127$)	0.86	0.50	1.5

Table 2.2: Parameters Relevant to A Wigner Crystal in Three Systems

experiments to detect a Wigner crystal have been performed in each system, and $\Gamma_0 = 127$). The value of $r_{sc} = 10$ comes from a specific calculation which accounts for the impurities and valley splitting in silicon which will be discussed later. This table reveals that theoretically the Wigner crystal will solidify at a higher density and at higher temperatures in a system of electrons on liquid helium than in either of the semiconductor systems, but that the Si-MOSFET system is the more favourable of the two semiconductors. This is particularly noticeable when comparing the experimentally studied densities to the cold melting densities. For $r_{sc} = 10$ in silicon, n_{CM} is very close to a typical experimentally attainable density of $n_{st} = 8 \times 10^{10}$ cm^{-2}. In GaAs/AlGaAs, however, where $r_{sc} = 37$ is a more appropriate value, n_{CM} is over two orders of magnitude lower than the usual experimental density regime, and even with $r_{sc} = 10$, experimental densities must still be lowered by an order of magnitude. This is why most studies using GaAs/AlGaAs have focused on the extreme quantum regime, where the cyclotron energy is lower than the zero-point motion energy, and thus cold melting will occur at higher electron densities. The table also explains why the cold melting transition in the quantum regime was inacessible using electrons on liquid helium where the maximum electron density was on the order of $10^{10}\,cm^{-2}$.

Since its larger effective mass and smaller dielectric constant allow silicon to have a higher value of n_{CM} than GaAs/AlGaAs for the same theoretical value of r_{sc}, it seems strange that the focus in the search for a Wigner crystal moved from the former system to the latter. This makes more sense, however, when the electronic mobilities, μ_t of the two systems are compared. Few Si-MOSFETs have $\mu_t > 10^4\,cm^2/Vs$, whereas modern GaAs/AlGaAs heterostuctures often have $\mu_t > 10^6\,cm^2/Vs$. Higher mobility is ususally associated with a decrease in the influence of impurity potentials. Reducing

the influence of impurities makes it possible to attain lower values of n_{st} before all the electrons are lost to single particle localization and the formation of any correlated electron state becomes impossible. The value of $n_{st} = 8.0 \times 10^{10}\,\mathrm{cm}^{-2}$ which is given for silicon in Table 2.2 is for the remarkably high quality Si-MOSFETs which were used in this work. Thus, until the experiments described in this thesis began, no strong evidence had been observed for the Wigner crystal existing in Si-MOSFETs and those who were interested in finding the Wigner crystal in a semiconductor based 2DES were studying GaAs/AlGaAs heterostructures in the extreme quantum limit using high magnetic fields.

The first evidence for the existence of a Wigner crystal in GaAs/AlGaAs heterostructures came from radio-frequency spectroscopy measurements made in 1988 [And88]. These measurements relied upon the fact that a Wigner lattice (because of its rigidity to shear) has a gapless magnetophonon mode, whereas an electron liquid does not. The frequency versus wavevector relationship for this magnetophonon branch can be expressed as:

$$\omega_- = (2\pi\mu_s)^{0.5} q^{1.5} B^{-1} \tag{2.55}$$

where μ_s is the static shear modulus of the crystal and q is the wavevector of the magnetophonon. An rf electromagnetic wave of wavevector q and frequency ω introduced in the plane of the electrons will couple to the magnetophonon mode when the two frequencies match (i.e. $\omega = \omega_-(q)$). This can be detected by a decrease in amplitude of the transmitted electromagnetic wave at those frequencies which are harmonics of $\omega_-(q)$. Since this initial set of measurements, other experimental work using different techniques has taken place. These have included all the electronic transport measurements, using both AC and DC techniques [Wil88, Gol90, Jia90b, Jia91, Wil91, Paa92, Gol93], which were carried out first on charge density waves and then on the classical Wigner crystal. Although the distinct noise frequency spectrum associated with a perfect lattice or CDW was not directly observed, all of the other characteristics of a CDW or classical Wigner crystal were seen. Figure 2.14, for instance, shows a plot of R_{xx} on a logarithmic scale versus T^{-1} on a linear scale (an Arrhenius plot) using data taken from a GaAs/AlGaAs heterostructure at high magnetic field [Jia90b]. The linear fit indicates that R_{xx} has an exponential dependence on T^{-1} and thus fits the temperature activated conduction model which has previously been associated with a pinned CDW. This figure also conveniently shows the influence of the magnetic field. At filling factor $\nu = 0.21$ the temperature dependence of R_{xx} is strong, whereas at $\nu = 0.302$, the temperature dependence is much weaker (i.e. the activation energy is lower). This is typical of the behaviour

observed in the extreme quantum limit. The Wigner crystal exists only at very high magnetic fields starting just below $\nu = 1/5$, so the decrease in intensity of the characteristic signatures at higher filling factors is expected. These experiments have also shown, however, that the $\nu = 1/5$ quantum Hall effect plateau is still observed if the magnetic field is increased slightly, but that the strongly insulating behaviour associated with the crystal becomes dominant as the magnetic field increases beyond $\nu = 1/5$. This demonstrates that, in accordance with the fact that the Laughlin wave function has a lower energy ground state at specific fractional filling factors, the Wigner crystal is actually reentrant with the Laughlin wave function of the QHE at $\nu = 1/5$, melting and then recrystallizing as the magnetic field is increased. The densities at which the Wigner crystal forms near $\nu = 1/5$ correspond to $r_{sc} \approx 2$ [And88], indicating the stabilizing effect of the magnetic field.

The two other techniques used have been photoluminescence [Buh91, Tur92, Gol92] measurements, in which an additional line has been observed in the luminescence spectrum in the region of the phase diagram where the Wigner solid is believed to exist, and cyclotron resonance [Bes92, Sum93], in which long wavelength cyclotron excitations have also shown new absorptions which indicate a change in the electron ground state coincident with the same region in phase space. Since photoluminescence measurements are sensitive to the energy spectrum of electrons in the 2DES, and it can be demonstrated [Fer93] that this spectrum is unique for electrons in a periodic potential (ie. in the crystallized state) subject to a magnetic field, the appearance of new lines in the magnetoluminescence spectrum provides convincing evidence for the formation of a Wigner crystal. In fact one group [Kuk94] has actually used time resolved luminescence measurements to probe the precise structure of the periodic potential and claims that it has the form of an ideal triangular lattice, just as is predicted for a Wigner crystal.

It was about two years after the initial discovery of the Wigner crystal in GaAs/AlGaAs heterostructures that it was also observed in Si-MOSFETs. These experiments took place in the quantum regime using electronic transport techniques on extremely high mobility Si-MOSFETs. An unusually strong insulating behaviour was noticed when measuring R_{xx} versus B in these MOSFETs at low temperatures and electron densities. This phenomenon was unusual because the insulator was reentrant with the IQHE R_{xx} minima at $\nu = 1$ and $\nu = 2$, and the peak resistance could change by up to seven orders of magnitude with as little as a 30% change in n_s. Further experiments used a variety of different electronic transport measurements, including AC and DC conductivity [DIo92a, DIo92b, Pud93a], Hall measurements [Pud93b], and noise spectra [Pud94a]. All of these pointed to

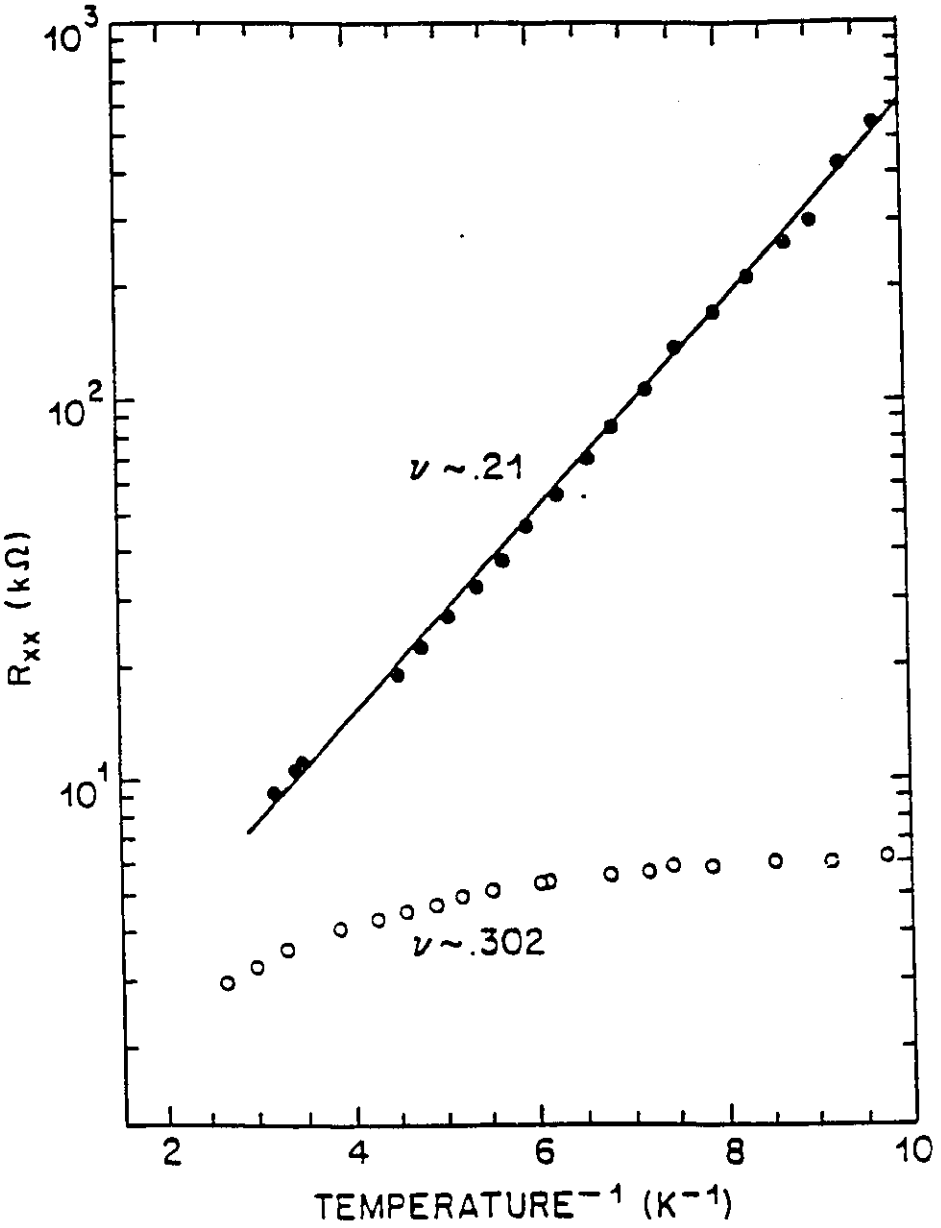

Figure 2.14: The data points (measured from a GaAs/AlGaAs heterostructure in the extreme quantum limit [Jia90b]) show a temperature activated conductivity at $\nu \approx 0.21$. The dependence of the phenomenon on magnetic field is indicated by the much weaker behaviour at the lower field of $\nu \approx 0.302$.

the existence of a Wigner crystal under the right conditions in these very high quality MOSFETs. It is only very recently that Si-MOSFETs were also used to study the extreme quantum limit [Dol94]. The basic nature of the experiments performed on Si-MOSFETs has been similar to those performed on CDWs, electrons on liquid helium and GaAs/AlGaAs heterostructures. The differences are in how the data are used to determine different trends and how they are then compared to theoretical predictions.

2.5.3 Models: Ideal to Real

The simple predictions made using the CDW analogy and even those made for the rf spectroscopy, photoluminescence or cyclotron resonance methods have generally assumed that the Wigner crystal can be treated as an ideal system. Most theoretical predictions of r_{sc} and T_{M0} have also used assumptions which are difficult to apply to real systems. The effect of non-zero temperature should be taken into account when comparing an experimental n_{sc} with a theoretical r_{sc}, and the role of disorder in the sample is also critical. The fact that the Wigner crystal is assumed to be pinned by impurities when studying its electronic transport properties means that experimentalists have implicitly assumed that disorder plays some important role, but its precise effects on the crystal have seldom been quantified. It is particularly important that the differences between disorder in the GaAs/AlGaAs heterostructures (where the introduction of a spacer layer limits the short range potential fluctuations) and the Si-MOSFETs (where the ionic impurities and surface roughness defects are close to the 2DES and the short range potential fluctuations may be very large) be accounted for. A better understanding of the effects of spin and valley splitting on the Wigner crystal would also be useful.

Some attempts have been made to include these effects. Platzman and Fukuyama [Pla74] were able to calculate the melting temperature corresponding to any electron density, as described above, but their calculations did not include disorder. In 1983, A. G. Eguiluz, A. A. Maradudin and R. J. Elliott considered a two dimensional Wigner crystal in the presence of a random array of pinning centers [Egu83]. They were able to calculate the spectral density for the phonons of the electron lattice both with and without a magnetic field and concluded that at low temperatures ($T < 1\,\mathrm{K}$), the presence of the pinning centers helped to stabilize the crystal by lowering the mean square displacement of electrons in the lattice from their equilibrium positions which could cause the crystal to melt. They assumed a uniform random distribution of pinning centers, however, and this is not the case in a real

sample where the pinning centers are associated with impurities which are unlikely to have an even distribution throughout the sample. More recent calculations of the phonon dispersion curves for a two dimensional Wigner crystal have attempted to use more detailed models of the pinning potential [Nor92, Ale94], but a precise phonon spectrum does not reveal under what conditions the crystal will melt unless the melting mechanism is understood. Two quite different mechanisms have typically been considered. The first is melting through an increase in the lattice vibration until the ratio of the mean square displacement of an electron from its equilibrium position ($< \delta r^2 >$) to the square of the interelectron spacing exceeds some critical value. This is called the Lindemann criterion and for $T = 0\,\mathrm{K}$ it is given by [Egu83]:

$$\frac{< \delta r^2 >}{a^2} \approx 0.0256 \tag{2.56}$$

Since this ratio can be calculated from the phonon spectral density, it does allow an estimate of the critical melting value of a (and thus of r_{sc}) to be made for zero temperature, although it does not determine if the transition from a solid to a liquid should be first order, with discontinuous changes in some of the measurable state variables, or second order, with the state variables changing continuously. At higher temperatures the mean squared lattice vibration in two dimensions becomes infinite and other quantities such as the relative displacement of adjacent electrons must be used in place of δr. Under these conditions, the results of using the Lindemann criterion may not be reliable [Chu91a]. The second mechanism involves melting at finite temperatures through the creation of dislocation pairs or grain boundaries in the crystal. This mechanism is associated with the Kosterlitz-Thouless criterion which states that the crystal will melt when the interaction energy between a pair of dislocations is $E_{dis} \approx 2k_BT$. This allows a phase boundary to be established wherever the dislocation energy can be calculated. If Kosterlitz-Thouless melting is the appropriate mechanism then the transition should have some first order characteristics [Gla90].

One study which is useful in understanding the effects of impurities in real systems is that of I. M. Ruzin, S. Marianer and B. I. Shklovskii [Ruz92], in which they numerically calculate the magnitude of the pinning potential as a function of the distance between the Wigner crystal and a charged impurity. They find that impurities in a delta-doped layer located far from the 2DES ($d >> a$) create only very weak pinning. This means that the correlation length in the Wigner crystal, ξ, is much larger than the spacing between impurities, and thus the Wigner crystal is quite rigid. Impurities located close to the 2DES, however, cause strong pinning so that the correlation length

becomes smaller than the inter-impurity spacing and the phase coherence of the crystal is broken at each impurity site. They also note that acceptors cause much stronger pinning than donors. This type of calculation is important, because it shows which types of impurities have to be considered in a realistic model.

Another important factor in Si-MOSFETs is that the electrons can exist in either of two valleys which are separate in $\vec{k}$-space. The valley splitting energy is large even at zero magnetic field, $\Delta_v = 2.4 + 0.6B$ (measured in Kelvin), so that it is important to consider what effect this may have on the formation of a Wigner crystal. One possibility is to treat the valleys as two separate electron systems which are coupled only through the Coulomb force. This is similar to a pair of electron lattices in adjacent quantum wells, a system which was considered by L. Swierkowski, D. Neilson and J. Szymanski [Swi91], who showed that the value of n_{CM} in such systems was increased by a factor of 3.4 due to the increased Coulomb coupling between the electrons in the two wells. Since each valley in a Si-MOSFET only contains half of the total electrons, this means that the cold melting concentration for the total two valley system would actually increase by a factor of 6.8 [DIo92a]. This leads to values of n_{sc} which are as high as $4.5 \times 10^{10}\,\mathrm{cm}^{-2}$ in Si-MOSFETs (calculated using $r_{sc} = 37 \pm 5$), coming fairly close to the regime where the characteristic signatures of a Wigner crystal have been observed.

The most realistic model yet developed for Si-MOSFETs, however, is probably that of S. T. Chui and B. Tanatar [Chu93c]. They use Monte Carlo calculations which take into account the valley degenaracy and the effects of impurity scattering to calculate the ground state energies of the quantum fluid (electron gas or liquid) and the Wigner crystal in fully polarized, partially polarized and unpolarized states. Their choice of impurity densities and other parameters is compatible with the Si-MOSFETs discussed in this thesis. Specifically, they chose a donor concentration of $10^{10}\,\mathrm{cm}^{-2}$ located at a setback distance of $300\,\mathring{A}$, and a surface roughness potential with a correlation length of $100\,\mathring{A}$ and a fluctuation in surface height of $3\,\mathring{A}$. The results indicated a transition from a partially polarized fluid to a fully polarized solid at $r_{sc} = 10$ (the value quoted in Table 2.2). For Si-MOSFETs it is currently the most realistic value of r_{sc} available because it results from direct ground state energy calculations which take into account the specific parameters of the Si-MOSFET.

2.5.4 Phase Slip and Dislocation Pair Models

S. T. Chui also helped to develop a self consistent model of the conduction mechanism in the Wigner lattice which explains (and makes specific predictions about) the activation energy, the electric field threshold and the Hall resistivity [Chu91a, Chu91b, Chu93b]; three key components of the electronic transport signature of a Wigner crystal. This model assumes that the most important charge carrying excitations in the crystal are bound dislocation pairs and that melting occurs through a Kosterlitz-Thouless mechanism involving dislocation pairs rather than individual dislocations. A single dislocation is best visualized as a shift of the electrons from their equilibrium lattice positions which has greatest magnitude near the centre of the dislocation but decreases with increasing distance. It can be described in terms of a displacement field. A dislocation pair consists of two dislocations with displacement fields in opposite directions, so that any electron which is far from the pair will be almost unaffected since the complementary displacement fields will cancel each other. This is similar to the way in which the field of an electric dipole decreases faster with distance than that of either of its two component charges taken individually. Thus the energy to create a dislocation pair is smaller than that required to create a single dislocation and this is what binds the dislocation pair together. Since there is a net displacement of charge in the area between the two dislocations, it is possible to move this locally altered electron density and thus conduct charge by moving the bound pair as a unit.

This model is remarkably similar to the model of transport through phase slips in a charge density wave [Chu93b]. The picture of the entire CDW (or a segment of it) sliding coherently past the impurities once $E > E_T$ is too simplistic. In a disordered sample where the number of impurities is large and the length of the phase coherent segments small, it is important to consider what happens at the boundaries between different coherent segments. This question is addressed by the phase slip model, in which the CDW is assumed to have its phase pinned at impurity sites and is analyzed in terms of the energy required to cause the phase to change by 2π at such a site. The energy can be provided either thermally or by the external electric field and corresponds physically to a sudden relaxation of the strained CDW shown in Figure 2.11 to an unstrained position where the pinning centre has moved by one wavelength to the left with respect to the CDW. This is equivalent to moving the edges of the coherent segment by one wavelength to the right with respect to the fixed impurity. Thus, in one dimension the phase slip model is directly equivalent to the model of a sliding CDW with multiple

coherent segments.

In two dimensions the loss of phase coherence occurs in both directions, so the CDW is divided up into phase coherent islands, pinned by impurities. In a Wigner crystal this is equivalent to the idea of a single sample containing many small crystallites, each of which is separately pinned. Here the phase slip is propagated through the sample as a 2π phase soliton, a stable, moving phase shift. If there are many solitons present, each one represents a moving site of local displacement and they can aggregate into multi-soliton complexes which are equivalent to dislocations [Bra91, Nad92]. It is then possible to view the sliding CDW (or Wigner crystal) as a single coherent unit containing moving phase soliton complexes (dislocations) or as a group of sliding phase coherent islands (crystallites) separated by phase dislocations. Both viewpoints are equivalent.

Since the phase slip model is so similar to the dislocation pair model of charge transport which was created specifically for a Wigner crystal, it should be possible to use some of the previous work on phase slippage to make predictions about the electronic transport characteristics of the Wigner crystal. Most of the previous work on phase slippage, however, has focused on narrow superconducting channels or on quasi one dimensional charge density waves. In order to compare the predictions of the phase slip model with the experimental results, it is necessary to express them in a framework that is compatible with a two dimensional Wigner crystal. Certainly, phase slippage is characterized by non-linear $I-V$ curves and thermally activated conduction. It is also associated with both broad band and narrow band noise. All of these characteristics, however, can also be determined from the simpler model of a coherent CDW sliding past impurities. To get more precise predictions it is necessary to use the expression which has been developed [Bor86, Gil86, Nad92] for the current carried in a CDW undergoing phase slippage, which can be written as:

$$I_{CDW} = I_0 \exp\left[\frac{-F}{k_b T}\right] \sinh\left[\frac{neV}{k_b T}\right] \tag{2.57}$$

where I_0 is the total current provided by a single 2π phase slip, F is the energy barrier which must be overcome in order to shift the phase by 2π, V is the voltage applied across the distance corresponding to a 2π phase slip and n is the average number of electrons in the unit cell of the CDW over which each phase slip event nucleates. In the framework of a two dimensional Wigner crystal some of these quantities need to be redefined. I_0 remains the basic unit of current, but now it is related to an integration of the phase slips occuring across the entire width of the two dimensional sample, so it

is more accurately defined as a constant which gives the magnitude of the current rather than as a fundamental property of the microscopic physics. The phase slip energy barrier can be identified with Δ, the experimentally determined thermal activation energy. The voltage corresponding to a 2π phase slip is equal to the applied electric field multiplied by the distance over which the electron phase changes by 2π (ie. the interelectron spacing) so that $V = E_{app}a$. Finally, the number of electrons in a unit cell of the CDW can be assumed to be equal to the average number of electrons in a coherent crystallite in the Wigner crystal, since all the electrons in each crystallite must undergo phase slippage together or the crystallite will be destroyed. Making these substitutions and dividing Equation 2.57 by $V_{xx}f$ where V_{xx} is the potential measured between a pair of longitudinal contacts such as source and drain or P1 and P3 and $f = W/l$ is the geometrical factor given by the width of the sample divided by the length between the pair of contacts used to measure V_{xx} results in the equation:

$$\sigma_{xx} = \frac{1}{\rho_{xx}} = \frac{I_x l}{V_{xx} W} = \frac{I_0 l}{V_{xx} W} \exp\left[\frac{-\Delta}{T}\right] \sinh\left[\frac{neaV_{xx}}{lk_b}\frac{1}{T}\right] \tag{2.58}$$

This single equation makes a specific prediction about the relationship between σ_{xx} and V_{xx} (or alternatively $E_{app} = V_{xx}/l$) and the relationship between σ_{xx} and T. Since all three of these quantities can be varied and measured experimentally, this should give an excellent idea of whether or not the phase slip model is valid for a given set of experimental conditions. If it is valid, then this will be strong evidence for the existence of a CDW or Wigner crystal under those conditions.

The dislocation pair model developed by S. T. Chui and others for the two dimensional Wigner crystal [Chu91a, Chu91b, Chu93b] also makes some specific experimentally verifiable predictions. S. T. Chui and K. Esfarjani calculated that the energy required to create a single dislocation pair (experimentally identifiable with the thermal activation energy, Δ) followed the form:

$$\Delta \propto (\nu_0 - \nu) \tag{2.59}$$

where ν_0 is the critical filling factor at which the crystal melts. At constant magnetic field, $\nu \propto n_s$ and thus:

$$\Delta \propto (n_{sc} - n_s) \tag{2.60}$$

where n_{sc} is the experimentally measured crystal melting concentration at that magnetic field. In a later paper [Chu93b], it was noted that the application of an external electric field may be sufficient to unbind a dislocation pair

into two independent dislocations. In order for this to happen, the dislocations will have to tunnel through the energy barrier Δ. The effective decay length associated with this type of quantum mechanical tunneling is given by:

$$d \approx \left(\frac{\hbar^2}{2m_d\Delta}\right)^{0.5} \tag{2.61}$$

Here m_d is the effective mass for the moving dislocation wave ($m_d \approx 0.1m^*$). The electric field threshold, E_t associated with depinning the crystal will occur when the energy gained from the field in breaking apart a dislocation pair (proportional to $E_t d$) is larger than the energy required to create a dislocation pair (Δ), and thus large numbers of unbound, charge carrying dislocations will occur, creating the sudden increase in current associated with the threshold field. Since these dislocations can also be viewed as the boundaries between crystallites, this is the point at which the individual crystallites begin to slide. Mathematically this can be expressed as:

$$E_t d = E_t \left(\frac{\hbar^2}{2m_d\Delta}\right)^{0.5} \propto \Delta \tag{2.62}$$

Rearranging and substituting in the result of Equation 2.60 gives:

$$E_t \propto \Delta^{1.5} \propto (n_{sc} - n_s)^{1.5} \tag{2.63}$$

This is another theoretical prediction which should be easy to test experimentally.

The final important prediction of the dislocation pair model relates to the Hall resistivity. For a Hall voltage created by the Lorentz force acting on mobile dislocation pairs and through them on the whole electron lattice, the Hall resistivity should have the form [Chu93a]:

$$\rho_{xy} = \frac{\rho^0_{xy}}{1+x} = \frac{B}{en_s(1+x)} \tag{2.64}$$

This means that the Hall resistivity will be equal to its classical value, ρ^0_{xy} divided by a scaling factor, $(1+x)$, where x is related to the material parameters of the sample and is typically very small (on the order of 10^{-18}). For experimental purposes, this means that ρ_{xy} in the Wigner crystal phase should retain its classical value regardless of what happens to ρ_{xx}.

These three predictions of the dislocation pair transport model give some more specific transport characteristics of the Wigner crystal than those which are generally inferred from the CDW analogy. The phase slip model, which

is roughly equivalent to the dislocation pair model, also provides a specific relationship between σ_{xx}, E_{app} and T which can be experimentally verified. These two models are quite realistic and if their predictions are observed in a system it would lend support to the existence of a Wigner crystal in that system.

Chapter 3. Experimental Details

3.1 The Samples

The experiments discussed in this thesis all made use of a set of extremely high quality Si-MOSFET samples, which were fabricated at the Institute of Microelectronics in Chernogolovka and were brought to Canada by Dr. V. M. Pudalov. These unique samples, which were produced during the early 1980's, have electronic mobilities, μ, among the highest ever recorded for silicon [Kru91a]. They allow electron transport measurements to be performed in the normally inaccessible regime of extremely low electron densities.

3.1.1 Detailed Description

Fourteen samples from different wafers were studied, but three were determined to be the most suitable (in terms of their electronic mobility, contact resistances, and overall homogeneity of the two dimensional layer) for studying electronic properties in the dilute electron regime. All of them followed the same basic design. A SiO_2 layer of approximately 0.2μm was deposited over a lightly p-doped Si substrate. The contacts were established in a Hall bar geometry with a source, drain, and three pairs of Hall contacts by standard phosphorous diffusion. Then a final Al metallization layer was used to create the gate and to provide an Ohmic contact between the bonding wires and the sample. The sample itself was rectangular, with a length of 5.0 mm and a width of 0.8 mm. Thus the distance between adjacent longitudinal contacts was 1.25 mm, giving a length to width ratio for conductivity calculations of 1.6. Connecting the samples to the external electronics was accomplished by ultrasonically bonding gold wires to the eight contact pads and the gate, and then to gold plated posts on a 12 pin TO-8 header which fit the probe.

It is possible to obtain a sense of how these samples compare with those manufactured by other groups over the past decade by examining the relevant peak electronic mobilities, μ_p, as well as the values of electron density and temperature at which those mobilities occur. The significance of these numbers and details of how they were calculated will be given later, but

Authors	μ_p cm^2/Vs	n_s at μ_p 10^{11} cm^{-2}	T K
Wheeler [Cha80]	22000	15	4.5
Furneaux [Fur87]	33000	1.8	0.5
Kruithof [Son90]	25000	6.0	1.1
Pudalov (Si-2)	22800	4.1	0.025
Pudalov (Si-5)	40800	3.2	0.025
Pudalov (Si-11)	33100	4.1	0.025

Table 3.1: Characterization Data for High Mobility MOSFETs

Table 3.1 lists the appropriate values and shows immediately that these Russian MOSFETs had peak mobilities which were among the highest of any Si-MOSFETs ever produced.

3.1.2 Fabrication

It is noteworthy that only a select few of the countless Si-MOSFETs studied during the last three decades have had peak mobilities over 2.0×10^4 cm^2/Vs. This is why it is useful to study the fabrication process used to produce these high quality MOSFETs. G. H. Kruithof [Kru91a] provides a detailed discussion of the fabrication steps required, and private communications with V. M. Pudalov and R. G. Wheeler have revealed that they used similar techniques to those used by Kruithof.

Cleanliness is the single most important issue in fabricating high mobility Si-MOSFETs. Every stage of sample processing must be designed to isolate the sample from impurities which can degrade the two dimensional layer. This means that the sample should be protected from the atmosphere as much as possible, since atmospheric oxygen atoms can oxidize the silicon in an uncontrolled fashion. It is also vital to use the highest purity chemicals possible (impurity levels, especially of the light metallic ions which easily diffuse into the oxide, should ideally be in the parts per billion range) in all of the wet etching stages and to rinse the sample thoroughly between each etch. The water used for rinsing the sample between etches must also be very pure, as it is very easy for ions such as sodium or potassium to diffuse into the silicon, creating Coulomb impurities. To help keep the chemicals and gases used in each stage separate it is also beneficial if separate sets of vessels can be used for each chemical etch and subsequent rinses, and separate furnaces

for each high temperature stage. Cleaning the furnace tubes and rods before each use with some type of fluorine or chlorine derived chemical such as HF or CH_3-CCl_3 further helps to reduce the possibility of contamination from stray atoms which may be present in the furnace walls.

Another important issue is that of minimizing the surface roughness at the interface between the silicon and the SiO_2 oxide layer. Ideally, the interface should be as flat as possible. This was accomplished by repeatedly laying down a thin oxide layer and then removing it before the final oxidation began. Doping the contacts to be strongly n-type by implanting phosphorous atoms is also a delicate step, since care must be taken not to let the phosphorous diffuse into the main channel under the gate, where a small amount could affect the doping over a wide area of the lightly doped p-type silicon. Finally, it was important during the metallization of the gate and contact pads with aluminum not to allow a charge to build up on the gate as it was being formed because this could cause an electrical break down through the oxide layer, permanently ruining the MOSFET. It was also essential to try and minimize any external contaminants present during metallization as these could easily diffuse into the oxide, lowering the mobility or even creating a break down path from the gate to the silicon.

After the initial fabrication, there are also concerns with how the samples are handled. The samples must be stored in a static free environment and once they have been connected to an external circuit they must be carefully protected from any sudden electrical discharges. Failure to observe these precautions can result in a breakdown of the insulating oxide which separates the gate from the silicon surface, severely damaging the MOSFET. Even a voltage spike as low as 100 Volts between the gate and the two-dimensional electron layer may cause shifts in the impurity distribution near the interface which will shift the threshold gate voltage dramatically. Although the samples usually recovered from a shock of this second type after a few weeks of equilibration at room temperature, it could still terminate an experimental run prematurely. Even with all possible precautions, the samples could not be maintained at base temperature for more than a month before their contacts began to degrade and they had to be warmed up to room temperature for a few days to recover. The precise reason for this slow degradation is not known, but it should be noted that others have observed a long term degradation of their samples [Kru91a] even when stored at room temperature. In these Russian samples such degradation during storage was not a problem, since the peak mobilities changed by less than 10% between 1990 and 1993. The reproducibilities of the threshold gate voltage, insulator capacitance, etc.

were also very good from one cooldown to the next, provided that the samples were cooled down very slowly (taking approximately four hours to cool from 300K to 77K) while applying a constant gate voltage (usually $V_g = 9.0$ Volts).

3.1.3 Characterization

All of the samples were characterized in several stages. First, a potential of 9.0 volts was applied between the gate and one contact, and the resistances across all possible contact pairs and between the gate and each contact were checked at room temperature. For these measurements a voltmeter was set with a limiting current of approximately 1 μA so as not to damage the sample. In order to pass this stage of testing, a sample had to have at least two contact pairs (preferably source-drain and one other pair) which had path resistances below 1MΩ and could not have any gate-contact pathways with resistances below 100MΩ.

After successful room temperature testing, a sample would be cooled in approximately 4-6 hours from room temperature to 77K and then from 77K to 1.7K in an additional 2-4 hours. More detailed characterization could then begin, starting with the determination of the electron density in the 2DES as a function of the gate voltage. This was accomplished using the traditional assumption [And82] of treating the insulator in the MOSFET structure as a parallel plate capacitor with the gate and the inversion layer as the complementary charge accumulators. The threshold gate voltage, V_t, is defined as the potential when inversion first occurs and thus the effective voltage across the capacitor is obtained by subtracting V_t from V_g, the applied gate potential. The capacitance is then given by $C_{ins} = Q/(V_g - V_t)$, where Q is the charge of the entire inversion layer. This leads to a simple formula for the carrier density:

$$n_s = \frac{Q}{eA} = \frac{C_{ins}}{eA}(V_g - V_t) \tag{3.1}$$

The area of the sample and the capacitance of the insulating layer are fixed during sample fabrication, but as noted previously, the threshold voltage is sensitive to the distribution of impurities near the interface and thus it can easily be affected by sudden electrical discharges or even by the way electronic traps are filled as the sample is cooled. Thus, although the slope of the n_s versus V_g graph should not change significantly with time, the threshold voltage should be checked every time the sample is cooled down from room temperature and whenever it is suspected that the sample may have undergone a shock.

Two methods of determining the insulator capacitance and threshold voltage were used. The first was to set the gate voltage and then measure the resistance between a pair of longitudinal contacts as a function of transverse magnetic field. An experimental plot of such a measurement using contacts P1 and P5 from sample Si-11 is shown in Figure 3.1. It shows clear resistance minima at the integer filling factors which correspond to the Integer Quantum Hall Effect, with arrows indicating which filling factor corresponds to which minimum. This correspondance is usually made by manually measuring the resistance across a pair of Hall contacts when the longitudinal resistance trace is observed to be at a minimum and deriving the filling factor from the value of the Hall resistance plateau. In this case, four Landau levels were filled at $B = 1.16$T, two at $B = 2.35$T, and only one at $B = 4.68$T. Since each magnetic quantum level can contain $2.418 \times 10^{10}B$ electrons per square centimeter, where B is the transverse magnetic field measured in Tesla, a filling factor of ν at a magnetic field B corresponds to a carrier density of $2.418 \times 10^{10}B\nu$ cm^{-2}T^{-1}. Defining the electron density in the 2D layer as n_s, measured in cm^{-2}, gives:

$$\nu = \frac{1}{B}\frac{n_s}{2.418 \times 10^{10}} \tag{3.2}$$

Thus a plot of filling factor versus inverse magnetic field will produce a straight line whose slope defines the electron density for the specific gate voltage used. Repeating the process for different gate voltages produces multiple points on the graph of n_s versus V_g and allows a complete characterization of n_s as a function of V_g.

Sometimes, however, the minima can be quite large (See $\nu = 1$ in Figure 3.1), which leads to a high uncertainty in the values of B corresponding to those filling factors and ultimately to a high uncertainty in C_{ins} and V_t. Since the experiments described in this thesis took place in the very dilute electron density regime where $V_g - V_t$ was often less than 0.1V, it was vitally important that the value of V_t be established accurately. To accomplish this, a second method was employed, in which the magnetic field was held constant while the gate voltage was varied. This also leads to a graph where the longitudinal resistivity has minima corresponding to each integer filling factor, but now the density can be varied over a wide range in a single experiment. Figure 3.2 shows such a plot of longitudinal resistance versus gate voltage for sample Si-31. There are five large, distinct minima, with smaller minima half-way between them becoming prominent at lower gate voltages. The larger minima correspond to Landau levels with the spin and valley splittings unresolved whereas the smaller minima are due to the spin splitting.

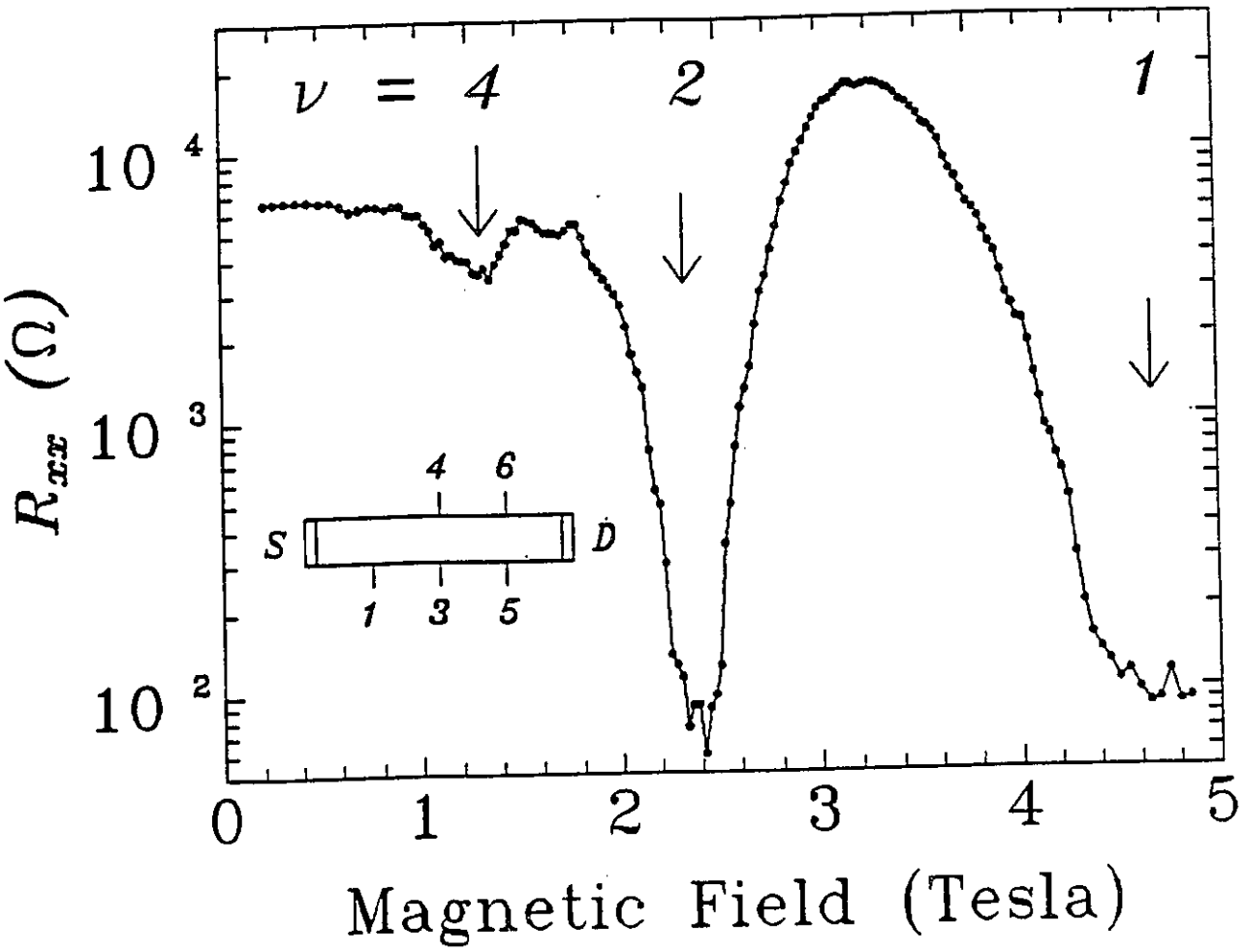

Figure 3.1: The resistance between a pair of longitudinal contacts on Hall bar sample Si-11 (with geometry as indicated in the inset) is shown as the magnetic field is varied from 0 to 5 Tesla. The numbered arrows indicate the filling factors corresponding to the three Integer Quantum Hall Effect resistivity minima.

At the lowest gate voltages, valley splitting induced minima start to become visible between the spin split levels. The minimum number (starting from the lowest gate voltage available and counting both spin and valley splitting induced minima) can now be plotted versus gate voltage. This is essentially a graph of ν versus V_g and it should have an intercept at V_t where no Landau levels are filled, and a slope, c, given by:

$$\nu = \frac{1}{B}\frac{n_s}{2.418 \times 10^{10}} = c(V_g - V_t) \tag{3.3}$$

Therefore,

$$n_s = 2.418 \times 10^{10} Bc(V_g - V_t) \text{ cm}^{-2} \tag{3.4}$$

Here B is the constant magnetic field at which the gate voltage was swept and c is the slope in units of V^{-1}. Figure 3.3 shows a typical 'fan' diagram of the type often used to characterize n_s. There are three different lines, at different magnetic fields, so that the average slope and intercept can be used to determine C_{ins}/eA and V_t.

The only drawback to this method of determining n_s is that it may be difficult to tell whether the spins and valleys are truly resolved or not, and thus the slope may be wrong by a factor of two or four. This potential problem can easily be avoided by using different magnetic fields for the different lines on the fan diagram. If any of the slopes are incorrect the lines will not converge at the same value of V_t. Alternatively, the density at one of the minima on each resistance versus gate voltage curve can be verified with a manual Hall resistance measurement and that point can then be used to determine whether the slope needs to be adjusted.

3.1.4 Mobility

Once the electron density in the inversion layer had been properly characterized for a given sample, the electronic mobility was determined by a single experiment at zero magnetic field in which longitudinal resistance was measured while the gate voltage was swept over a wide range. In this type of measurement each value of V_g is converted to a corresponding value of n_s, and the zero field resistance, R^0_{xx}, is converted to an electronic transport mobility, μ_t, using the formula:

$$\mu_t = \frac{e\tau_t}{m^*} = \frac{\sigma^0_{xx}}{en_s} = \frac{1}{\rho^0_{xx} en_s} = \frac{1}{\frac{W}{l} R^0_{xx} en_s} \frac{\text{cm}^2}{\text{V s}} \tag{3.5}$$

Here, τ_t is the mean scattering time for electrons in the two dimensional layer, m^* is their effective mass in the plane of the inversion layer, σ^0_{xx} is the

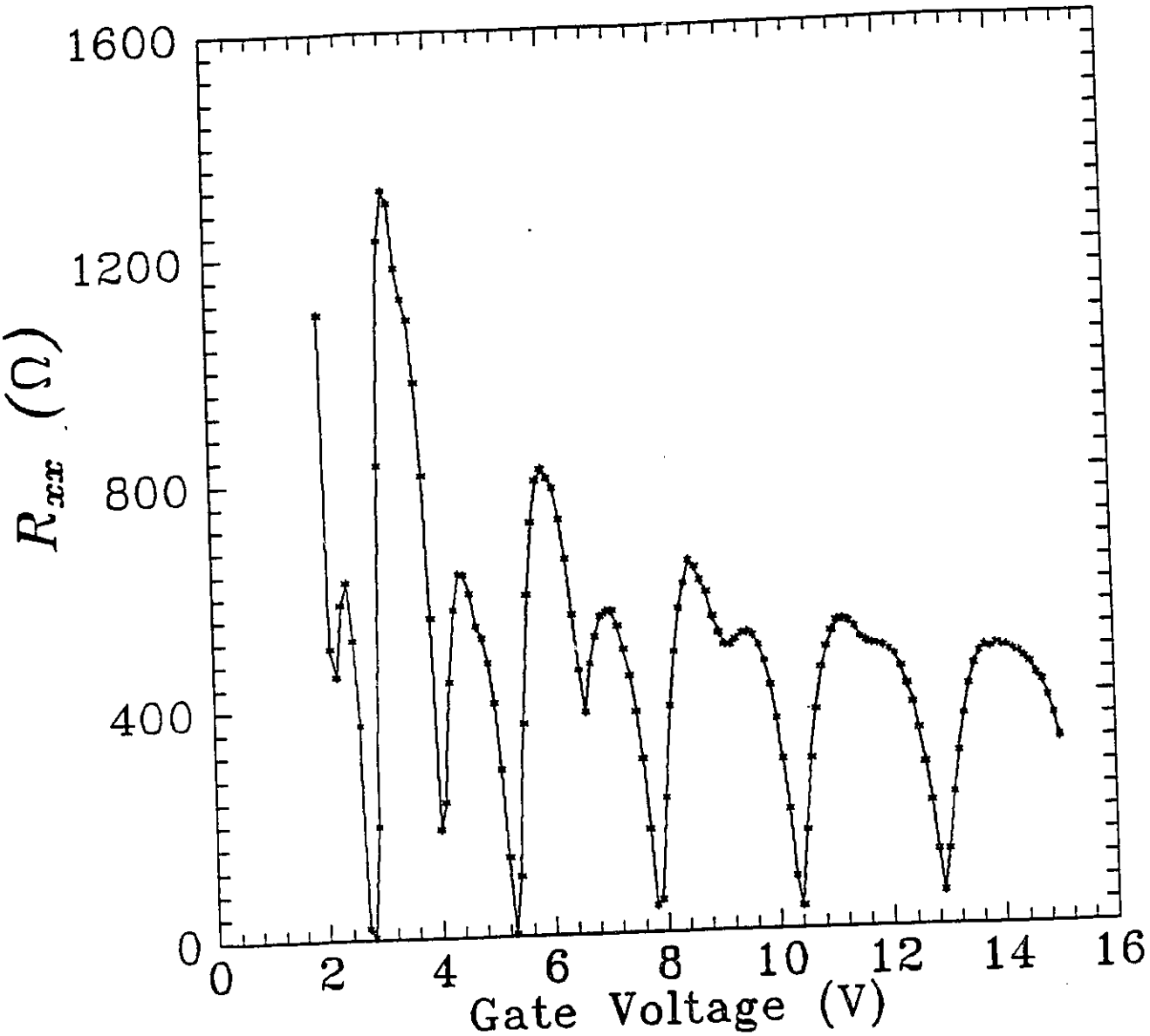

Figure 3.2: The longitudinal resistance is plotted as a function of applied gate voltage for sample Si-31 at $B = 4.0\,\mathrm{T}$. Both spin splitting and, at the lowest gate voltages, valley splitting can be observed.

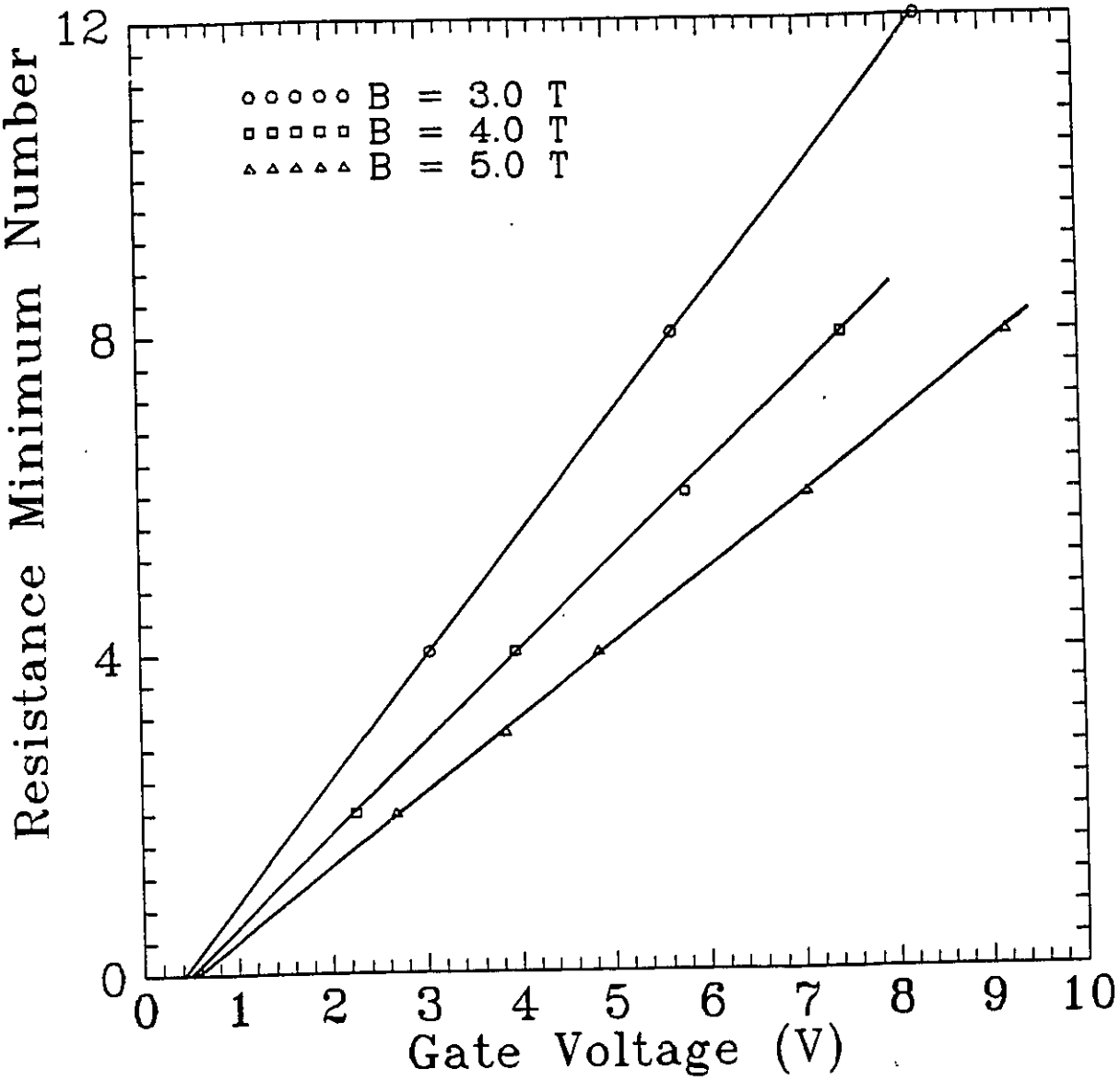

Figure 3.3: This is a plot of minimum number with spin and valleys resolved versus gate voltage for sample Si-5 at three different magnetic fields.

longitudinal conductivity in the 2DES at zero magnetic field, ρ_{xx}^0 is the zero field resistivity, l is the distance between the two contacts used to measure R_{xx}^0 and W is the width of the sample. Figure 3.4 demonstrates the geometry, and also gives a typical example of mobility versus carrier density curves at two different temperatures. Note that, as has previously been observed in high mobility Si-MOSFETs at low temperatures [Cha80, Gol86a, Kru91b], the mobility rises as the temperature decreases. This is because τ_t in this regime is dominated by scattering from Coulombic impurities near the inversion layer and by surface roughness scattering from the silicon-oxide interface, and both of these scattering mechanisms are screened more effectively as the temperature decreases. Also note that the peak mobility, μ_p, is limited by a sharp decrease on the low carrier density side and by a slow fall off at higher electron densities. This is because the lower density regime is dominated by Coulomb scattering and the higher density regime by surface roughness scattering, and these have different dependences on carrier density.

After a sample had been characterized at $T = 1.7\,\mathrm{K}$, it was cooled to 23 mK, the base temperature of the dilution refrigerator, and μ_p and V_t were rechecked. The best samples were selected on the basis of their low temperature mobility (20000 $\mathrm{cm}^2/\mathrm{Vs}$ or greater) and their threshold gate voltage ($0 < V_t < 1\,\mathrm{V}$).

3.1.5 Sample Quality

Although the transport mobility is the simplest way of comparing the quality of different samples, it is not sufficient by itself. There are other factors which may be important, depending upon the purpose for which the samples are required. The Russian MOSFETs studied here, for instance, have peak mobilities which are an order of magnitude higher than those of typical industrially produced Si-MOSFETs, but they are still two orders of magnitude below the mobilities of the best GaAs/AlGaAs heterostructures. This suggests that the heterostructures would make a far better candidate for studying the possible formation of the Wigner crystal. The silicon samples, however, have a higher effective mass and lower dielectric constant, which enhances the stability of the Wigner crystal. Furthermore, the regime where the electron solid forms is not the same regime where peak mobilities occur.

Excellent references on the detailed scattering mechanisms for electrons in silicon inversion layers [Cha80, And82, Gol85] and GaAs/AlGaAs heterostructures [Har85, Bas88, Col91] are available. These reveal that in the high electron density range of $2 - 3 \times 10^{11}\,\mathrm{cm}^{-2}$ and up, the peak mobility in Si-MOSFETs is limited mostly by surface roughness scattering from the

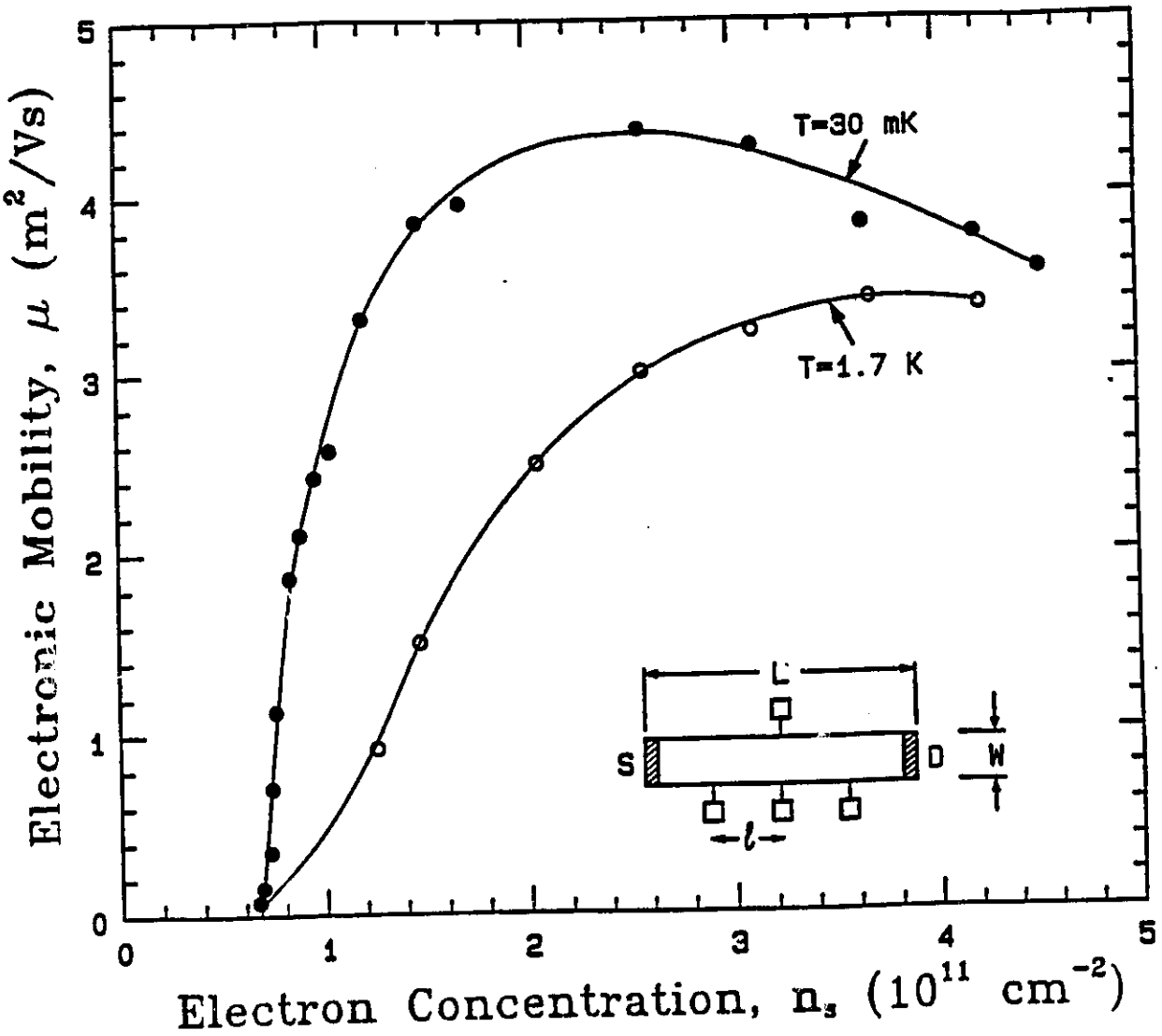

Figure 3.4: Electronic mobility is plotted versus electron density at two different temperatures for sample Si-5. Note the increase in mobility as temperature is decreased.

imperfect interface, whereas in GaAs/AlGaAs heterostructures the limiting scattering mechanism is Coulombic scattering from remote ionized impurities. In the very low density regime where a Wigner crystal may form, however, the mobility drops sharply and ionized impurity scattering is dominant in both systems. Table 3.2 shows the relevant data for comparing the Si-MOSFETs and GaAs/AlGaAs heterostructures in this regime. In the silicon samples, the bulk p-dopants ($N_A \approx 7\times10^{14}\,\mathrm{cm}^{-3}$) are expected to be homogeneously distributed and thus the number of them in a typical two dimensional layer of 50-100 $\AA$ will be insignificant. Charged impurities in the oxide layer created by improperly matched silicon atoms (eg. Si^+ or SiO^+) or by diffused metallic ions such as Na^+ do tend to be concentrated within a few atomic layers of the interface. The number of these impurities can be estimated (See Section 2.3.1) and for the Russian MOSFETs is $n_i \approx 4-5\times10^{10}\,\mathrm{cm}^{-2}$. The setback distance, d, is the average distance of these impurities from the centre of the confining well which defines the 2DES, and for an electron density of $n_s = 8\times10^{10}\,\mathrm{cm}^{-2}$ and assuming $N_A = 7\times10^{14}\,\mathrm{cm}^{-3}$ this gives $d = 45-80\AA$. The potential fluctuations caused by these impurities may be detrimental to the formation of an electron solid, but since the appropriate length scale for discussions of the Wigner crystal is the Bohr radius, $a_B = \hbar^2\kappa/m^*e^2$, it is necessary to convert the setback distance into units of the Bohr radius before deciding on the importance of the fluctuations. In a Si-MOSFET the Bohr radius is about 20 $\AA$, so the setback distance to Bohr radius ratio is approximately 2-4. It is now possible to compare with a typical high quality GaAs/AlGaAs heterostructure, which will have a highly doped layer ($n_i \approx 10^{12}\,\mathrm{cm}^{-2}$) separated from the 2DES by a spacer layer (500 $\AA$ is a reasonable thickness). The setback distance in units of the Bohr radius (approximately 100 $\AA$) is thus 5. This is similar to the setback distance in the Si-MOSFET, but the ratio of impurities to 2D electrons (n_i/n_s) is more than an order of magnitude less in the MOSFET. Thus, when the specific conditions of the dilute electron density regime are considered, it is easy to see how these Si-MOSFETs are comparable to the best GaAs/AlGaAs heterostructures, despite the large difference in their transport mobilities.

3.2 Equipment

3.2.1 Refrigerator and Magnet

The equipment used in all of the experiments was built around a TL400 top-loading Oxford Instruments dilution refrigerator, which has a base temperature of 23 mK and a cooling power of 400 μW at 100 mK. The top loading

	Si-MOSFET	GaAs/AlGaAs Heterostructure
n_i (cm^{-2})	$4-5\times10^{10}$	$\approx 10^{12}$
d ($\AA$)	45-80	≈ 500
$\frac{n_i}{n_s}$	≈ 0.5	≈ 12
$\frac{d}{a_B}$	2-4	≈ 5

Table 3.2: Quality Comparison Data (Si vs GaAs/AlGaAs)

probe is inserted directly into the helium mixture, using a vacuum lock. The probe is fitted with two copper thermal anchors; a screw thread anchor which is linked to the top of the inner vacuum can at 4.2K, and a frictional conical anchor at 1K which separates the dilution unit from the insert. In order to prevent thermal shorts to room temperature, there is an insulating displacer between the two anchors and another below the conical anchor. The total volume of the ^{3}He-^{4}He mixture is 319 L with a mixing ratio of 10% ^{3}He. The long glass tail which extends into the magnet has a fifteen millimeter diameter access space and the probe is configured to allow 1 Lakeshore semiflexible coaxial cable and up to 18 40-gauge wires to be connected. The wires are arranged in twisted pairs to reduce thermal emf and exit from the top of the probe using a shielded 18 pin Fischer connector. For experiments with MOSFETs, the coax was typically used to control the gate voltage, since a stable gate voltage is essential to any experiment at low electron density.

The thermometry in this refrigerator relied upon two calibrated carbon Speer resistors. These were initially calibrated below 50 mK using a primary nuclear orientation thermometer. Above 50 mK they were calibrated against a cerium magnesium nitrate magnetic thermometer which had itself been calibrated and checked by the National Institute of Standards and Technology (NIST) using the NIST superconducting fixed point diode scale. Later, the calibration was verified using a Lakeshore germanium resistor which was calibrated against a different cerium magnesium nitrate magnetic thermometer. Finally, the base temperature was checked using ^{60}Co nuclear orientation thermometry. The Speer resistors were subject to drift with the thermal cycling of the probe and there were also temperature uncertainties due to magnetic field induced eddy currents or to thermal gradients within the mixing chamber, but the total uncertainty of any temperature measurement varied from as little as 2% at base temperature to as much as 10% in the temperature region from 50 mK to 150 mK, where thermal stability was most difficult to maintain.

The system also contained an Oxford Instruments superconducting magnet. This was the company's first hybrid Nb_3Sn and NbTi magnet with a persistent switch which could achieve $B = 15$ Tesla at $T = 4.2$K. Due to the large hoop stress on the magnet's coils, it was not equipped with a lambda plate to further increase the field at $T = 2$K, or with a cancellation coil to allow zero field measurements. The large size of the magnet also meant that the dilution refrigerator had to be scaled accordingly. The bottom half was contained in the lab, but the top half protruded above the roof of the building in a specially built wooden hut. This put it in a direct line of sight with at least one commercial radio transmitter, and since the probe was coupled inductively to the cryostat, it could act as a radio frequency antenna. In most experiments, this was not a problem, but in the very low electron density regime, the Si-MOSFET becomes highly resistive and rf pickup can be greatly magnified. It was therefore necessary to implement some standard techniques for reducing noise.

3.2.2 Instrumentation

One of the best ways to reduce the impact of noise on the experiments was to integrate data for long periods of time. This was primarily useful in combatting rapidly varying noise sources such as intrinsic Johnson noise in the highly resistive sample and rf pickup noise. There were also slowly varying changes in the background potentials measured from the samples (corresponding to noise frequencies of millihertz or sometimes even microhertz), and these were removed by repeatedly zeroing and/or inverting the polarity of the current. These complications meant that a large number of measurements were usually necessary to obtain a single experimental point when the sample was in its highly resistive state, and thus it was preferable to use a computer to control as much of the equipment as possible.

Figure 3.5 shows a schematic representation of the computer control system used. Every instrument required in an experiment was connected by a GPIB cable to an IOTECH 488/HS Bus Extender. If a piece of equipment was not GPIB compatible, then a GPIB compatible voltage source was used to trigger it or a GPIB compatible digital multimeter (DMM) was used to read its analog output. All the instruments and the bus extender were grounded together on racks, which were kept separate from the table containing the computer, plotter and printer. The only connection to the computer was through the parallel cable linking the two bus extenders, which helped to minimize the influence of noise from the computer CPU on the instruments. The controlling computer was also connected to a GPIB compatible

HP 7550A pen plotter and a HP 82905B printer. Two different CPU's were used at different times. The first was a HP 9816S computer which had the traditional HP advantage of direct and simple access to the GPIB bus from within programs. The second was an IBM compatible 80386 based machine, using LabWindows software. It had the advantages of a superior user interface and the ability to output the data on DOS diskettes which could then be easily transferred to other computers for further processing. The actual instruments used for most of the measurements were not unusual. They included Stanford Research SR560 pre-amplifiers and SR530 lock-ins, as well as HP 3458A DMM's, and a Keithley 230 V-Source.

3.2.3 Noise

A variety of techniques were used to reduce the effects of noise on the measurements. Different grounding and shielding arrangements were tried, until finally one was found which was very successful. This consisted of a series of battery operated instruments inside a shielded metal box near the top of the cryostat. This box was linked by heavy grounding cable to the cryostat frame and also to a copper pipe which carried signals between the instruments in the box and the racks of GPIB controlled instruments in a different part of the lab. A deep grounding spike was also connected to the copper pipe, just before it reached the instrument racks. The goal was to perform the most sensitive measurements in a magnetically and electrically shielded environment close to the cryostat using only battery operated instruments which would be less likely than equivalent line powered equipment to introduce 60 Hz noise or grounding loops into the measuring circuit. Where line powered equipment was necessary, it was kept at a distance from the cryostat both to protect the equipment from the magnetic field, and the sample from any noise produced by the equipment. The shielding box was made of 0.25 inch steel plate and had dimensions of 26 by 26 by 25 inches. It contained two sliding shelves to hold any circuit components and instruments desired, and had several small holes of varying diameter to allow electrical connections to be made. A shielded cable connected the Fischer connector on the probe to a shielded wiring box inside the main box. The various circuit components and battery operated devices such as electrometers or pre-amplifiers were connected to this wiring box, and any outputs which had to be transmitted to the remote instrument racks were connected to coaxial cables which then passed through a separate output hole and straight into the copper shield pipe. Every circuit component or device inside the main box was contained within a separate metal (usually copper or aluminum) case, and all of these

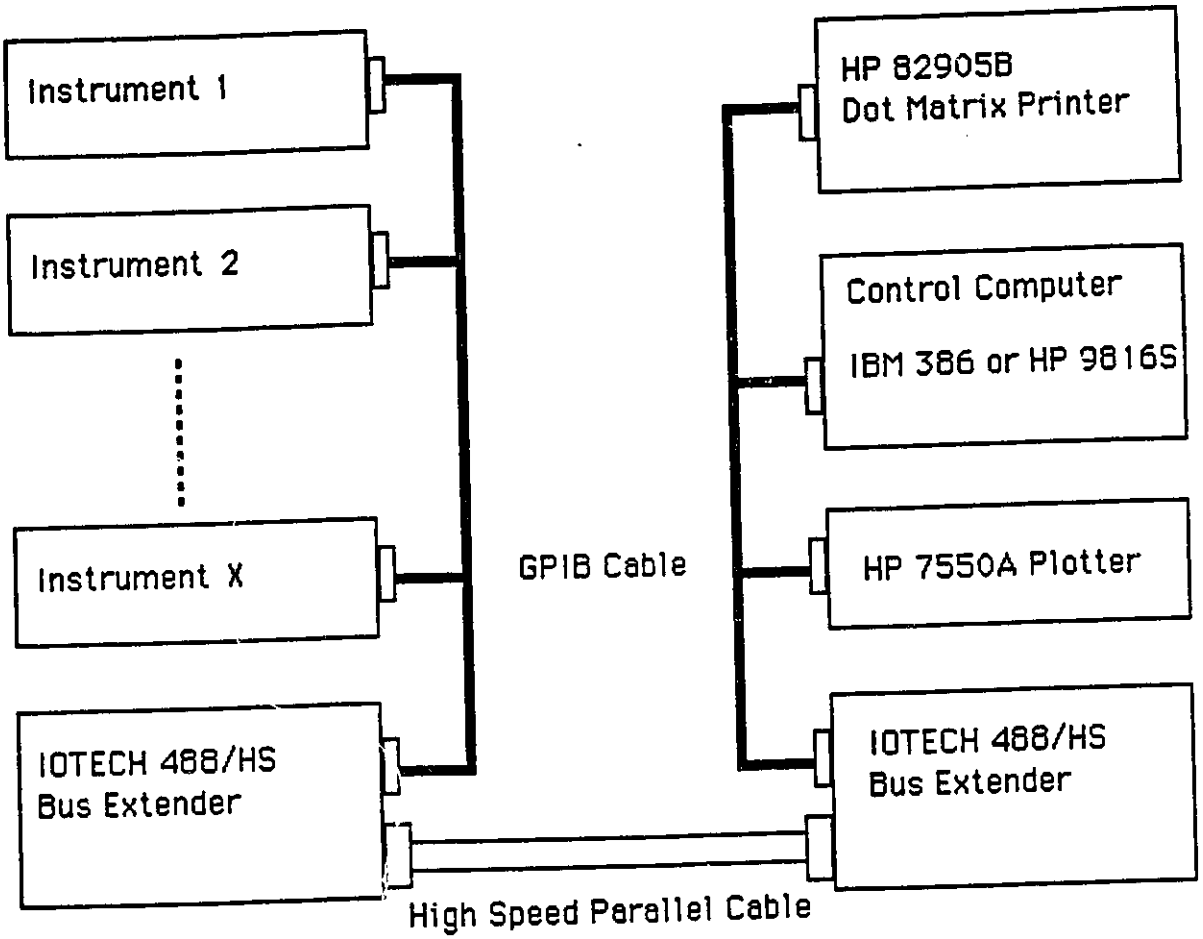

Figure 3.5: A Schematic of the Computer Control System

cases were grounded together to a common reference point which also served as the ground for the sample. Only coaxial cables were used to connect the circuit components inside the box. All cables outside the shielded box (including the copper pipe) were linked to a separate grounding point on the outside of the box. An example of how this arrangement worked for a typical measurement of a voltage versus current curve can be seen in the block diagram of Figure 3.6.

A battery operated symmetric DC electrometer with input resistance of $10^{14}\,\Omega$ was constructed to measure all DC potentials in the very low electron density regime, when the sample resistance was in the gigaohm to teraohm range. A high impedance current source was also constructed using high precision resistors of up to $10^{12}\,\Omega$ to control currents down to femtoamperes. A battery operated voltage source was used for the gate when it was being held constant, but during sweeps of the gate voltage a Keithley 230 V-source had to be used. Since the Keithley V-source was known to inject noise along with its DC potential, a means was sought to break the direct link between the V-source (which was used both to sweep the gate and to drive the current source during current sweeps) and the sample. The solution came in the form of an optical isolator which proved successful in passing DC signals while virtually eliminating associated noise. One of these isolators was built into the current source and another was created for general use, simply by inserting it instead of a coaxial coupler anywhere in a circuit.

The total effect of all these measurement techniques was to reduce noise to insignificant levels during almost all of the experiments. Depending upon the resistance of the sample, the time of day, and various other factors, the rms noise value as measured by the HP DMM or a digital oscilloscope in the instrument racks varied from around 10μV to 10mV. The signal being measured at the same point was often as high as 1.0V. If a high precision measurement was required, the noise could be further reduced by integrating the signal for long periods as described previously. The only time when the noise was a serious problem was when a noise spectrum was being measured from the sample itself. This AC noise was masked by a large DC signal and it became necessary to use a sequence of notch filters to isolate each point in the frequency spectrum and then integrate the data for many hours in order to obtain a single intensity value.

The importance of these efforts to reduce the noise and their ultimate success is demonstrated very well by Figure 3.7. The top graph shows an $I-V$ curve (current sweep) from sample Si-8 taken in 1991 before the noise reduction efforts had been completed. The non-linear $I-V$ caused by the existence of a threshold voltage is barely distinguishable from the other jumps

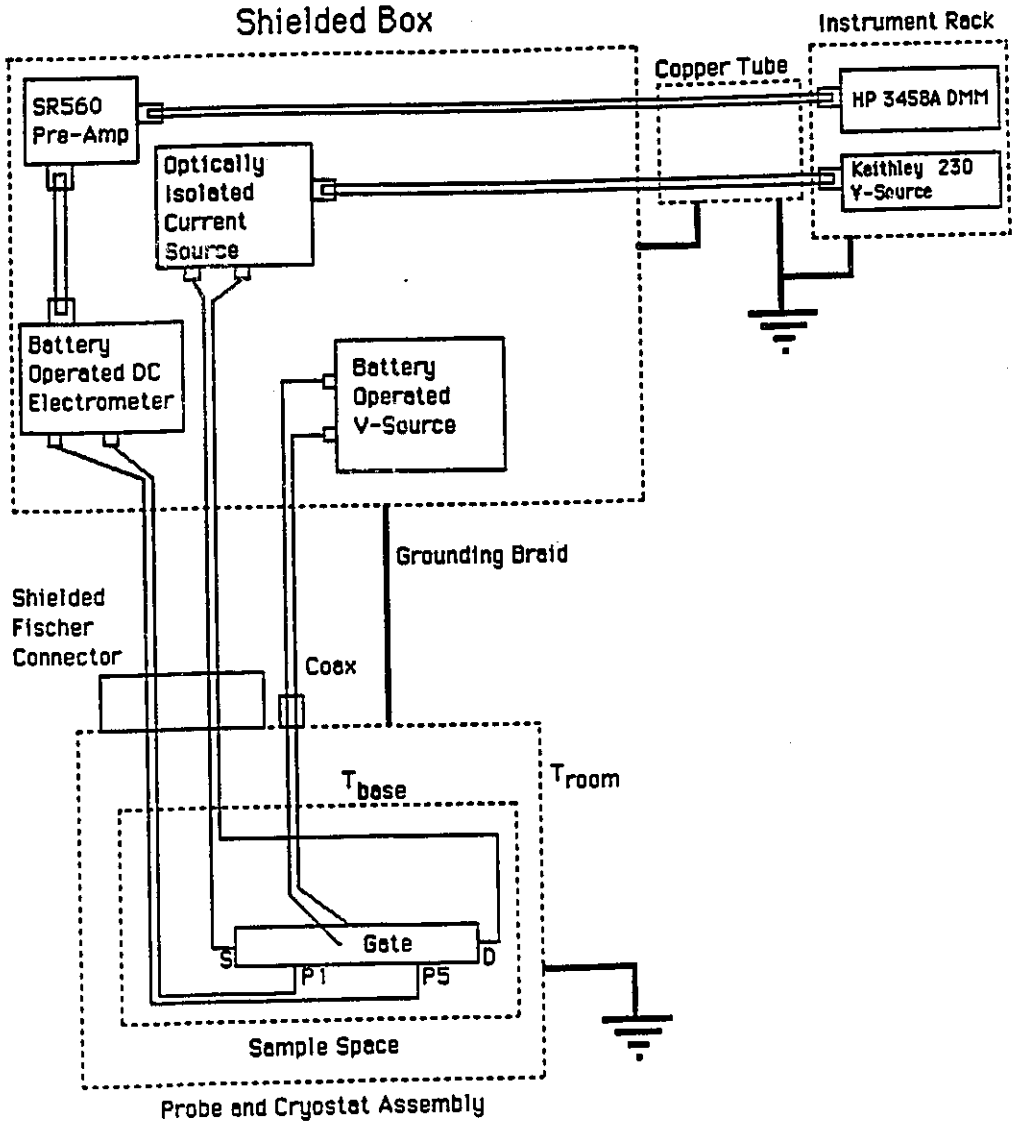

Figure 3.6: This block diagram shows the typical arrangement used to obtain an $I-V$ curve by varying the current and measuring potential.

in the graph caused by short time scale noise, and it is not centered at (I,V) = (0,0) because of the slowly varying offset potentials. The lower graph shows a similar curve taken in 1993 after the noise reduction procedures had been fully implemented. Sample Si-11 is used for the second graph because sample Si-8 was severely damaged in the Summer of 1991 and thus was not available for direct comparison, but the two samples were of comparable quality so the difference between the curves is not due to the change in samples. Obviously the lower curve is much smoother and the offset potential problem has been solved, indicating that the noise reduction techniques were effective.

3.2.4 Sample Effects

There were also two experimental considerations which were related directly to the samples themselves. The first of these was associated with Joule heating. The power injected into the two-dimensional electron system during a measurement, $P = IV = I^2R$, could become large when the sample was in its high resistance state, especially at the far end of a current sweep when I was at a maximum. If the 2DES was unable to transfer this power as heat to the bulk of the sample and from there to the refrigerator itself, the electron temperature in the 2DES would begin to rise and this could significantly change the conduction characteristics by thermally exciting electrons into conducting states. The $I - V$ curves in the low electron density regime did show a threshold field behaviour, where the differential resistance decreased sharply once the applied field (created by an applied current) exceeded a critical threshold value. This was similar to results obtained by H. W. Jiang and his coworkers [Jia91] in GaAs/AlGas heterostructures, which they attributed to Joule heating. There are several differences, however, which indicate that their measurements and ours did not occur in the same regime.

Figure 3.8 shows a graph of the threshold field, E_t, and the corresponding dissipated power, P_t, for sample Si-5 as a function of inverse Landau level filling factor (which is equivalent to magnetic field). The threshold field changes with filling factor, and so does the dissipated power, which contrasts with the results of Jiang et. al. in which the threshold occurred at a constant value of the dissipated power. It is also important to note that their threshold power was $P_t \approx 200\,\mathrm{pW}$; two orders of magnitude higher than the threshold powers observed in Figure 3.8. Thus with picoWatt peak powers being dissipated in a sample completely immersed in the liquid ^{3}He/^{4}He mixture of a dilution refrigerator with microWatt cooling power, it seemed very unlikely that Joule heating could be affecting the electrons significantly. This was verified when

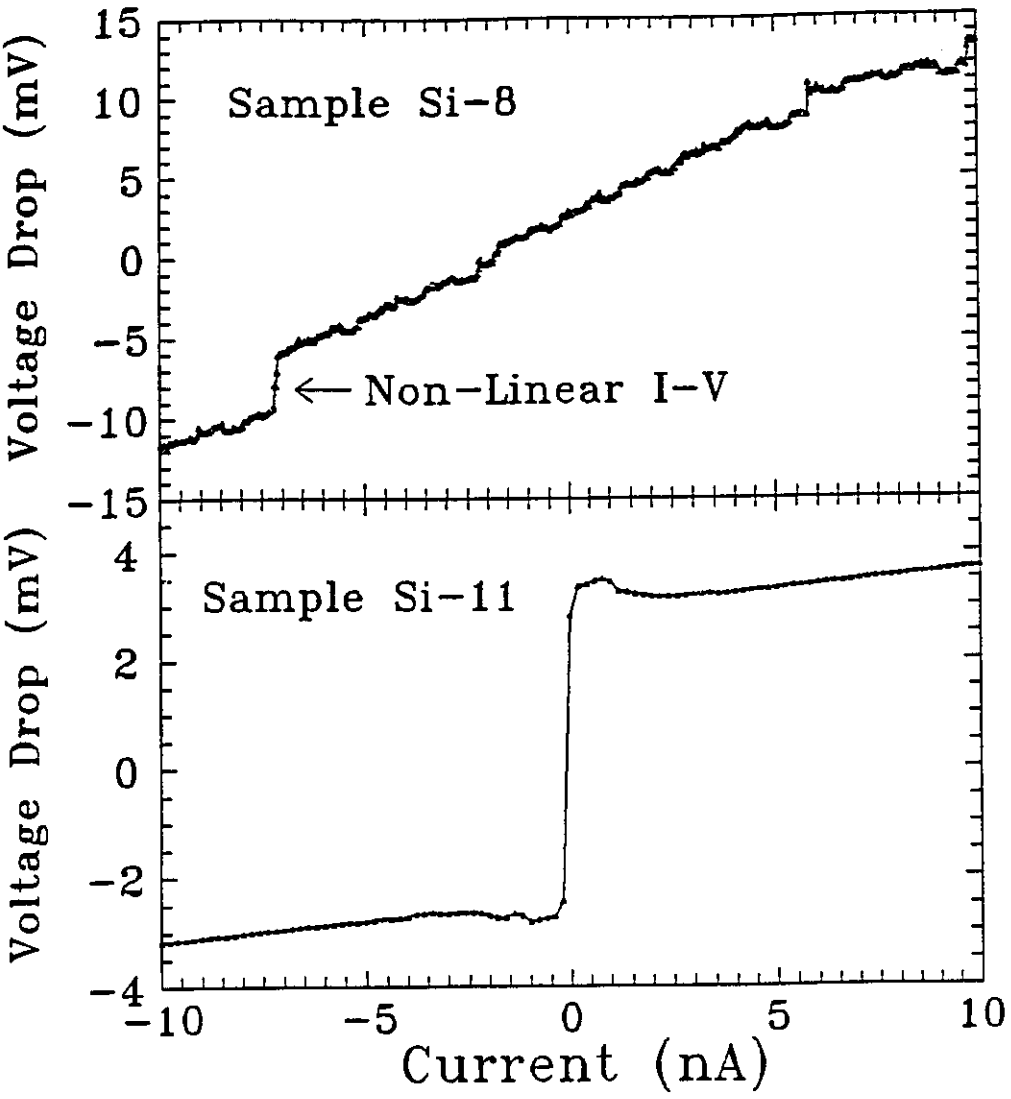

Figure 3.7: The upper $I-V$ curve in this figure was taken before the noise reduction techniques were introduced using sample Si-8 and the lower one afterwards using sample Si-11. The second curve is much smoother and the non-linearity has moved from its incorrect position in the lower left to be centred around (I,V) = (0,0).

the measurements were repeated at different temperatures and the threshold field and corresponding dissipated power for any given value of magnetic field and electron density were observed to be independent of temperature. If Joule heating was occurring, then the amount of dissipated power required to overheat the sample should have decreased as the temperature was raised. Thus, provided that very low dissipated powers were used, Joule heating was not an issue. It was necessary, however, to ensure that a stable and noise free current source was used to drive the $I - V$ sweeps in this very low current regime, which was one of the reasons why a commercial current source was not used in these experiments.

The next concern which had to be addressed was related to the contacts on the samples. All of the connections between the samples and the outside world involve two stages; the first from the two dimensional layer to the heavily n-doped region in the Si, and the second from that region to the Al film which forms the contact pad. These connections could give rise to highly resistive and possibly non-Ohmic phenomena such as Schottky barriers or tunneling junctions. The non-linearities were easy to identify in the initial characterization of the samples, so that any non-Ohmic contacts could be detected and ignored during the actual experiments. The resistance of the contacts was more difficult to dismiss, since many of the measurements used the same pair of contacts both to control the current and measure the voltage drop across the sample (two contact geometry) so that the experimentally measured resistance was actually the sum of the resistances of the 2DES and of the two contacts. One way to solve this was to use one pair of contacts to control the current (usually source and drain) and a different pair to measure the voltage (four contact geometry). Figure 3.9 illustrates a single $I - V$ curve taken from sample Si-11 using both the two contact and four contact geometries. The straight line obtained by subtracting the two contact voltage from the four contact voltage is the linear $I - V$ characteristic due to the constant resistance of contacts P1 and P5. Thus, these contacts were Ohmic and influenced the shape of the $I - V$ curve because of their contribution to the two contact voltage. Fortunately, this does not affect the experimental value of the threshold voltage, which is obtained by extrapolating the linear portion of the $I - V$ curve back to $I = 0$, because the contribution due to the resistance of the contacts is zero at zero current. Thus, even for the relatively low sample resistances shown in this figure, the contacts do not hide the existence of a non-linear $I - V$ curve or affect the value of V_t, the physically important parameter.

At lower electron densities, the sample resistances could become very high and it was observed that four contact measurements showed more noise than

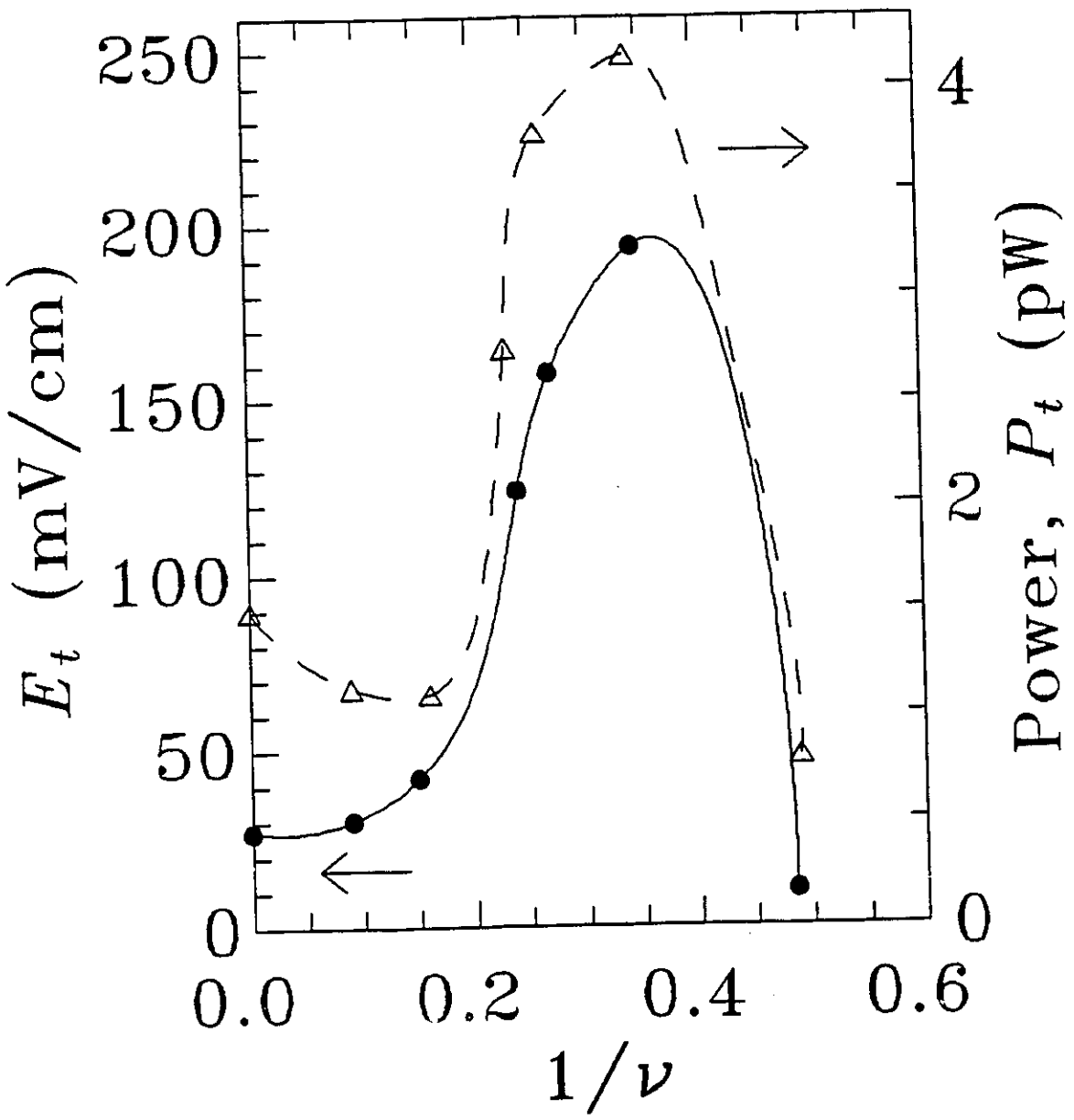

Figure 3.8: The threshold electric field, E_t, and corresponding dissipated power, P_t, are shown for several different values of the inverse filling factor, $1/\nu$. The data comes from sample Si-5.

equivalent two contact measurements. This was probably related to the extra rf pickup and possible ground loops which were inherent in using additional contacts, but in order to verify that there was no fundamental difference between experimental values obtained using the two and four contact techniques in the high resistance regime, a comparison was made of two magnetic field sweeps on sample Si-11. Figure 3.10 shows this graph of resistance, R_{xx}, versus magnetic field where the electron density is low enough that the first of the large resistance maxima which will be discussed in subsequent chapters becomes visible as the magnetic field varies between 1.1 and 1.5 Tesla. The values obtained from the two contact measurement are indicated by open circles and have error bars which are too small to show effectively on this scale. The four contact resistances are indicated by the filled circles, and their error bars are shown. It is evident that, within the limits of the error bars, the two contact and four contact techniques give the same values in both the low resistance and high resistance regimes. Considering this and the previous figure, it seems that the contacts are not significantly affecting the experimental results. A standard procedure was adopted to use two contact measurements in the highly resistive regime where noise was a concern and four contact measurements when the sample resistance was very low.

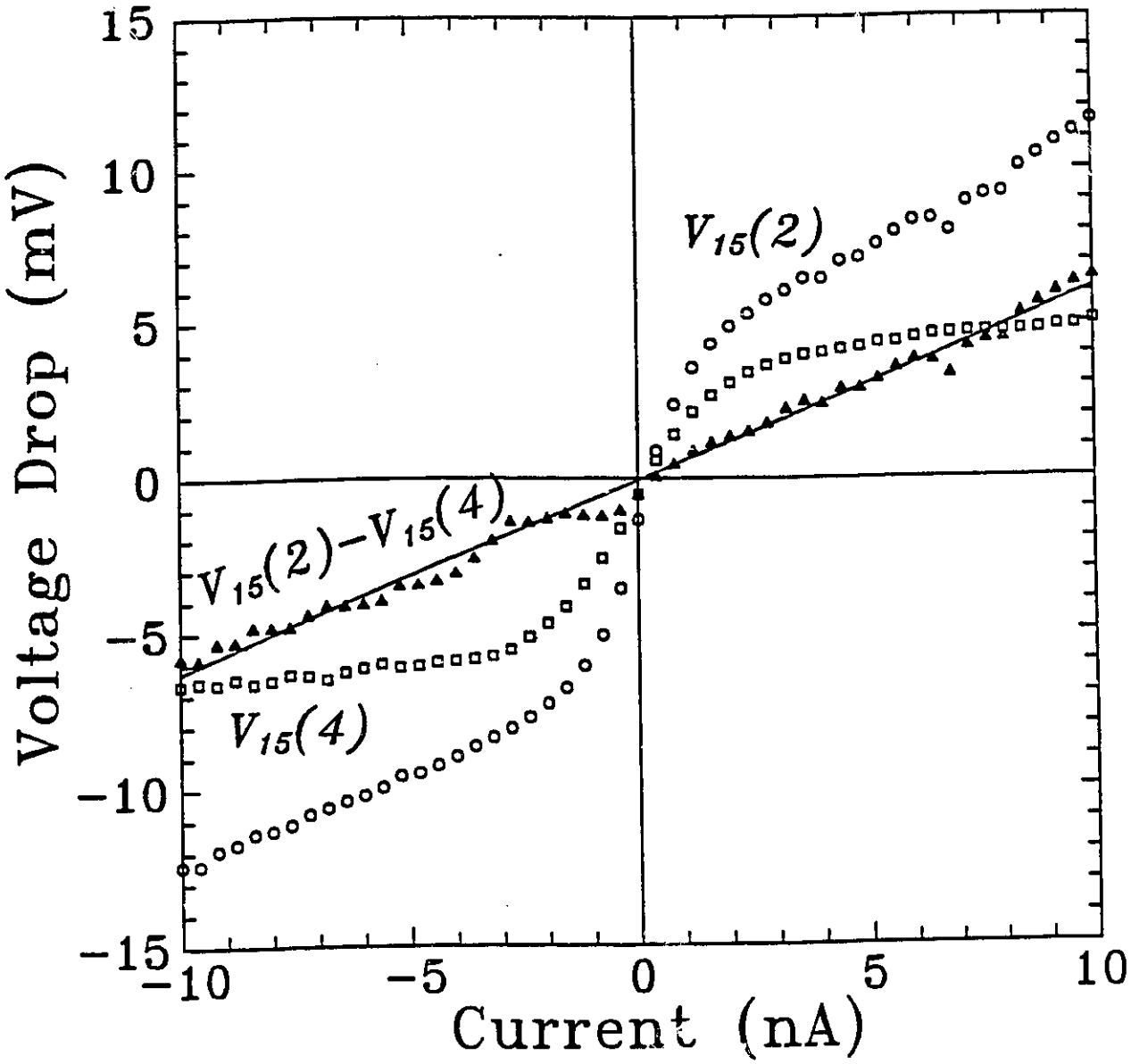

Figure 3.9: Two $I-V$ curves taken from sample Si-11 are plotted. $V_{15}(2)$ represents a two contact measurement and $V_{15}(4)$ represents a four contact measurement. The difference between them is a straight line, indicating that contacts P1 and P5 are Ohmic.

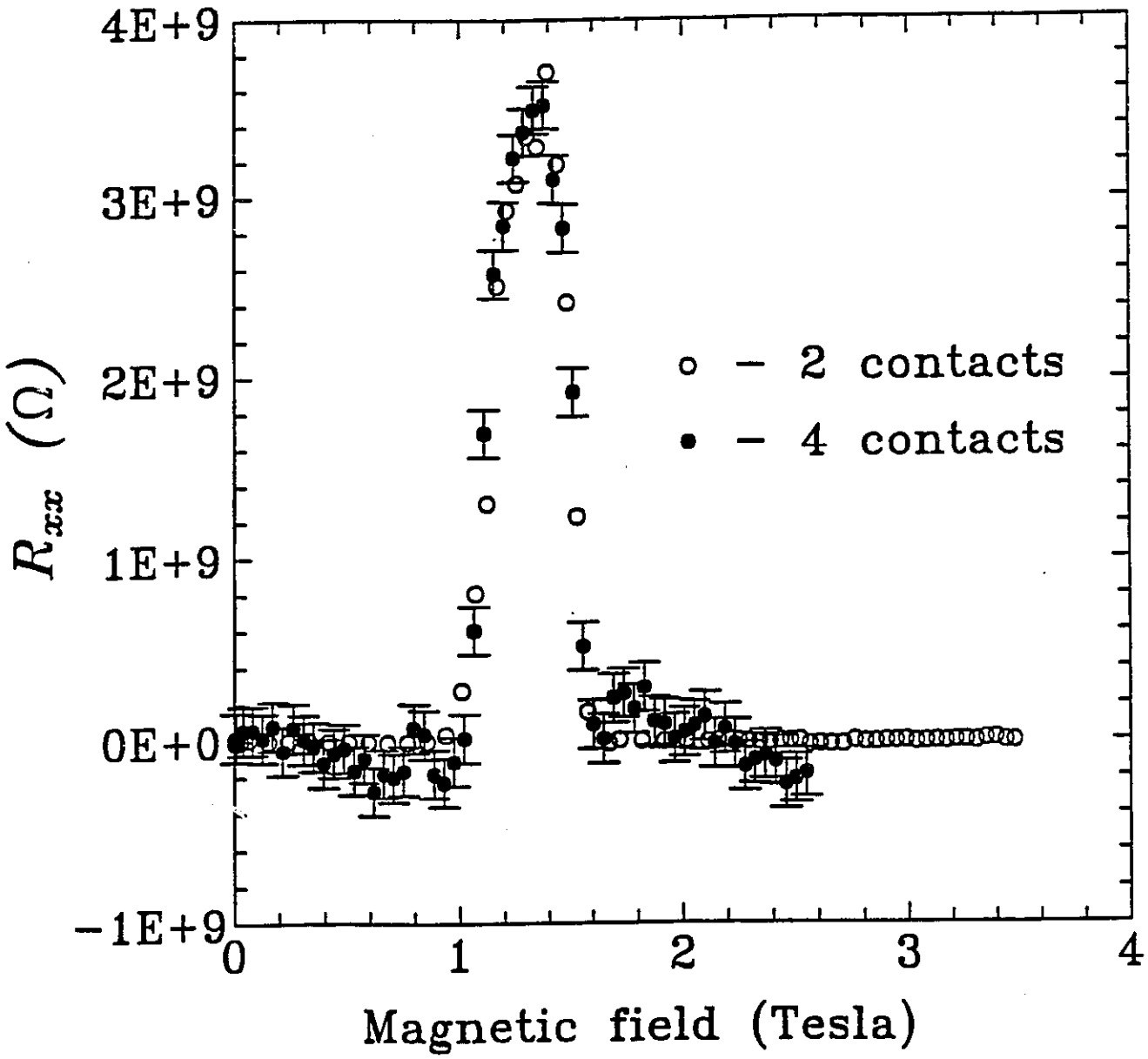

Figure 3.10: This graph of resistance versus magnetic field shows that a two contact technique (open circles, error bars too small to show) and a four contact technique (filled circles) give the same resistance values in both the low resistance and high resistance regimes.

Chapter 4. Experiments Under Magnetic Field

4.1 Introduction

This chapter describes a variety of experiments which were carried out on several extremely high mobility Russian Si-MOSFET samples between 1990 and 1993. These experiments were motivated by the observation of an unexpectedly high resistance in the direction of the applied current, R_{xx}, occurring in these samples at half-filled Landau levels when the electron density was very low. The chapter will focus on experiments carried out in the presence of a magnetic field applied perpendicular to the plane of the inversion layer in the Si-MOSFET. The measurements described are firstly those in which the sample resistance (both R_{xx} and R_{xy}) is measured as a function of magnetic field and secondly those in which the voltage between a pair of contacts on the sample is measured as the current through the sample is varied with the magnetic field held constant. Both of these measurements are carried out at different temperatures and electron densities, and each of these variables can be used in developing an understanding of the phenomenon. After presenting the basic experimental data and the temperature and electron density dependences, some of the characteristic quantities (such as thermal activation energy and critical density) which can be derived from these dependences will be discussed and the data will be interpreted within the framework of existing single particle and collective excitation models. Except for the temperature dependences, the data shown in this chapter were taken at or near base temperature in the range $20 - 35\,\mathrm{mK}$, as indicated on the individual plots.

4.2 Magnetic Field Sweep

Figure 4.1 shows R_{xx} as a function of magnetic field at various carrier densities, n_s, in units of $10^{10}\,\mathrm{cm}^{-2}$. This data was taken from sample Si-11 at constant temperature $T = 25\,\mathrm{mK}$. The Shubnikov-de Haas oscillations in the upper trace ($n_s = 11.1 \times 10^{10}\,\mathrm{cm}^{-2}$) have deep minima at filling factors $\nu = 1$ and 2, typical of the integer quantum Hall effect, and maxima whose amplitude does not exceed the classical Hall value of B/en_s. As the electron

density is decreased, however, the value of R_{xx} at the maxima starts to rise as seen in the lower traces. The IQHE induced minima at $\nu = 1$ and $\nu = 2$ are still present, as is a shallower minimum at $\nu = 6$. These minima are separated by very high R_{xx} maxima occurring at $\nu \approx 1.5$ and $\nu \approx 2.7$. The $\nu = 4$ minimum is washed out by the shoulder of the $\nu \approx 2.7$ maximum. The peak values of the maxima are in the gigaohm range while the minima are still in the $10^2 - 10^4\,\Omega$ range. At lower densities the minima also begin to rise, but even at $n_s = 6.98 \times 10^{10}\,\mathrm{cm}^{-2}$, where the $\nu \approx 2.7$ maximum has reached one teraohm, the $\nu = 1$ and $\nu = 2$ minima are still discernible. In these resistance ranges, noise becomes a concern due to the input impedance of the low-noise preamplifiers, the very low source-drain currents needed to avoid heating and the ohmicity of the contacts themselves. The smoothness of the curves indicates that noise is not a serious problem and the error bars on the lower graph fall within the size of the points. Contrary to the usual increase of the Shubnikov-de Haas oscillatory amplitude with magnetic field, the $\nu \approx 2.7$ maximum is stronger than the $\nu \approx 1.5$ maximum and becomes anomalously large at higher carrier density than its $\nu \approx 1.5$ counterpart. The appearance of anomalously high R_{xx} maxima depends on the mobility or the inherent disorder of the sample . Only samples Si-5 and Si-11, with peak mobilities of 40800 $\mathrm{cm}^2/\mathrm{Vs}$ and 33100 $\mathrm{cm}^2/\mathrm{Vs}$ respectively, show the pattern of well defined R_{xx} minima at the integer filling factors separated by extremely high R_{xx} maxima. The lower mobility Si-11 exhibits onset of the anomaly in the R_{xx} maxima at higher electron density and has higher R_{xx} peak values for a given carrier density than sample Si-5.

In order to understand the origin of the anomalous R_{xx} maxima, it is necessary to measure the Hall resistance R_{xy} at these half-filled Landau levels and also check that appropriately quantized plateaus in R_{xy} occur where minima in the longitudinal resistance R_{xx} are found. Measuring R_{xy} when R_{xx} is extremely large is quite difficult, however, because R_{xy} retains its classical value and thus can be many orders of magnitude lower than R_{xx}. Since the noise levels tend to rise with R_{xx}, long averaging times are required (up to an hour per point) when measuring R_{xy}. The contacts on the samples are not perfectly aligned and they are finite in size which produces an admixture of both R_{xx} and R_{xy} components in the resistance measured between Hall contacts. This leads to a dramatic increase in the nominal Hall resistance near the R_{xx} maxima. To solve this problem, two sets of measurements are performed with the magnetic field applied in opposite directions. Since the Hall voltage changes sign when the magnetic field is reversed but the longitudinal voltage does not, the relatively small R_{xy} component can be extracted by subtracting one set of nominal Hall resistances from the other. The result

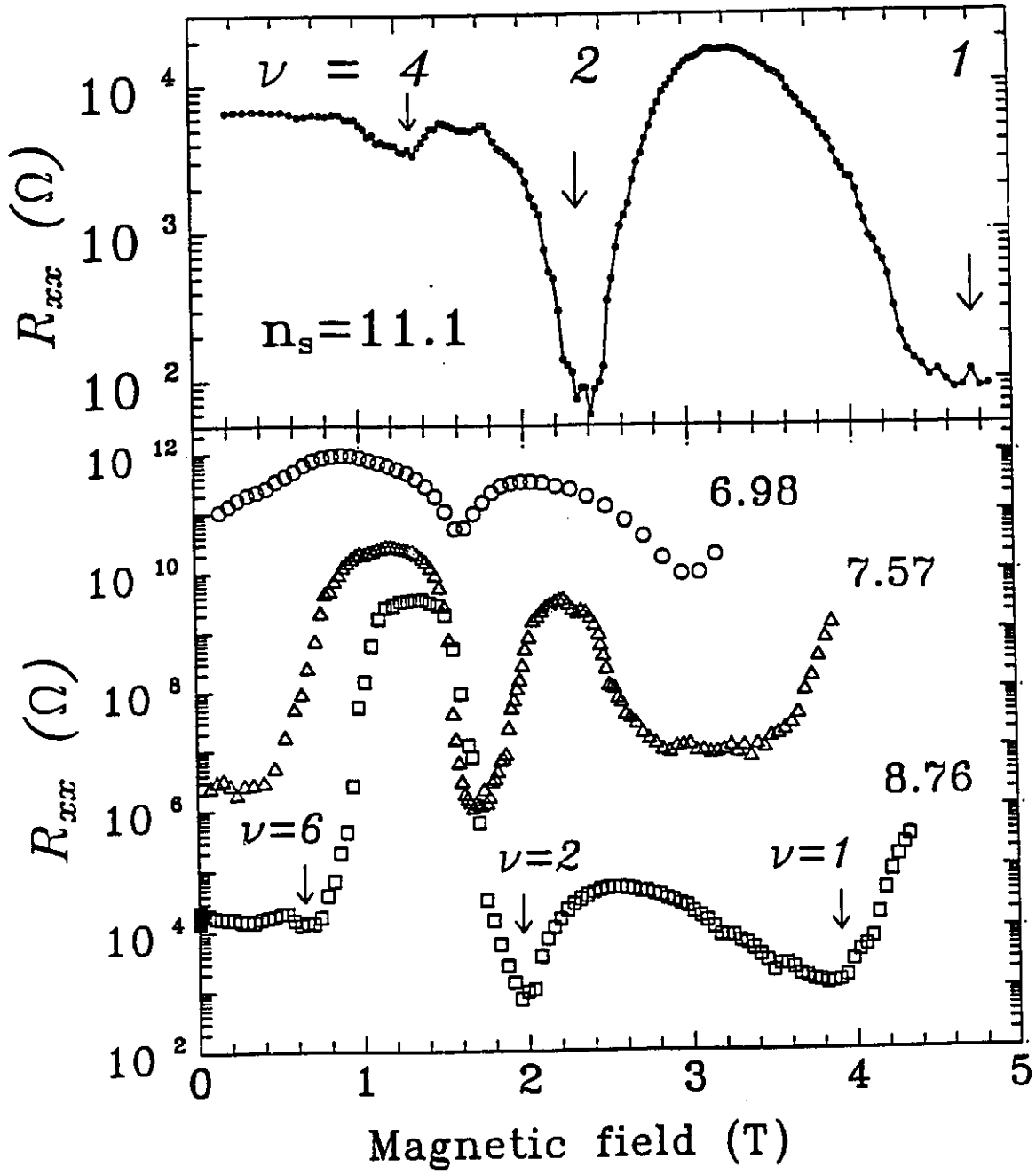

Figure 4.1: Measurements of the resistance between contacts P1 and P5 for sample Si-11 are shown as a function of magnetic field. The numbers near each curve are electron densities in units of $10^{10}\,\mathrm{cm}^{-2}$. The upper part of the graph shows the normal R_{xx} curve which is observed at high n_s, and the lower part of the graph shows three traces of the anomalous R_{xx} occurring at lower values of n_s. The arrows indicate the filling factors of the minima for the $n_s = 8.76 \times 10^{10}\,\mathrm{cm}^{-2}$ trace.

is illustrated in Figure 4.2. Here, both R_{xx} and R_{xy} are shown for sample Si-5 at a density of $n_s = 9.33 \times 10^{10}\,\mathrm{cm}^{-2}$. As expected from the integer quantum Hall effect, the Hall plateaus at $\nu = 1$ and $\nu = 2$ occur at the nominal quantized values h/e^2 and $h/2e^2$ within 1%. The Hall resistance between the plateaus follows the classical Hall line of $R_{xy}^{2D} = \rho_{xy} = B/en_s$, even when R_{xx} is at its peak values. The individual points marked with error bars correspond to measurements made at a very low current of 0.25 nA, whereas the solid line corresponds to measurements made using a current of 10 nA. The significance of these two different current values lies in the existence of two conduction regimes which will be seen later in the non-linear $I-V$ characteristics. These regimes correspond to transport both below and above a threshold electric field.

The results of a more detailed investigation of R_{xy} near $\nu = 6$ are shown in Figure 4.3. This set of measurements comes from sample Si-5 at a density of $n_s = 8.34 \times 10^{10}\,\mathrm{cm}^{-2}$. The experimental evidence for a plateau with a resistance of $h/6e^2$ (within 8%) is clear. The full line and the data points with error bars are experimental results taken at the same current values as the larger field sweeps in Figure 4.2. Having established that only the longitudinal resistance at the Shubnikov-de Haas maxima become anomalously large below some critical carrier density, it is important to study both the carrier density and the temperature dependences of this anomaly.

4.3 Density and Temperature Dependence of R_{xx}

The carrier density can be varied easily in a MOSFET by changing the bias voltage applied to the top metal gate. A measurement of the resistance at a given R_{xx} maximum as a function of the electron density requires an adjustment of the magnetic field to keep the filling factor constant. Figure 4.4 shows R_{xx} versus electron density, n_s, taken from the $\nu \approx 1.5$ and $\nu \approx 2.7$ maxima of sample Si-5. Both curves show a low resistance value at high densities which increases exponentially as the density decreases below a critical value, n_{sc}. Below $8 \times 10^{10}\,\mathrm{cm}^{-2}$, the exponential increase changes to a much slower variation. These measurements give a first indication of a critical carrier density which, for the $\nu \approx 2.7$ maximum, lies between $10 \times 10^{10}\,\mathrm{cm}^{-2}$ and $11 \times 10^{10}\,\mathrm{cm}^{-2}$. Similarly, n_{sc} for the $\nu \approx 1.5$ maximum lies between $9 \times 10^{10}\,\mathrm{cm}^{-2}$ and $10 \times 10^{10}\,\mathrm{cm}^{-2}$.

As the bath temperature is decreased, the magnitude of R_{xx} at the maxima increases, as shown in Figure 4.5. This figure shows three R_{xx} versus magnetic field curves taken from sample Si-5 at a carrier density of

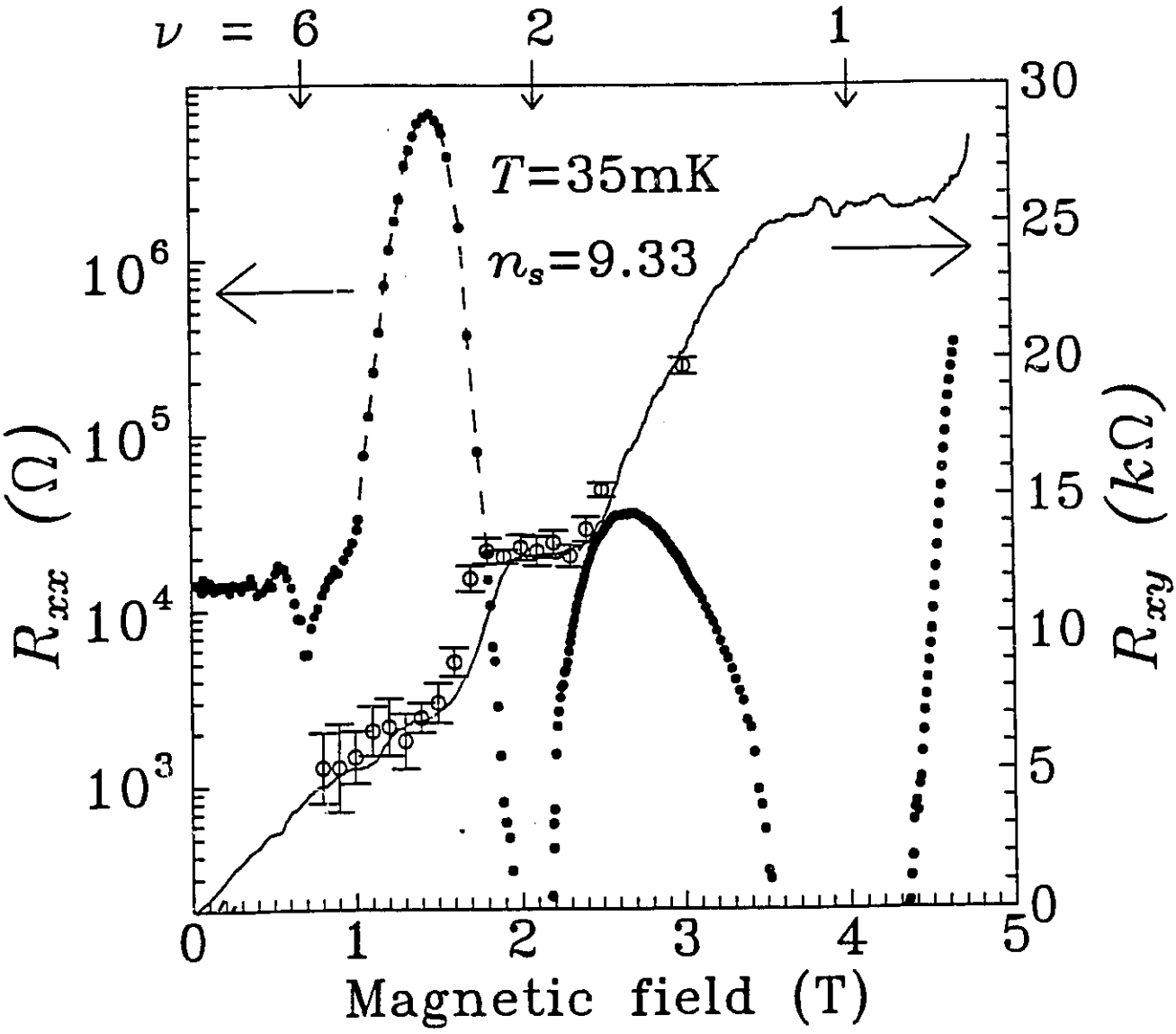

Figure 4.2: R_{xx} and R_{xy} are plotted as a function of magnetic field. The measurements come from sample Si-5 at electron density $n_s = 9.33 \times 10^{10}\,\mathrm{cm}^{-2}$. The Hall resistance near the $\nu \approx 2.7$ maximum was measured both below the electric field threshold ($I = 0.25\,\mathrm{nA}$, open circles with error bars) and above it ($I = 10.0\,\mathrm{nA}$, solid line).

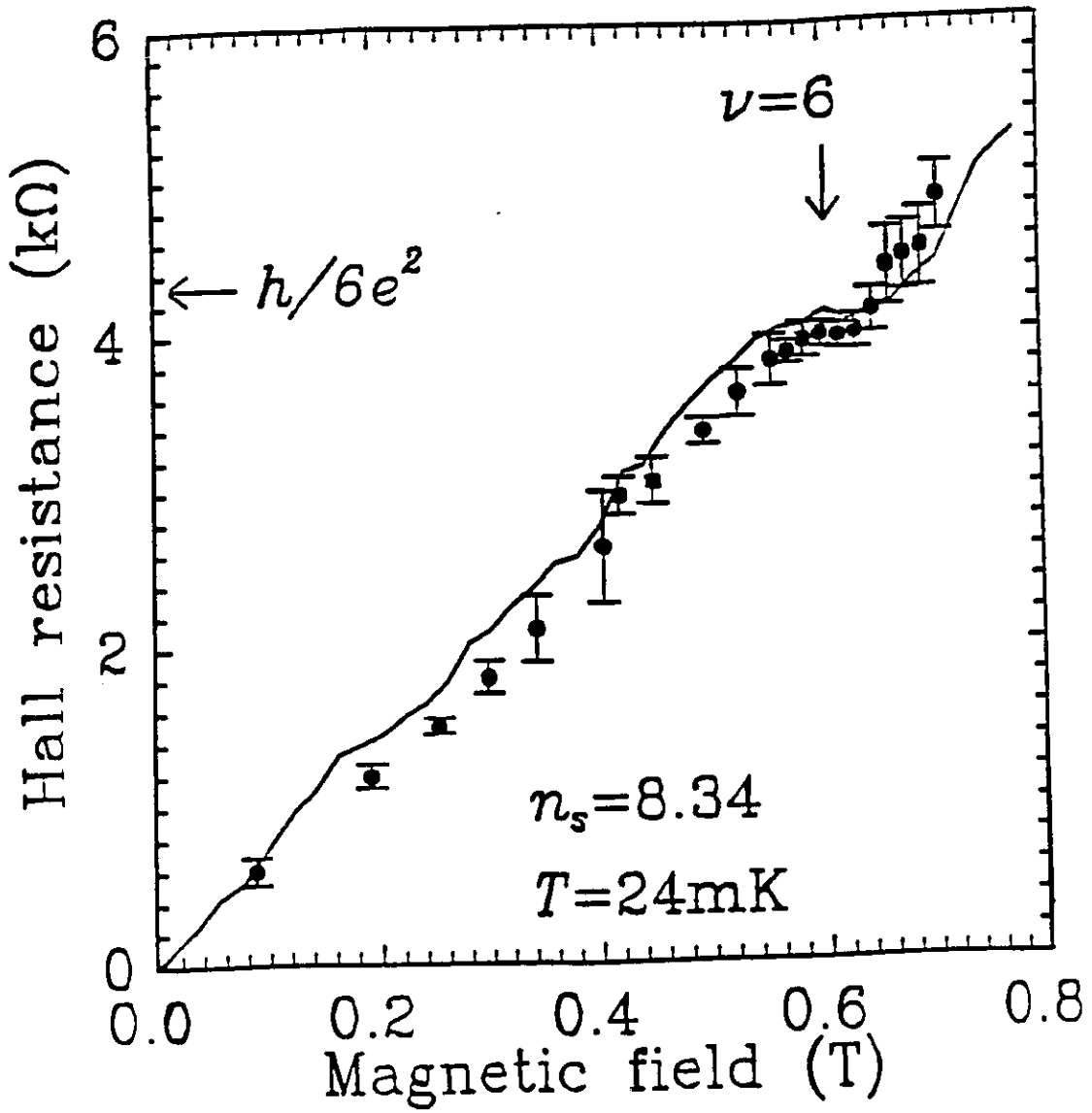

Figure 4.3: The Hall resistance is plotted as a function of magnetic field for sample Si-5 from zero field to just beyond filling factor $\nu = 6$. The electron density is $n_s = 8.34 \times 10^{10}\,\mathrm{cm}^{-2}$.

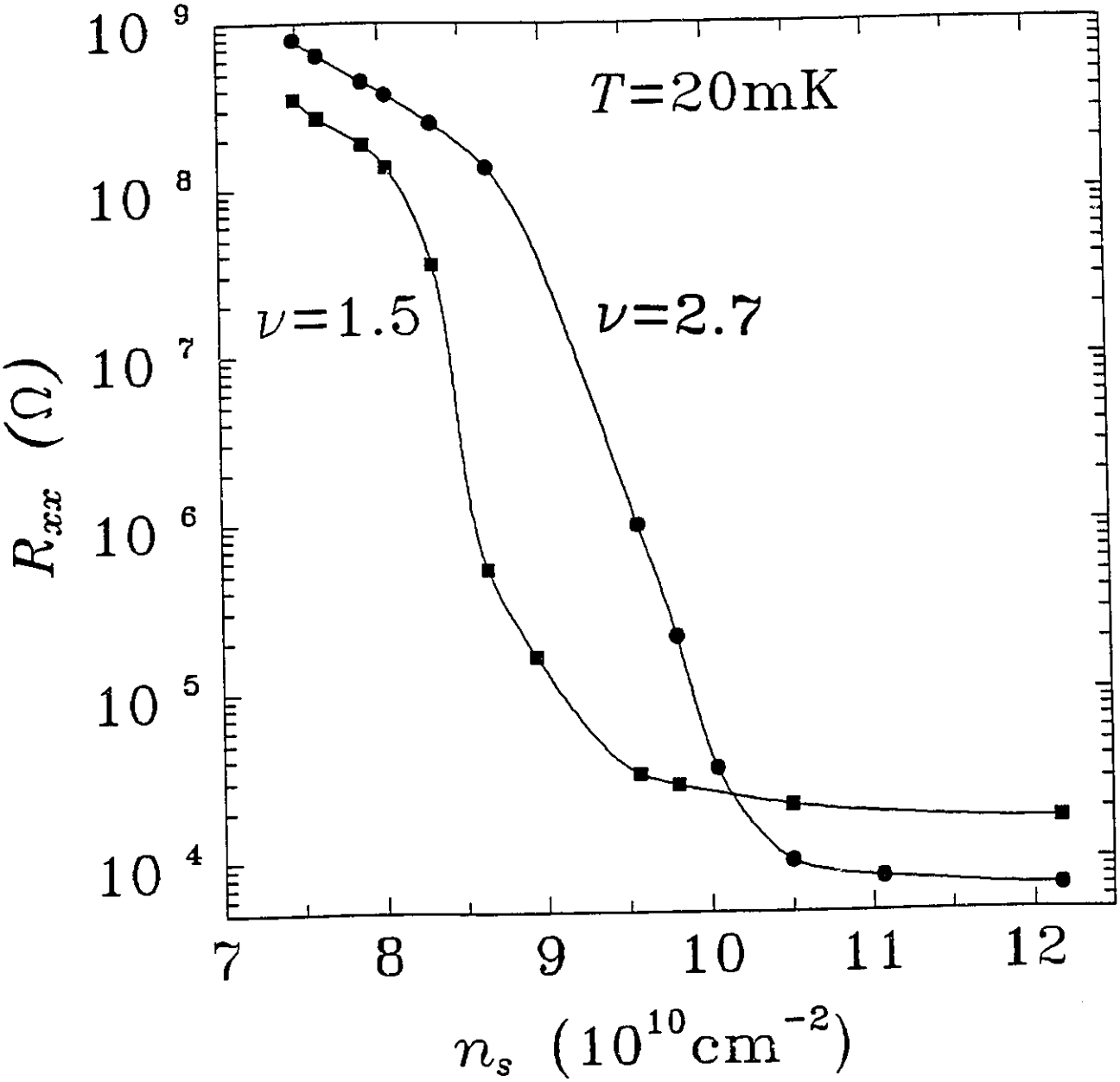

Figure 4.4: The peak resistances from the two resistance maxima at $\nu \approx 2.7$ and $\nu \approx 1.5$ in sample Si-5 are plotted as a function of electron density.

$n_s = 8.7 \times 10^{10}\,\mathrm{cm}^{-2}$ at the temperatures indicated on the graph beside each curve. The contrast between the temperature dependent behaviour at the maxima and the minima is shown more clearly in Figure 4.6, which shows three plots of ρ_{xx} versus B taken from sample Si-5 at an electron density of $n_s = 8.34 \times 10^{10}\,\mathrm{cm}^{-2}$. The temperatures for each curve are indicated on the graph. The minimum at $\nu = 6$ is quite pronounced and the resistivity at this point decreases with decreasing temperature. In the lowest temperature curve, a minimum is also visible at $\nu = 10$, but it is not sufficiently developed at base temperature to follow its temperature dependence. The maximum between these two minima is visible in all three curves, and it shows a peak R_{xx} which increases with decreasing temperature. This behavior does not agree with the normal behaviour in the high density regime, where the amplitude of the Shubnikov-de Haas oscillations increases slowly with decreasing temperature (as predicted by Equation 2.26 and as exhibited by these samples at electron densities well above n_{sc}) rather than increasing exponentially as it does here. It also contrasts with the behaviour exhibited when localization effects become dominant, in which the amplitude actually decreases with decreasing temperature.

Measuring the peak value of R_{xx} for a single maximum at a large number of different temperatures allowed a graph like Figure 4.7 to be plotted. The points on this graph are R_{xx} values from the $\nu \approx 1.5$ peak of sample Si-5 at an electron density of $n_s = 7.07 \times 10^{10}\,\mathrm{cm}^{-2}$. Since R_{xx} is a maximum (and thus $R_{xx} \gg R_{xy}$), $\sigma_{xx} \propto \rho_{xx}^{-1}$. The points can then be plotted on a logarithmic scale against inverse temperature (an Arrhenius plot), so that the straight line represents the best least squares fit of the form: $R_{xx} \propto \rho_{xx} \propto \sigma_{xx}^{-1} \propto \exp(\Delta/T)$. In this case, the thermal activation energy which is calculated from the fit is $\Delta = 0.64\,\mathrm{K}$, where the energy is expressed in Kelvin because Boltzmann's constant has been dropped from Equation 4.1 to simplify calculations. Figure 4.7 shows that this model fits the data points very well in the temperature range from approximately 160 mK to 500 mK. Lower activation energies were obtained with data taken in the low temperature range from 25 mK to 150 mK and with data taken using above threshold source-drain applied currents.

Figure 4.8 shows a plot of Δ (in units of Kelvin) versus electron density for the $\nu \approx 2.7$ maximum of sample Si-5. The best fit to the points is a straight line. This linear dependence of Δ on n_s is also observed for the $\nu \approx 1.5$ maximum in sample Si-5 and in both maxima from sample Si-11. The extrapolation of the best linear fit to zero activation energy gives the critical carrier density below which the anomalous R_{xx} maxima develop. In Figure 4.7, this occurs at $n_{sc} = 9.6 \times 10^{10}\,\mathrm{cm}^{-2}$.

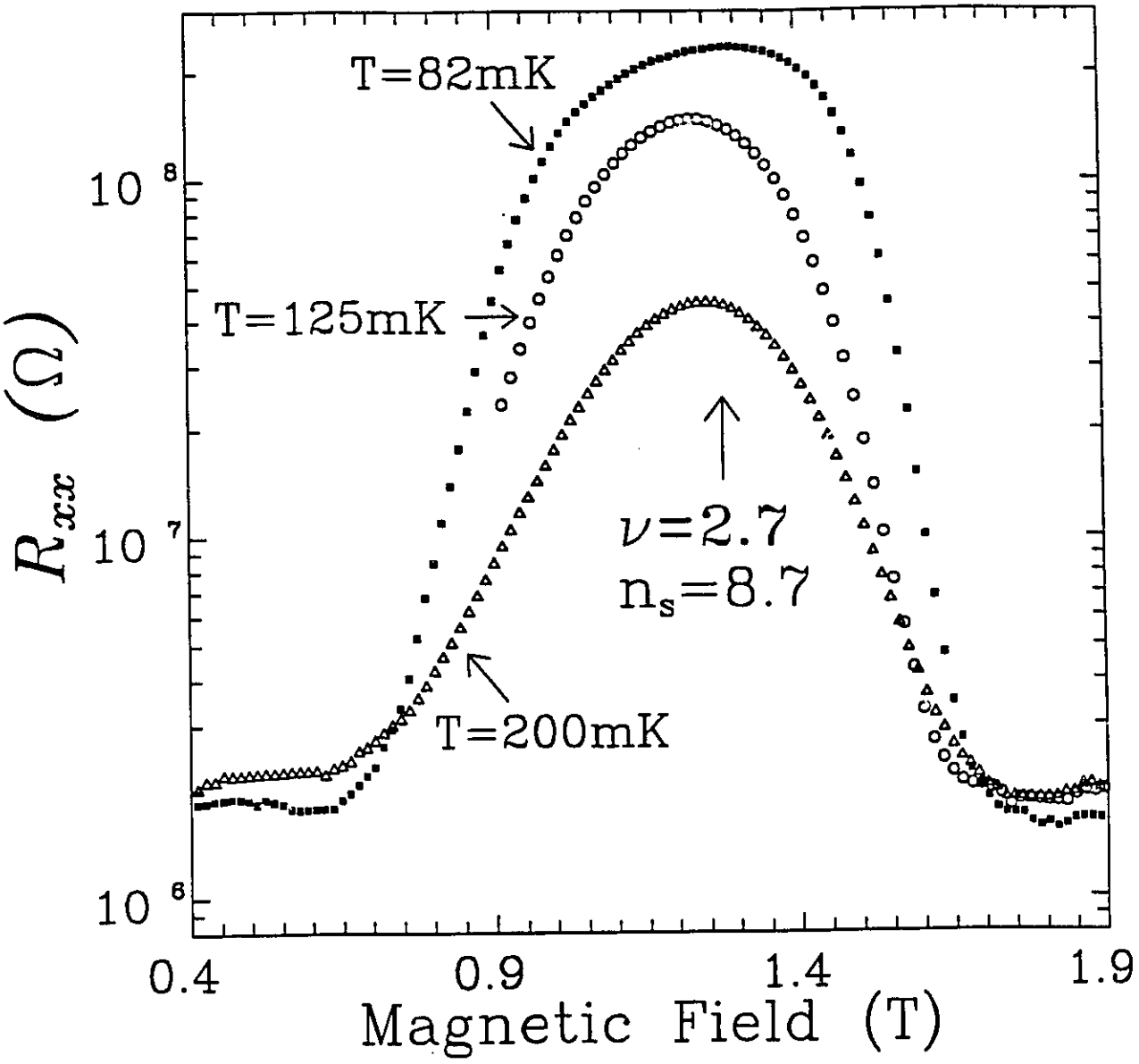

Figure 4.5: Three R_{xx} versus B curves are shown in the vicinity of the $\nu \approx 2.7$ maximum from sample Si-5 at an electron density of $n_s = 8.7 \times 10^{10}\,\mathrm{cm}^{-2}$. The temperature corresponding to each curve is indicated on the graph.

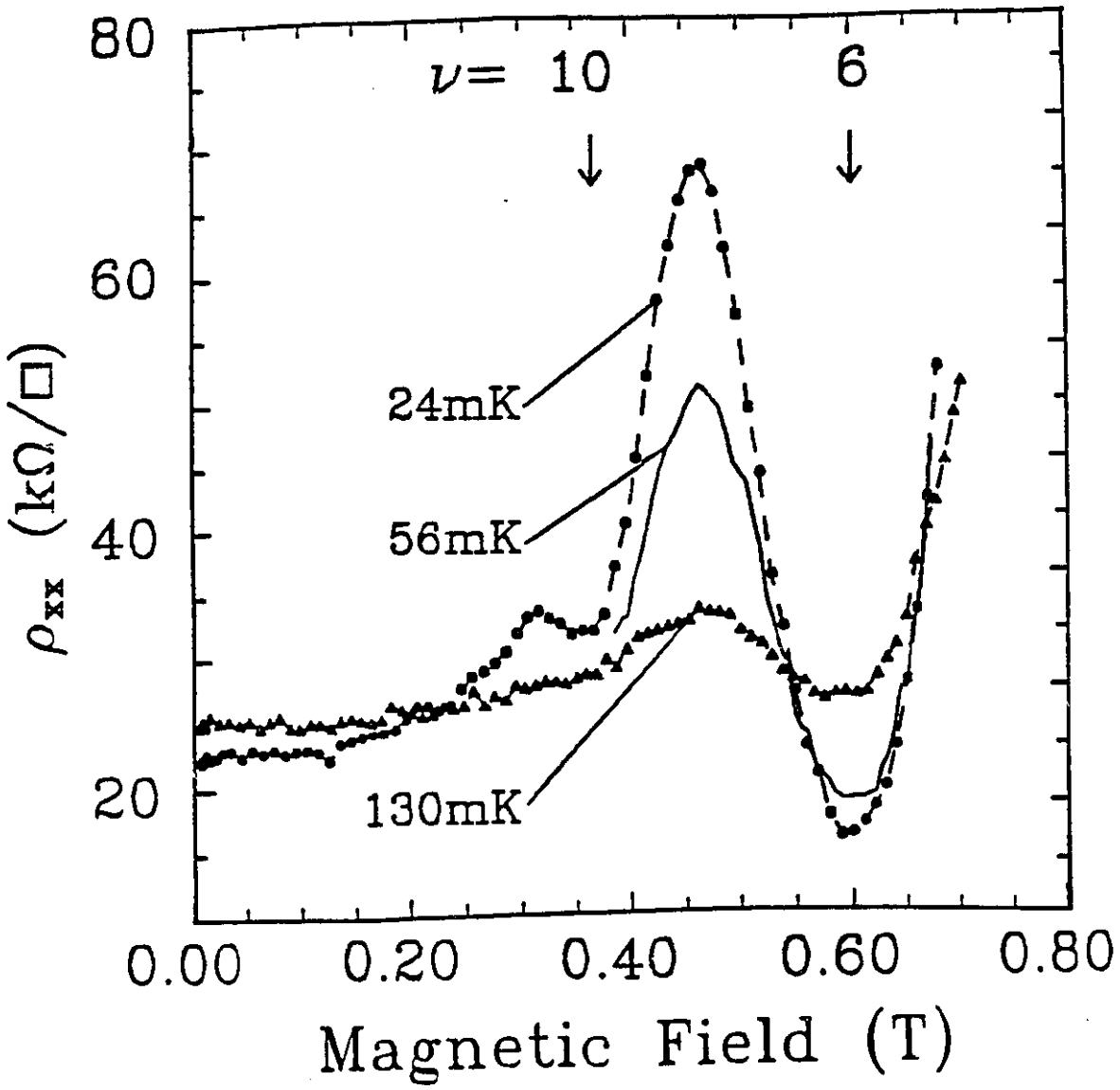

Figure 4.6: This figure shows three ρ_{xx} versus B curves in the region from zero field to just beyond filling factor $\nu = 6$. The electron density is $n_s = 8.34 \times 10^{10}\,\mathrm{cm}^{-2}$ and the temperatures are as indicated on the graph. There is a metallic minimum at $\nu = 6$ and an insulating maximum between $\nu = 6$ and $\nu = 10$.

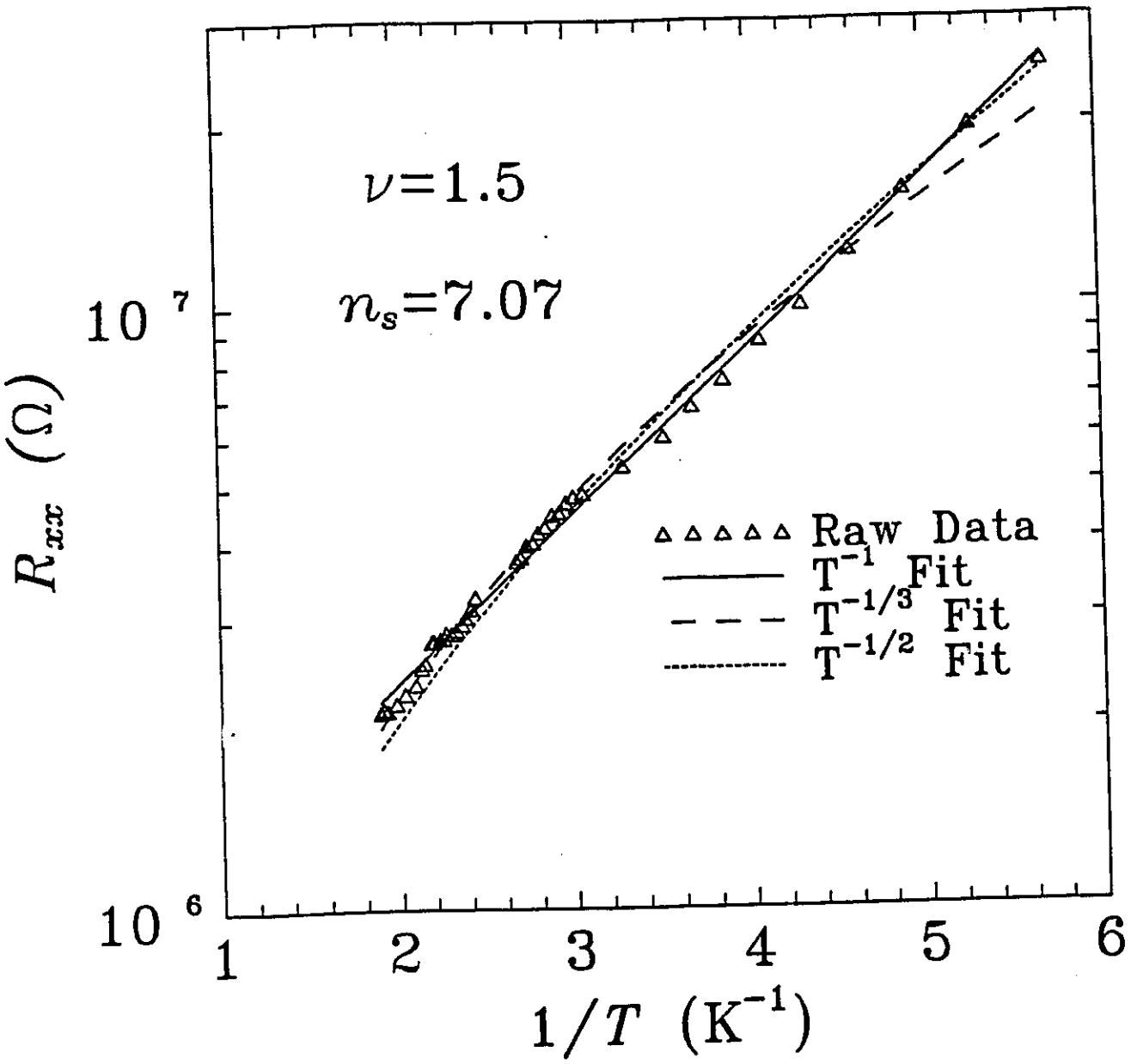

Figure 4.7: This logarithmic plot shows the peak resistance from the $\nu \approx 1.5$ maxima in sample Si-5 as a function of inverse temperature for $n_s = 7.07 \times 10^{10}\,\mathrm{cm}^{-2}$. The linear ($\rho_{xx} \propto exp(\Delta/T)$) fit is valid over the entire temperature range shown and gives an associated activation energy of $\Delta = 0.64\,\mathrm{K}$. A $T^{-1/3}$ fit and a $T^{-1/2}$ fit are also shown.

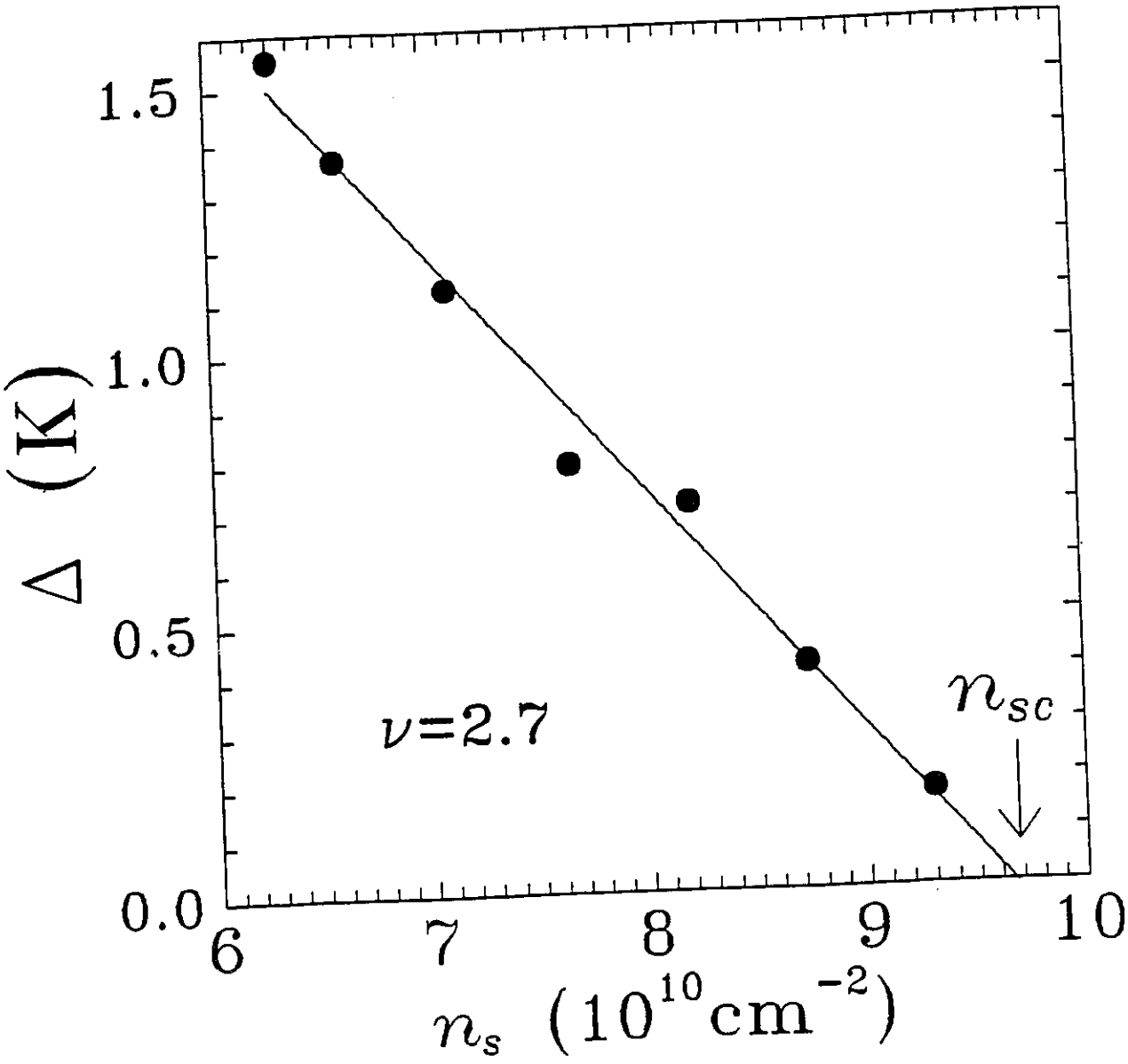

Figure 4.8: This graph shows a plot of the activation energy versus electron density from the $\nu \approx 2.7$ maximum of sample Si-5. The straight line is the best linear fit to the data and the critical density is $n_{sc} = 9.65 \times 10^{10}\,\mathrm{cm}^{-2}$.

4.4 $I-V$ Characteristics

One of the transport signatures of the CDW or the pinned Wigner crystal discussed in Chapter 2 is the non-linear $I-V$ characteristic from which a threshold voltage or field can be extracted signalling a change in the conduction mechanism. The $I-V$ curves are measured by setting the magnetic field so that R_{xx} is at a maximum for a given electron density and then varying the current and measuring the longitudinal voltage between two contacts. This yields V as a function of I. The current is swept from negative to positive values (usually with subtraction of the $I=0$ value from each point) so that dc current offsets and asymmetry effects can be easily detected if they are present.

Typical $I-V$ curves taken at the $\nu \approx 2.7$ maximum from sample Si-11 are shown in Figure 4.9, with the electron density corresponding to each $I-V$ curve indicated in units of $10^{10}\,\mathrm{cm}^{-2}$. Each curve shows a distinct threshold voltage beyond which the differential resistance decreases sharply. The $I-V$ curves appear linear both below and above the threshold, but the slopes (ie. the differential resistances) are much higher below the threshold than above it. The threshold field can be estimated by extrapolating the linear region above the threshold back to $I=0$. The current range must be selected with care as demonstrated by the inset in the figure. Extrapolating back to $I=0$ using only the inset would suggest that $V_t \approx 1.8\,\mathrm{mV}$ for this curve, whereas the main graph suggests that $V_t \approx 3.7\,\mathrm{mV}$. Thus a sufficiently wide range of current must be covered to correctly determine the threshold, while a detailed sweep of the low current range is required to correctly determine the differential resistance below threshold. As the electron density decreases, the threshold voltage increases while the differential resistance above the threshold does not change significantly.

An alternative way of determining V_t is to plot the differential resistance, $dR = dV/dI$, versus the voltage drop. The threshold voltage can be precisely identified as the point above which the differential resistance is constant. In Figure 4.10, the threshold voltage is $V_t = 3.9\,\mathrm{mV}$ and is indicated by an arrow. This figure also reveals that the differential resistance above the threshold is greater than $10^5\,\Omega$ which is much larger than the typical resistances of $10^3\,\Omega$ or $10^4\,\Omega$ observed in a two dimensional electron gas at this temperature. The Hall resistance, R_{xy}, as noted previously, is not affected by changes in the applied current and has a linear $I-V$ curve over the current range studied.

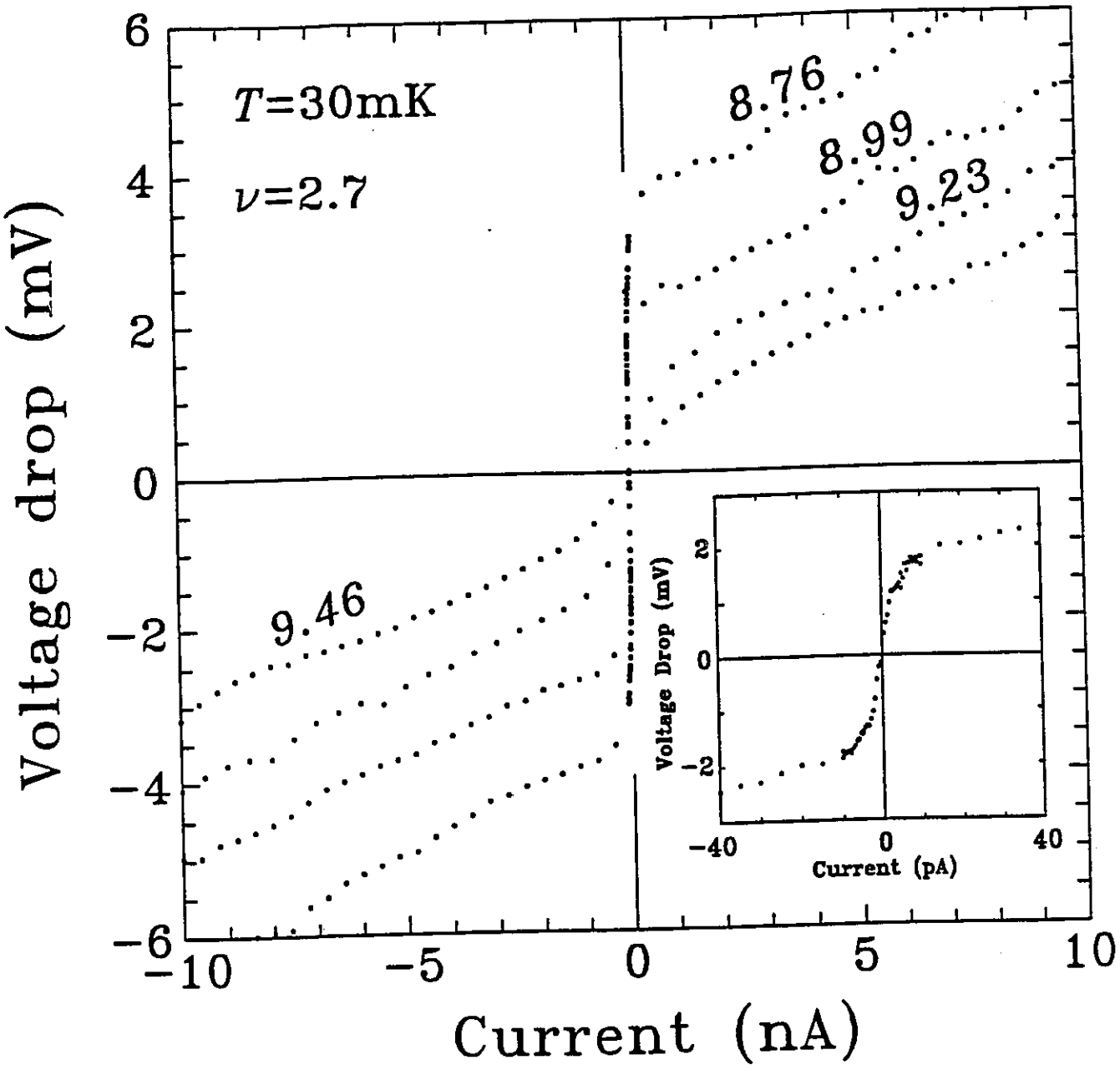

Figure 4.9: The voltage drop across contacts P1 and P5 on sample Si-11 is plotted versus the applied current at a filling factor of $\nu \approx 2.7$. The numbers beside each curve indicate the electron density in units of $10^{10}\,\mathrm{cm}^{-2}$. The inset shows an expansion of the lowest density curve near the origin.

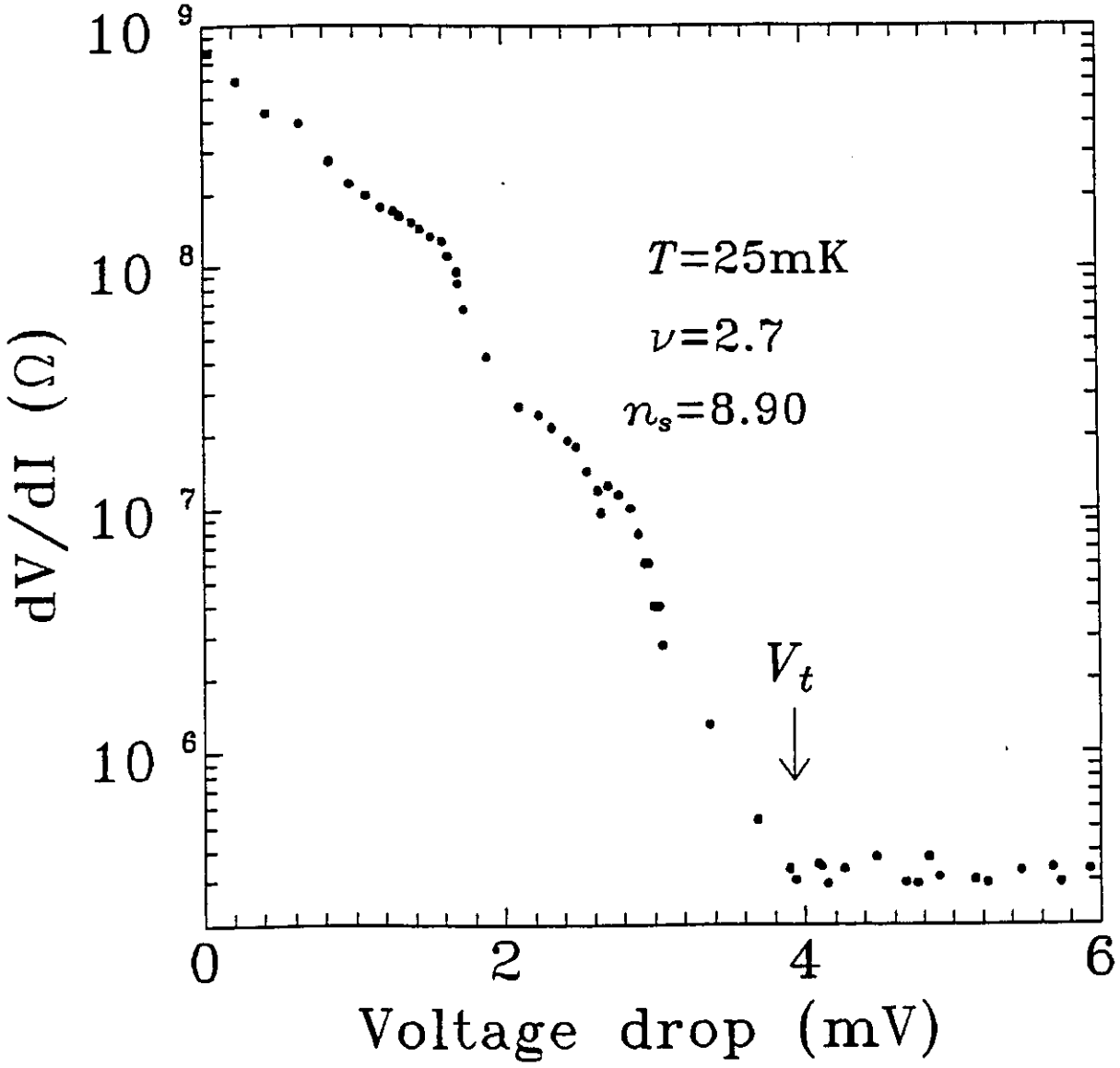

Figure 4.10: The differential resistance, $dR = dV/dI$, is plotted versus voltage drop for an $I-V$ measurement on sample Si-11 at the $\nu \approx 2.7$ maximum for electron density, $n_s = 8.90 \times 10^{10}\text{cm}^{-2}$. The threshold voltage is indicated by the arrow.

4.5 Density and Temperature Dependence of $I - V$

From Figure 4.9, the dependence of the threshold voltage on the carrier density can be inferred: the threshold voltage and thus the threshold field, E_t, increases with decreasing density. The dependence of the threshold field on carrier density is shown in Figure 4.11 for the $\nu \approx 2.7$ maximum of sample Si-11. The solid line is a third degree polynomial fit to the data which leads to the determination of a critical electron density where the threshold field becomes zero at $n_{sc} = 9.82 \times 10^{10}\,\mathrm{cm}^{-2}$. Within the accuracy of the measurement the threshold field decreases smoothly to zero at the critical carrier density.

The changes in the $I - V$ curves as the temperature is varied are also significant. Figure 4.12 shows three $I - V$ curves taken from the $\nu \approx 2.7$ maximum of sample Si-11 at electron density, $n_s = 7.75 \times 10^{10}\mathrm{cm}^{-2}$. The temperature at which each curve was measured is indicated in mK next to the curve. The lines have been drawn as a guide to the eye. With increasing temperature, the $I - V$ curves become smoother; the differential resistance below the threshold is thus smaller requiring a larger current to be applied before the threshold field can be reached. All three curves converge towards the same asymptote as the current increases meaning that the threshold field itself has not been affected by the change in temperature. Thus, decreasing temperature tends to increase R_{xx} both below and above the threshold, but it does not significantly change the threshold field value, even though it smoothes out the transition. The fact that the three curves are converging towards the same asymptote means that far above the threshold there will be very little temperature dependence in R_{xx}. This means that the thermally activated conduction mechanism dominates below the threshold and may exist as some sort of parallel conduction mechanism for a short range of potentials near the threshold, but eventually it is completely displaced by the electric field mediated conduction mechanism which creates the threshold field.

In summary, the temperature and carrier density dependences of the longitudinal resistivity and the $I - V$ characteristics indicate that there is a critical density below which an insulating phase, with its extremely high R_{xx}, occurs and that this phase features non-linear current-voltage characteristics and thermally activated conduction. The insulating phase disappears above a critical temperature of ~ 500 mK depending on the sample and the carrier density. The critical density at which the insulator appears is higher for the $\nu \approx 2.7$ maximum than for the $\nu \approx 1.5$ maximum in each sample. At any

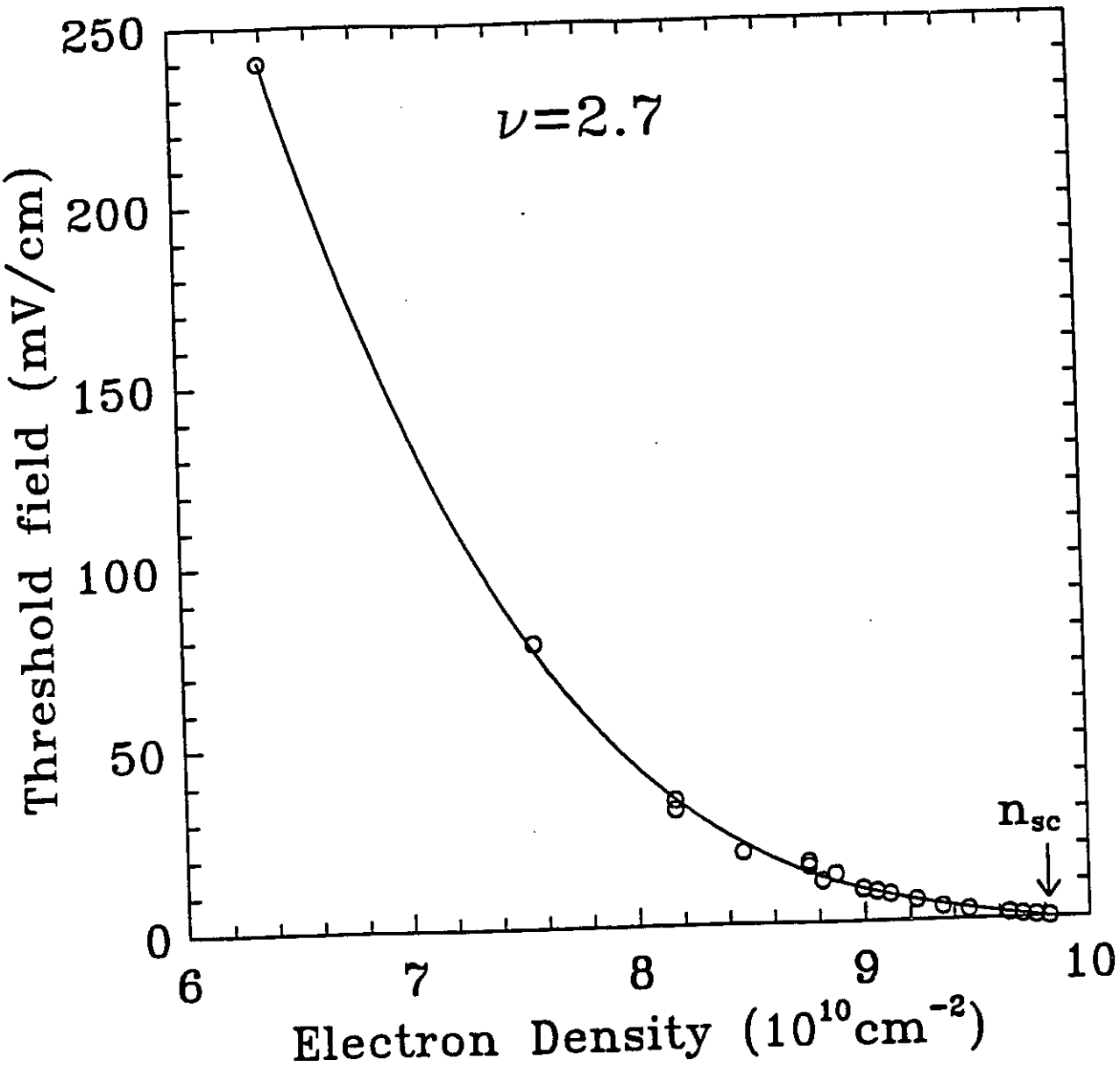

Figure 4.11: The electric field threshold, E_t, is shown for several values of electron density at the $\nu \approx 2.7$ maximum of sample Si-11.

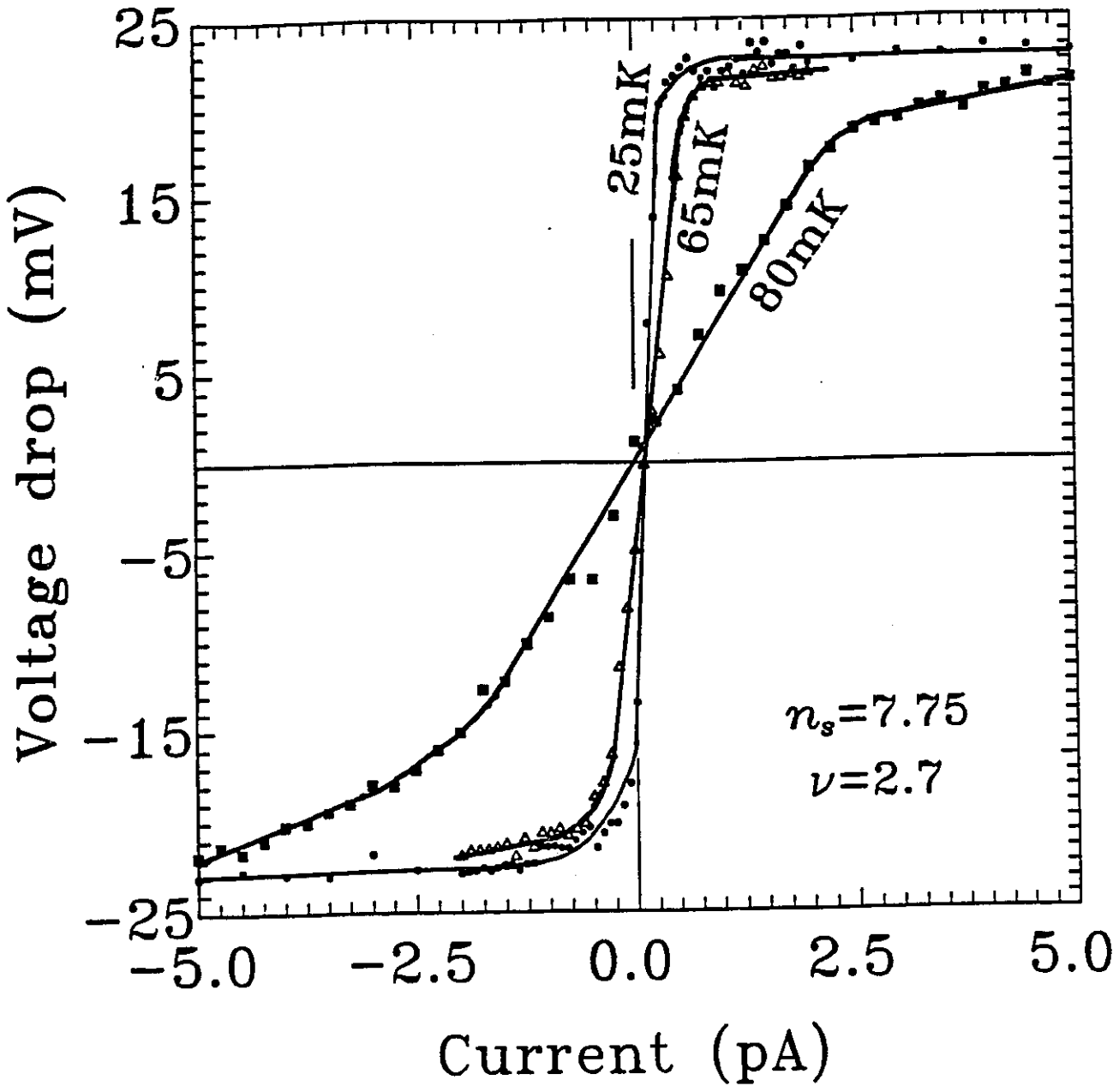

Figure 4.12: Three $I-V$ curves for sample Si-11 at filling factor $\nu \approx 2.7$ and an electron density of $n_s = 7.75 \times 10^{10}\text{cm}^{-2}$ using contacts P1 and P5 are shown. The temperatures corresponding to each curve are indicated on the graph.

given filling factor, the critical density is lower for the higher mobility sample, Si-5, than for sample Si-11. The non-linear $I-V$ characteristics and the thermally activated transport do not vanish immediately if the filling factor deviates slightly from one of the maxima, but the critical carrier density decreases as the magnetic field moves away from the R_{xx} maximum and towards an integer quantum Hall effect minimum. For a given carrier density below n_{sc}, the activation energy and the threshold field decrease smoothly as the filling factor deviates from the position of one of the R_{xx} maxima.

4.6 Discussion and Analysis

The experimental data show that below a critical carrier density, n_{sc}, the magnitude of the Shubnikov-de Haas maxima at half filled Landau levels between filling factors ν=1 and 2, 2 and 6, and 6 and 10 increase exponentially with decreasing density and temperature. This behavior is typical of an insulating state with $R_{xx} \gg 10^6\,\Omega$ and $dR_{xx}/dT < 0$. The minima corresponding to the quantum Hall effect at $\nu = 1, 2$ and 6 exhibit metallic behavior, however, becoming deeper (or closer to zero) with decreasing temperature and having $R_{xx} < 10^3\,\Omega$ and $dR_{xx}/dT > 0$. This means that as the magnetic field is swept, the magneto-transport data shows a set of reentrant metal-insulator transitions. Some of the questions which must be addressed are whether these transitions are single particle or collective in nature and whether the data are consistent with known conduction mechanisms in the localization regime.

4.6.1 Review of Localization

The first possibility to consider is that some form of localization is the underlying mechanism for the metal-insulator transition at low magnetic fields. One question which must then be addressed is whether or not localization is likely to occur at half filled Landau levels. At filled Landau levels, where the quantum Hall effect is observed, the electron gas can be treated as a non-interacting ensemble of carriers. The Landau levels are broadened by disorder due to impurities so that the density of states is not composed of delta functions but is Lorentzian in shape with a narrow band of extended, current carrying states at the center of each Landau level where the density of states is a maximum and a tail of localized, non-current carrying states (a mobility gap) where the density of states is a minimum between Landau levels. The presence of some disorder is crucial to the observation of the

quantum Hall effect since otherwise there would be no states in the gap between Landau levels. When the Fermi energy lies in the mobility gap, the current carrying states are filled below the Fermi energy and empty above it. There can be no elastic scattering of the carriers and the current flow is dissipationless despite the presence of disorder. Thus the longitudinal resistance is vanishingly small and the Hall resistance is quantised as $h/\nu e^2$ over a plateau whose width depends on the sample disorder. At finite temperature there is a small longitudinal resistivity due to hopping processes between localized states at the Fermi level. As the Fermi level passes through the core of extended states at the center of the Landau level, the longitudinal resistance becomes appreciable and the Hall resistance changes from one plateau to another. Thus the quantum Hall effect is understood in terms of localization of carriers by the impurity potential near filled Landau levels. This localization is made possible by the lack of screening which exists if a Landau level is fully occupied. It is therefore more likely for an insulating state to occur at a completely filled Landau level where there are fewer current carrying states than at a half filled Landau level where current carrying states are usually available and where the screening of impurity potentials should be effective. The experimental observation of a reentrant insulating state which occurs only at half filled Landau levels is therefore unexpected if it is to be explained in terms of a disorder induced single particle localization state.

The onset of the insulating phase at half filled Landau levels occurs below some critical carrier density, n_{sc}, which varies depending on the temperature, the filling factor and the sample mobility. A decrease in the density means an increase in disorder as the ratio of the carrier density to the fixed density of impurities becomes smaller. In samples of mobility lower than 25000 cm^2/Vs, the reentrant metal-insulator transitions are not observed. In the higher mobility samples where the reentrant metal-insulator transitions are observed, the increase in the longitudinal resistance at half filled Landau levels is more pronounced and occurs at higher carrier densities in the lower mobility (higher disorder) samples. This means that increasing sample disorder increases the strength of the insulating state, but above some critical disorder the phenomenon disappears. This is also unexpected in the framework of disorder induced localization, where increasing disorder should always increase the strength of the insulator.

Both Laughlin and Khmel'nitzkii [Khm83, Lau84] have suggested that with increasing disorder, the Landau level broadening exceeds the Landau level separation at low magnetic fields and the extended states become localized as their energy exceeds that of the Fermi level. The result is an insulating state. As the magnetic field is increased the Landau level separation increases

such that the extended states can sink below the Fermi level and return to the center of the Landau levels as a transition to the quantum Hall effect takes place. Such magnetic-field induced delocalization has been observed in disordered GaAs/AlGaAs heterostructures [Jia94] where the sample is in an insulating state at zero magnetic field. This process should be accompanied, however, by some anomaly in the longitudinal and possibly the Hall resistances as the extended states sink below the Fermi level. In the Si-MOSFET samples, the Landau levels are well separated at the fields where the reentrant metal-insulator transitions are observed and no evidence of the passage of extended states through the Fermi level is observed in either the Hall or the longitudinal resistance. As the magnetic field is further increased, the Landau level separation continues to increase and this mechanism should not produce any subsequent reentrant transitions. In the Si-MOSFET samples, however, there is clearly more than one reentrant metal-insulator transition as the magnetic field is increased.

Magnetic freeze out is another mechanism leading to an insulating state where carriers are localized as their magnetic length shrinks with increasing magnetic field. This localization depends on the magnetic field and both the carrier and impurity densities. Magnetic freeze out usually sets in quickly and permanently and the release of extended states with increasing magnetic field is unlikely. Thus magnetic freeze out can produce a metal-insulator transition, but not a reentrant metal-insulator transition in which the insulator is removed by increasing the magnetic field.

It is possible in very inhomogeneous samples to have long range variation in the impurity potentials such that carriers separate into individual macroscopic conducting regions separated by potential hills. It is conceivable that a highly inhomogeneous sample could display a series of reentrant metal-insulator transitions, since the carrier density would be different in different parts of the sample and thus they could become localized or delocalized at different values of the magnetic field. Since the Si-MOSFETs used in these experiments were patterned in Hall bars with lengths of several millimeters, the homogeneity issue is valid, and this must be taken into consideration when analyzing the experimental data.

As discussed in Section 2.3, there are two common approaches to localization, one developed by Mott and the other by Anderson. In the Mott approach, an electron at a donor impurity site has a large Bohr radius which encompasses many lattice sites of the underlying impurity sublattice. At low electron density, Coulomb repulsion keeps the electrons far apart and there is negligible overlap of their wavefunctions. The sample is thus insulating at zero temperature. At high density, the electrons overcome the Coulomb

repulsion so that there is wavefunction overlap and the material is metallic in its behavior. Since the scattering in Si-MOSFETs is mostly short ranged and thus the individual impurities will each generate separate potential fluctuations with different strengths, Mott localization, which requires an ordered sublattice of impurities is not likely here. The other approach is that of Anderson, where the electrons see random impurity potentials. This approach does not rely on any electron-electron interactions and allows each electron to be localized at a different impurity site, so it is a case of true single particle localization. For low electron density, the energy spread in the random potentials of the disordered system is large compared to the energy bandwidth and the material is insulating. The extended states are present, but are separated in energy from the localized states by a mobility edge. The metal-insulator transition occurs when enough electrons are added to push the Fermi level past the mobility edge from the localized to the extended regime. This occurs at the critical electron density which separates the high density metallic regime from the low density insulating regime. Thus both of these models do predict a critical electron density above which the insulating state becomes metallic.

In Section 2.3.3, the low temperature conductivity in Anderson localization was shown to have a distinctive temperature dependence caused by electrons trapped at impurity sites contributing to conduction through phonon assisted tunneling to states of similar energy. As the temperature is decreased, the resistance between neighbouring impurity sites increases and exceeds the resistance between remote sites with energies close to the Fermi level. Thus the distance the electron must tunnel through to get to a state within kT of the Fermi energy varies, giving this process the name "variable range hopping". For hopping in two dimensions with a constant density of states, the longitudinal resistivity varies with temperature as $\rho^0_{xx}(T) \propto \exp(T_0/T)^{1/3}$. If there is a Coulomb gap in the density of states, however, then the density of states in two dimensions increases linearly in energy away from the Fermi level and the temperature dependence of the resistivity is of the type $\rho^1_{xx}(T) \propto \exp(T_1/T)^{1/2}$. At higher temperatures conduction occurs through a direct excitation of particles to the percolation level at the center of the Landau level and $\rho^2_{xx}(T) \propto \exp(\Delta/T)$.

As single particle localization is impurity mediated, it is reasonable to expect that current-voltage characteristics will be non-linear since, in principle, the applied field can be made large enough to overcome the binding energy of the impurities. The threshold field at which this unbinding occurs can be estimated using the activation energy, Δ, obtained from the temperature dependence of the longitudinal resistance together with the electron mean

free path, l_i, and is given by $E_t^{sp} \sim \Delta/el_i$. Since the Ioffe-Regel criterion [Iof60] states that l_i must be smaller than the de Broglie wavelength, λ_{dB}, for strong localization to occur, this places a lower limit on E_t. For a typical activation energy of 1.0 K and assuming a temperature of $T \approx 100$ mK, the minimum field necessary to overcome the binding energy of impurities and depin electrons is $E_t^{sp} \approx 100$ V/m.

Another distinguishing feature of metal-insulator transitions is the behavior of the Hall resistance or the Hall conductivity in the insulating phase. Using scaling theory, Shapiro and Abrahams [Sha81] found that, in a disordered non-interacting two dimensional system, both conductivity components vanish at the critical point with $\sigma_{xy} \propto \sigma_{xx}^2$ so that the Hall resistance retains its classical value at the metal-insulator transition. Viehweger and Efetov [Vie91] arrived at the same result by considering a disorder induced metal-insulator transition for non-interacting electrons in the lowest Landau level and in the low frequency limit where $\sigma_{xy} \propto \sigma_{xx}^2 \propto \omega^2$. The Hall resistance remains finite in the insulating phase when $\omega \to 0$ and $\rho_{xx} \to \infty$. This is in contrast with the Mott insulator where the Hall resistance diverges. Kivelson and his coworkers [Kiv92] have coined the term "Hall insulator" to describe the behavior of electrons in the limit of zero temperature where $\rho_{xx} \to \infty$, $\sigma_{xx} \to 0$ and $\sigma_{xy} \to 0$ with ρ_{xy} retaining its classical value of B/en_s.

In considering the experimental data, all of these characteristics will have to be compared with the experimental results. It will also be useful to remember the energy scales for electrons in a silicon inversion layer. The cyclotron energy is 3 meV at $B = 5$ T and for a typical electron density of $n_s = 8 \times 10^{10}\,\mathrm{cm}^{-2}$ the Fermi energy is 0.5 meV and the Coulomb energy is 9 meV. All of these are much larger than the typical thermal energies of less than 0.1 meV for temperatures below 1 K.

4.6.2 Single Particle Localization and the Data

A comparison between the experimental data and the expectations of the single particle localization model will now be made. The anomalous longitudinal resistance occurs at the Shubnikov-de Haas maxima or at half filled Landau levels where a band of extended states resides and where the density of states is a maximum. Localization should be most effective at filled Landau levels where additional electrons would not be screened from an impurity potential. The insulating phase, however, sets in at half filled Landau levels and disappears at filled Landau levels where the quantum Hall effect sets in. If the insulating phase is mediated by single particle localization, then the number of charge carriers should decrease as electrons become bound at

impurity sites and are lost to the conduction process. The carrier density can be experimentally determined from the position in magnetic field of the middle of the Hall plateaus or the minima in the longitudinal resistance. As carriers disappear the minima should shift to lower magnetic fields. In Figure 4.13, the carrier density is shown as a function of the position in magnetic field of the centre of the R_{xx} minima for several different filling factors. The data was taken from sample Si-11 at a temperature of 24 mK. The straight lines through the sets of points indicate that the ratio of electron density to magnetic field is constant for a given filling factor. At filling factor $\nu = 1$, however, there is a shift of the resistance minima towards higher magnetic field below $n_s \approx 1 \times 10^{11}\,\mathrm{cm}^{-2}$. This signals an increase rather than the expected decrease in the carrier density. More recent measurements have revealed the same upward shift near filling factors 2 and 6 [Chu94] and an interpretation of this data will be given in Section 4.6.3. Thus the behavior of the carrier density is not characteristic of single particle localization. Moreover the density of impurities which are responsible for the scattering in this density regime is about an order of magnitude less than the carrier density in the 2D electron system and so the single particle localization limit of one impurity per carrier is never achieved in these samples. Another scenario to account for the large resistance is to consider the possibility that impurity scattering leads to a very short electron lifetime, τ_q, so that the Landau levels are broadened so much that they overlap ($\Gamma = \hbar/2\tau_q$) and the extended states float up above the Fermi energy. The well-defined Shubnikov-de Haas minima and quantum Hall plateaus observed experimentally, however, are characteristic of well separated levels.

The homogeneity of the samples must be checked to rule out the possibility of gross inhomogeneities leading to effective narrowing of the sample width and a corresponding increase in diagonal resistivity. This can be done by measuring the real and imaginary parts of the sample impedance where the real part of the impedance is inversely proportional to the conductivity, σ_{xx}, and the imaginary part is a measure of the capacitance which in turn is proportional to the area of the sample contributing to the conduction. These measurements were performed on some of the Si-MOSFETs used in this thesis [Kra91]. The results show that at half-filled Landau levels below the critical carrier density where the diagonal resistivity increases by several orders of magnitude, the diagonal conductivity develops minima, despite the fact that the maxima in the density of states should lead to R_{xx} maxima in the single particle picture. The imaginary part of the impedance shows that the capacitance does not change by more than 1% as the insulating phase

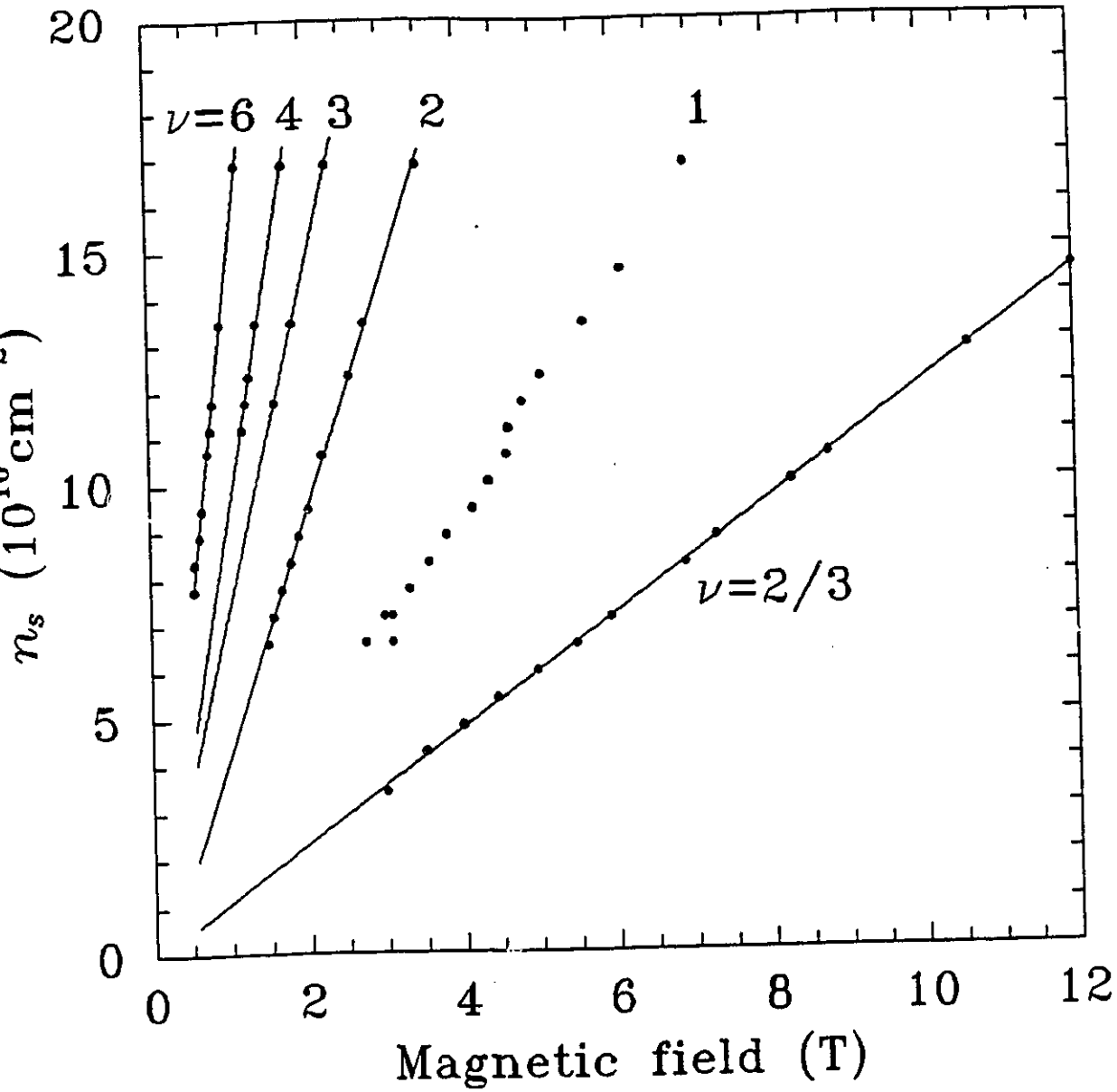

Figure 4.13: This figure shows the electron density versus the magnetic field at which the R_{xx} minima corresponding to various filling factors were centred. The data points were taken from sample Si-11 at a temperature of T = 24 mK. There is a significant discontinuity in the n_s versus B line for $\nu = 1$ which occurs at the same critical density as the reentrant insulator appears.

develops at half-filled Landau levels which demonstrates that the conductive area of the sample remains constant, ruling out the possibility of the development of large insulating islands narrowing the sample width. These measurements of the magnetic field dependence of the capacitance and the real part of the impedance also confirm that resistive contacts are not the source of the large resistance since the two measurements are very similar although the impedance is sensitive to contact effects while the capacitance is not. This is the same conclusion which was made in Chapter 3 by comparing the two and four contact transport data.

As there is no effective narrowing of the sample width, it is possible to use the data of the diagonal and Hall resistances to calculate conductivities above and below the critical carrier density. Contrary to Pruisken's scaling diagram shown in Figure 2.10, the appearance of the insulating phase at half-filled Landau levels marks the transformation of the maxima in σ_{xx} back to the origin $\sigma_{xx} = 0$ and $\sigma_{xy} = 0$ with $\sigma_{xy} \ll \sigma_{xx}$. The reentrant transitions occur between insulating states with $\sigma_{xx} \to 0$ and $\sigma_{xy}/\sigma_{xx} \to 0$ and integer quantum Hall states with $\sigma_{xx} \to 0$ and $\sigma_{xy} \to \nu e^2/h$. As the carrier density is decreased the scaling diagram collapses towards the origin as the minima in the quantum Hall effect are destroyed by increasing disorder and there is a crossover to a highly resistive phase over the entire range of magnetic field. Thus both diagonal and Hall conductivities vanish in the insulating phase and as seen in Figure 4.2, the Hall resistance does not change from its classical value as the diagonal resistance increases by several orders of magnitude. This is not typical of Mott localization or of magnetic freeze-out, where both ρ_{xy} and ρ_{xx} should approach infinity in the insulator phase as the temperature approaches zero.

Although the Hall resistance data is in agreement with the Viehweger and Efetov and Kivelson et. al. models of a metal-insulator transition with the Hall resistance keeping its classical value while the diagonal resistance becomes very large, the details of how this transition occurs experimentally are not consistent with these theoretical models. Figure 4.14 shows the theoretical phase diagram of the Hall insulator from zero field to $\nu = 1/2$. The x-axis is ρ_{xy} in units of h/e^2, which is equivalent to magnetic field, and the y-axis is ρ_{xx} in arbitrary units, representing the density of impurities in the sample, since ρ_{xx} increases as n_i increases. The regions inside the concentric semicircles, labelled $s_{xy} = n$, are regions in which there are n delocalized Landau levels below the Fermi level. Thus, the $\nu = 1$ plateau in ρ_{xy} is associated with the $s_{xy} = 1$ region, the $\nu = 2$ plateau with the $s_{xy} = 2$ region and so on. This model gives two very specific predictions. One is that ρ_{xy} in the insulating regions will retain its classical value and the other is that

any transition from $s_{xy} = n$ to the insulator must pass sequentially through $s_{xy} = n-1, n-2, n-3 \ldots$. This means that ρ_{xx} and possibly ρ_{xy} should exhibit some sort of anomalous behaviour every time a band of delocalized states rises in energy and passes through the Fermi energy. This prediction is obvious from the diagram because any path taken from $s_{xy} = n$ to the insulator (which is the area outside the $s_{xy} = 1$ semicircle) must pass through all the lower numbered states. In the experimental data from the Si-MOSFETs, there is no evidence for these multiple transitions as the diagonal resistivity increases smoothly from a quantum Hall effect minimum to a half-integer maximum and then back again. The Hall resistivity is even more convincing because it clearly shows no unusual characteristics whatsoever as the transitions from a quantum Hall state to an insulating state and back again are made.

Figure 4.15 is an experimental phase diagram taken from sample Si-5 at a temperature of 35 mK which shows the boundary between the metallic and insulating states as a function of both magnetic field and sample disorder. The disorder is represented by the ratio of the density of impurities to that of the electrons, n_i/n_s, and the magnetic field is represented by inverse filling factor, $\nu^{-1} = eB/n_{sc}h$. The boundary is denoted by n_i/n_{sc}, with the insulator lying above this boundary and the metal below it. The Hall insulator model assumes a single electron valley and spinless system and for the purpose of comparison, the two valley Si system is treated as a pair of stacked independent two dimensional systems. Thus, the region from $\nu = 0.5$ to $\nu = 2.5$ is attributed to $s_{xy} = 1$, the region from $\nu = 2.5$ to $\nu = 4.5$ to $s_{xy} = 2$, and the region from $\nu = 4.5$ to $\nu = 6.5$ to $s_{xy} = 3$. It is clear from the diagram that for certain values of n_s, there will be a transition from the insulating state directly to the $s_{xy} = 3$ state as the magnetic field is decreased. This was observed experimentally as a transition from the $\nu \approx 2.7$ maxima to the $\nu = 6$ IQHE minimum (corresponding to $s_{xy} = 3$). This deviates from the Hall insulator phase diagram because there is no evidence of the Fermi level passing through the $s_{xy} = 1$ and $s_{xy} = 2$ states before it reaches $s_{xy} = 3$. In fact, there is a smooth exponential rise of the diagonal resistivity and a gradual change of the Hall resistance (following the usual B/en_s dependence) as the magnetic field or the electron density are tuned. It is possible that the non-zero temperature at which the experiments are performed leads to a broadening of the anticipated features in ρ_{xx} since the theoretical limit of $T = 0$ and $\omega \to 0$ does not commute with the experimental conditions of $T \to 0$ and $\omega = 0$ [Zha92]. It is also clear that the Hall insulator model must be refined to treat disorder in a realistic way.

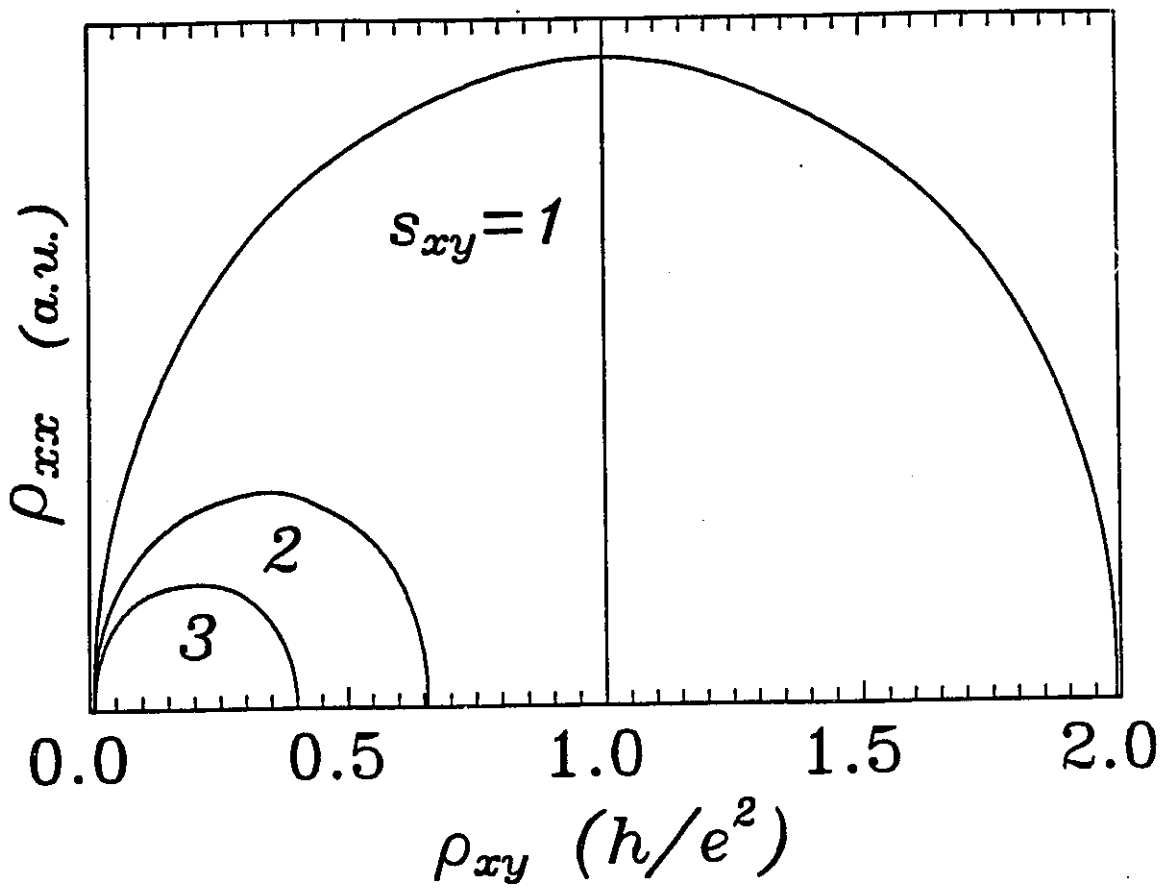

Figure 4.14: The theoretical phase diagram for the Hall Insulator is shown. The x-axis represents magnetic field (in terms of Hall resistivity) and the y-axis represents the disorder of the sample (in terms of longitudinal resistivity in arbitrary units). The concentric semicircles are regions with s_{xy} delocalized levels below the Fermi energy and the Hall insulator occurs in the area outside the $s_{xy} = 1$ semicircle. This graph is based on an original by D. H. Lee [Kiv92].

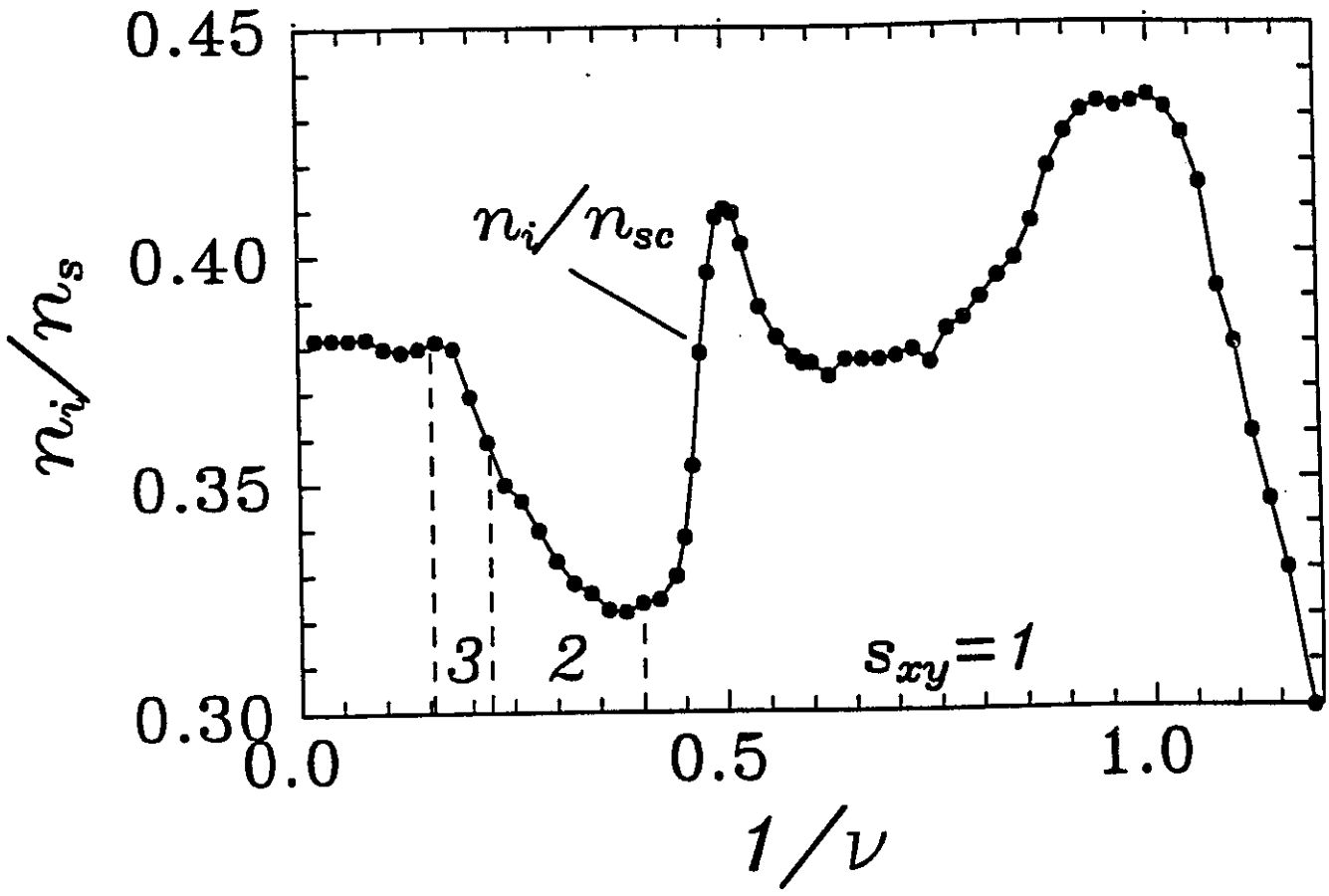

Figure 4.15: This figure shows the experimentally observed phase diagram of the insulator in terms of the density ratio of impurities to electrons (representing disorder) versus the inverse filling factor (which is proportional to magnetic field). The insulator occurs above the n_i/n_{sc} boundary. The data come from sample Si-5 at a temperature of 35 mK.

In considering single particle localization as the driving mechanism behind the metal-insulator transition at low magnetic fields, it is very important to determine the temperature dependence of the diagonal resistivity. As the $I-V$ characteristics are non-linear, both the regimes below and above the electric field threshold must be considered. The data shown in Figure 4.7 indicates that, below threshold, the temperature dependence of R_{xx} is best fitted by a $\exp(\Delta/T)$ model rather than by $\exp(T_0/T)^{1/3}$ or $\exp(T_1/T)^{1/2}$. The activation energy, Δ, reaches a maximum at half-filled Landau level and approaches zero in the quantum Hall effect regime at filled Landau levels, mimicking the magnetic field dependence of the resistivity. The activation energy also increases linearly below the critical carrier density down to $n_s \approx 7.5 \times 10^{10}\,\mathrm{cm}^{-2}$. Below this carrier density, the low temperature resistivity remains activated but Δ increases more slowly as n_s decreases. Above threshold, the resistivity decreases by several orders of magnitude, but remains much larger than that expected for a two-dimensional electron gas and the activated behavior persists with $\partial R/\partial T < 0$, albeit with a reduced activation energy. Since the typical thermal energies in this low temperature regime are more than an order of magnitude less than the Fermi energy, hopping conductivity should be the dominant conduction mechanism if the insulator is caused by single particle localization. The temperature dependence of the diagonal resistivity is, however, incompatible with this model.

The non-linear $I-V$ characteristics become broader as the temperature is increased in that the threshold differential resistance changes by orders of magnitude but the voltage drop across the sample remains unaffected. This is not expected in the single particle localization picture where dV/dT should be negative rather than zero. The magnitude of the threshold field must also be considered. As described previously, the Ioffe-Regel criterion for strong localization allows an estimate of E_t to be made. The minimum value of $E_t \approx 100\,\mathrm{V/m}$ which is calculated in this fashion is larger than the threshold fields extracted from the experimental data at any value of n_s, and is more than two orders of magnitude larger than the threshold fields extracted near n_{sc}.

The threshold field and activation energy exist over the same range of electron density, filling factor, and temperature, suggesting that the same process provides the carrier transport whether it is initiated by temperature or electric field. In fact, below the threshold electric field, the transport is activated with $\Delta \propto (E_t - E)$ and $(\partial\Delta(E)/\partial E) \sim \Delta_0/E_t$, where Δ_0 is the value of the activation energy as $E \to 0$. If single particle excitations are responsible for the conduction, then the derivative $\partial\Delta/e\partial E = \xi$ will be the characteristic length scale of this system. Calculating ξ reveals that it

is many times larger than the interelectron spacing, $a = (\pi n_s)^{-1/2}$, which means that the electrons are not strongly localized on individual impurity sites. For sample Si-11 at the $\nu \approx 2.7$ maximum for instance, data taken at an electron density of $n_s = 8.9 \times 10^{10}\,\mathrm{cm}^{-2}$ gives $a = 18.9\,\mathrm{nm}$ but $\xi = 24.9\,\mu\mathrm{m}$. Thus, ξ is a valid length scale for the conduction mechanism which is characterized by Δ_0 and E_t, but its magnitude suggests that the transport mechanism is not due to the excitation of single electrons from individual impurity potential wells. In conclusion, single particle localization does not provide either qualitatively or quantitatively a framework which accounts for the experimental data.

4.6.3 The Pinned Wigner Solid and the Data

Since single particle localization cannot readily account for the experimental data, a collective excitation or a many-body interaction approach should be considered. The energetics in this material system certainly favour many-body excitations since the Coulomb interaction energy is the dominating energy scale, with $E_{ee} = (\pi n_s)^{1/2} e^2/\kappa \sim 100\,\mathrm{K}$, being larger than either the cyclotron energy ($\sim 12\,\mathrm{K}$ at $B = 2\,\mathrm{T}$) or the zero-point oscillation energy ($\sim 6\,\mathrm{K}$). The experimental observation of enormous resistance growth with decreasing temperature and non-linear $I - V$ curves are both typical characteristics of a solid rather than a liquid or gaseous state. Two different descriptions of an electron solid are those of the Wigner solid (WS) and the charge density wave (CDW). Here it is assumed that the low magnetic fields used in these experiments, where the cyclotron energy is always less than the Coulomb interaction energy, act simply as a perturbation of the zero magnetic field "quantum regime" Wigner crystal, rather than driving the system into the "extreme quantum regime" (as discussed in Section 2.5). Since the reentrant metal-insulator transitions in Si-MOSFETs occur in higher Landau levels where $\nu > 1$ and the magnetic length is larger than the mean interelectron spacing, the charge density wave model may be a more appropriate description than that of the traditional Wigner crystal. In a magnetic field, however, the charge density wave is equivalent to a Wigner solid of quasi-electrons or quasi-holes. Impurities in the vicinity of the two dimensional electron layer also play a role by acting as pinning centers for the electron solid or the CDW. It will thus be most profitable to discuss the results in the framework of a pinned electron solid but to remember that the distinction between a pinned Wigner solid and a pinned charge density wave is extremely blurred in the regime where these experiments take place.

The pinned Wigner solid is expected to show insulating behavior because

of the Coulomb correlation potential which holds individual electrons in their respective lattice positions and the impurity pinning potential which prevents the crystal from moving coherently. The current-voltage characteristics will also show some non-linearity, since it should be possible to apply an electric field large enough to overcome the pinning impurity potential and force the solid lattice to slide coherently. This allows all the electrons in the Wigner crystal to participate in conduction and produces a significant reduction in the differential resistance. Since the ratio of the Coulomb correlation energy to the quantum zero point motion energy of electrons gets larger with decreasing electron density while the thermal kinetic energy of individual electrons gets larger with increasing temperature, the Wigner crystal should exhibit a critical melting density and a melting temperature, as discussed in Section 2.5. The existence of a critical density is observed experimentally in plots of R_{xx}, Δ, or E_t versus n_s (See Figures 4.4, 4.8 and 4.11) and a melting temperature is also observed in that all three of these signatures disappear at temperatures above $T \sim 500\,\mathrm{mK}$. As the carrier density is decreased and the effective disorder increases, the activation energy and threshold field values grow, underlining the importance of disorder and pinning.

The CDW model predicts that the magnetic field should have a stabilizing influence on the formation of the CDW which, however, should be suppressed at filled Landau levels. The experimental results show that the effect of the magnetic field is to favour the insulating state at half filled Landau levels and suppress its formation at integer filling factors until the carrier density is quite low ($n_s \approx 7 \times 10^{10}\,\mathrm{cm}^{-2}$). At filling factor $\nu = 1$ (corresponding to a magnetic field of approximately 4 Tesla), the valley splitting is approximately 4.8 K which is much larger than the level broadening and the insulating state does not set in until the carrier density is low and the relative disorder is high. At filling factor ν=2, the splitting at 1.7 Tesla is only 1.4 K and the insulating state appears at higher carrier density than at $\nu = 1$. An elegant formulation of this magnetic field dependence has recently been given [Chu94] which accounts for both the apparent competition between the Wigner solid at half filled Landau levels and the quantum Hall effect at filled Landau levels and the shift towards higher magnetic fields of the Landau level fan diagram below the critical carrier density. The chemical potential in the liquid or gaseous phase exhibits quantum oscillations as a function of magnetic field while the chemical potential of the solid is a monotonically decreasing function of magnetic field. In order to maintain the equality of the chemical potential in both phases, the boundary between these two phases must oscillate. Thus the reentrant transitions are not caused specifically by the quantum Hall effect or by changes in the energy gap but by whether the liquid or the solid

state has the lowest chemical potential at a given field.

Since the solid is pinned by impurities, it is also expected that there will be a mobility dependence, and that samples with fewer impurities will be more weakly pinned and thus show lower values of Δ and E_t at any given electron density. The presence of impurities close to the electron layer, however, can destroy the long or medium range order which is required for a correlated electron crystal. The sample is therefore polycrystalline with the size of the crystallites depending on the disorder or the number of pinning impurities. There is perfect coherence within each crystallite but phase dislocations occur at the borders between crystallites. As the carrier density decreases and the ratio of impurities to electrons approachs unity, the crystallites become smaller in size and eventually a glassy state ensues in which there is no long range order. Increasing the number of impurities still further (until $n_s/n_i \sim 1$) will totally destroy the Wigner crystal, because each electron will be pinned at a different impurity site. This is the single particle localization limit. Thus, the observed mobility dependence in which the insulator in the lower mobility Si-11 is pinned more strongly than in Si-5, and in which the even lower mobility Si-2 fails to show the characteristics of the other two samples, agrees with the predicted mobility dependence of a Wigner crystal.

The experimental $I - V$ curves are very similar to those previously observed for charge density waves in linear chain compounds, and depinning of a charge density wave by an electric field was described in detail in Section 2.4. This threshold is certainly not due to Joule heating (for reasons detailed in Section 3.2.4) and it is too small to be caused by depinning of localized electrons (as demonstrated by the calculation of E_t from the Ioffe-Regel criterion). The fact that the conduction mechanism is still activated after the threshold field has been exceeded and that the resistance is still at least two orders of magnitude too high to describe a non-interacting electron gas suggests that the conduction mechanism above threshold is due to a coherent sliding crystal.

As seen in the discussion of single particle localization the temperature dependence of the diagonal resistivity is not consistent with hopping mechanisms at low temperature but is typical of a thermally activated process. In the framework of the pinned WS, the current is carried by lattice defects such as vacancies, interstitials, or dislocation pairs, which does result in temperature activated conduction. S. T. Chui and his coworkers [Chu91a, Chu91b, Chu93a, Chu93b, Chu93c] have considered a pinned Wigner crystal with charge transport through the motion of extended defects. This model can be expressed in terms of a pinned charge density wave by assuming that the extended defects are the phase dislocations which occur

at the ends of coherent segments of a CDW or at the boundaries of individual crystallites in a Wigner crystal. Thus, it is already compatible with all the basic trends in the experimental data which were discussed previously. If charge transport below the threshold is by the motion of extended defects (dislocation pairs in this case) then ρ_{xy} should retain its classical value (as described in Section 2.5.3). This also accounts for the large values of ξ since dislocation motion involves many electrons being moved slightly from their equilibrium lattice positions and thus is a collective process. The $I-V$ characteristics are non-linear because beyond the threshold, the electric field is strong enough to break the dislocation pairs apart. This results in a constant stream of dislocation pairs being generated and then broken, creating a sudden decrease in the differential resistance. This is consistent with the observation that the nature of the transport mechanism is the same both below and above the threshold field (in that it occurs through the creation of dislocation pairs). The thermal activation energy in this model should depend linearly on $\delta n_s = (n_{sc} - n_s)$. This is in excellent agreement with the linear dependence of Δ on n_s in Figure 4.10. Finally, a quantum Monte Carlo simulation [Chu93c] of electrons in a real two dimensional system with impurities has been performed for this model. Using parameters appropriate for silicon (ie. two valleys, impurity scattering, etc.) a value of $r_{sc} \approx 10$ was obtained. This corresponds to a critical electron density of approximately $n_{sc} = 6.9 \times 10^{10}\,\mathrm{cm}^{-2}$ for a zero field Wigner crystal, which is fairly close (especially considering that a magnetic field should enhance the formation of the crystal) to the critical densities of $n_{sc} \approx 9 - 10 \times 10^{10}\,\mathrm{cm}^{-2}$ that were determined for Si-5 and Si-11 under a magnetic field. It seems that this model can successfully account for all of the experimental data.

In conclusion, the data is not well described by the standard localization models and the large length scale and small electric field suggest that it is more likely the result of some collective phenomenon than of single particle localization. It is true that if n_s is decreased sufficiently, then E_t will continue to increase and eventually ξ may approach the SPL limit (just as n_s will eventually approach n_i), but this is clearly not the case near n_{sc}. The fact that the metal-insulator transition is multiply reentrant as the magnetic field changes is unusual, but it is consistent with a Wigner crystal which is in equilibrium with a quantum fluid. A pinned Wigner crystal also explains the activated transport, non-linear $I-V$ curves, melting temperature, cold melting density and the mobility dependence of n_{sc}. A theoretical calculation of n_{CM} for a pinned Wigner crystal which uses parameters appropriate for these Si-MOSFETs is in reasonable agreement with the experimentally determined values of n_{sc}. Thus the best explanation for the anomalous R_{xx} and

all the other characteristics that were subsequently found is that a pinned Wigner crystal is formed near the half integer filling factors in these very high mobility Si-MOSFETs when the temperature and electron density are sufficiently low.

Chapter 5. Experiments At Zero Magnetic Field

5.1 Introduction

The previous chapter described a variety of experiments carried out in the presence of a perpendicular magnetic field. Anomalously high values of R_{xx} were observed with maxima near filling factors $\nu \approx 1.5$ and $\nu \approx 2.7$ when the electron density was below a critical value, n_{sc}, separated by integer quantum Hall effect minima at $\nu = 1, 2$ and 6. This insulating phase showed properties such as thermally activated transport and a threshold electric field, which suggested that it represented a Wigner crystal. If the electron density was decreased sufficiently the insulating phase began to spread over a wider range of magnetic field until eventually the integer quantum Hall effect minima and the metallic zero field state all became insulating. A detailed study of the zero magnetic field insulator was undertaken because it held the possibility of being the first manifestation in the quantum regime of a zero field Wigner crystal. This chapter focuses on the experiments carried out in the absence of a perpendicular magnetic field. The measurements described are of three types: measurements of the sample resistance (R_{xx} only since $B = 0$), measurements of voltage between a pair of contacts as the current through the sample is varied and measurements of the voltage noise between a pair of contacts as the current is varied. All of these measurements can be carried out at different temperatures and electron densities with the results used to calculate some typical parameters such as thermal activation energy and critical density. The addition of noise data and some calculations based on specific models are used to enhance the discussion of whether or not the zero magnetic field insulator which is observed at sufficiently low electron densities represents a quantum Wigner crystal. At zero magnetic field, all three samples tested showed similar characteristics, and a phase diagram is developed which shows the range of electron densities over which the Wigner crystal is expected to exist at zero magnetic field as a function of the sample mobility. Once again, all of the measurements except the temperature dependences were performed near base temperature, and unless otherwise noted the uncertainty in each measurement is within the size of the points on the graph.

5.2 Temperature and Density Dependence of R_{xx}

Figure 5.1 shows ρ_{xx} as a function of carrier density, n_s, taken from sample Si-2 at $T = 33\,\mathrm{mK}$. The current passed through the sample from source to drain had to be changed several times during this measurement because of the changing resistivity, but in each case the current was high enough that the resistivity could be measured accurately and low enough that the dissipated power was not sufficient to heat the sample. As n_s decreases below a critical density, n_{sc}, which lies between $10.0 \times 10^{10}\,\mathrm{cm}^{-2}$ and $10.5 \times 10^{10}\,\mathrm{cm}^{-2}$, ρ_{xx} increases exponentially over four orders of magnitude. For densities below approximately $8.5 \times 10^{10}\,\mathrm{cm}^{-2}$, the increase of ρ_{xx} with decreasing carrier density levels off. In the region above n_{sc} the sample has resistivities of a few kΩ (appropriate for a two dimensional electron gas at this temperature) but in the region below $n_s = 8.5 \times 10^{10}\,\mathrm{cm}^{-2}$ the resitivity values are over 100 MΩ, and the sample is strongly insulating.

Figure 5.2 shows Arrhenius plots of resistivity versus inverse temperature from sample Si-2 at three different electron densities (indicated in units of $10^{10}\,\mathrm{cm}^{-2}$ on the graph). These curves divide the temperature axis into two regimes. At higher temperatures ($500\,\mathrm{mK} > T > 120\,\mathrm{mK}$), each data set is characterized by a linear fit with a relatively high slope (and correspondingly high activation energy, Δ_1), whereas at the lowest temperatures a linear fit with much lower slope and activation energy, Δ_2, is more appropriate. The linear fits are least squares best fits of the form $\rho_{xx} \propto \exp(\Delta/T)$, identical to that used to calculate the activation energies in Chapter 4. The source-drain currents are low enough that the measurements are all below the threshold in the $I - V$ curves discussed in the next section. Measurements above the threshold result in much lower activation energies. Other common exponents (such as $\rho_{xx} \propto \exp(T_0/T)^{1/3}$ and $\rho_{xx} \propto \exp(T_1/T)^{1/2}$) associated with different transport models were also tried but the linear least squares best fit with two different temperature regimes most closely matched the data. Various combinations of two exponent models were also tried, to determine if a single curve could be made to fit all the data points, but none was found which could satisfactorily fit the entire curve. This suggests that the conduction mechanisms in the two different temperature regimes do not exist in parallel, but that each only exists over a limited temperature range.

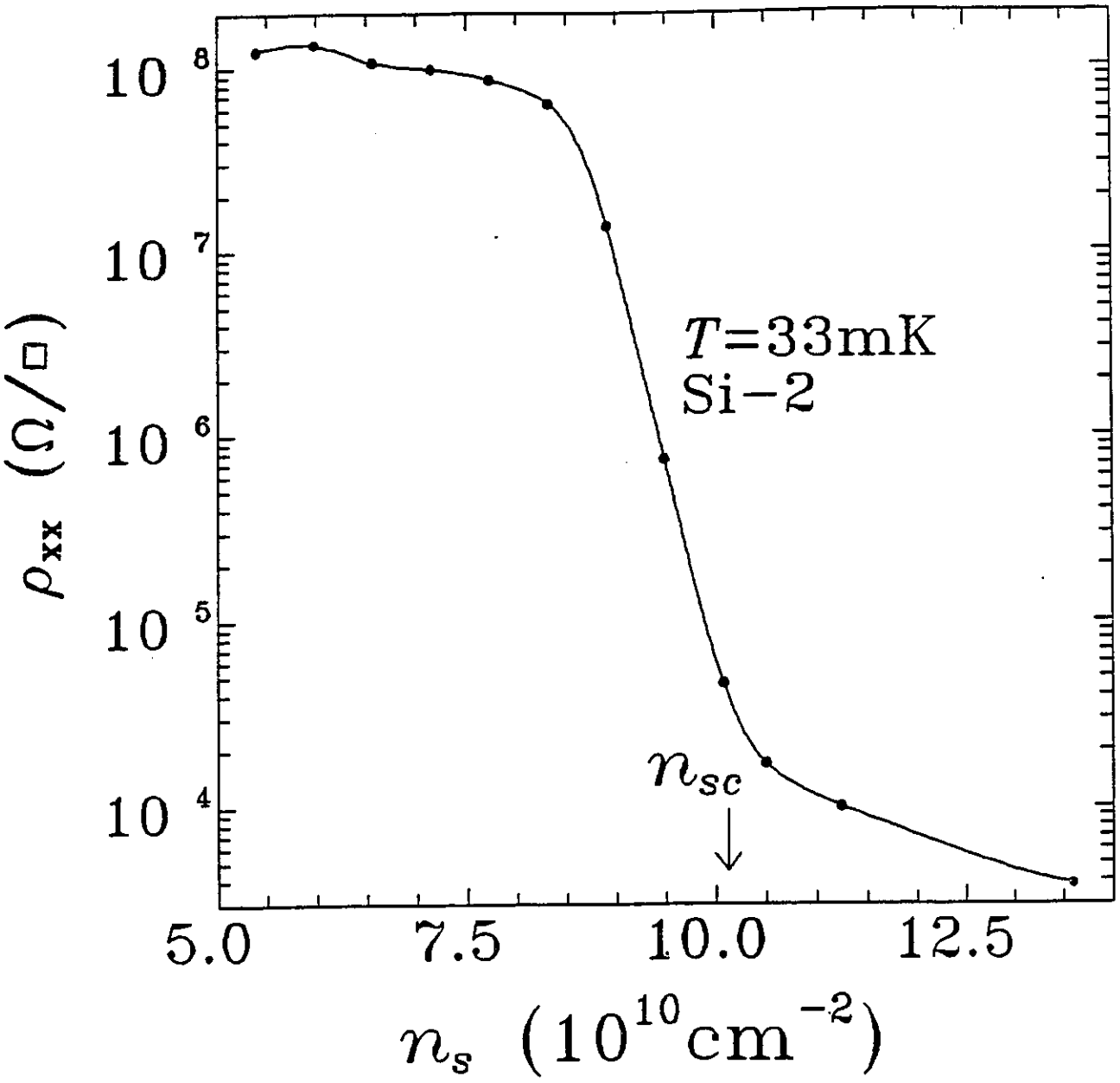

Figure 5.1: A graph of resistivity versus electron density taken from sample Si-2 is shown. A critical density, n_{sc}, below which the resistivity starts to increase exponentially is visible between $n_s = 10.0 \times 10^{10}\,\mathrm{cm}^{-2}$ and $n_s = 10.5 \times 10^{10}\,\mathrm{cm}^{-2}$.

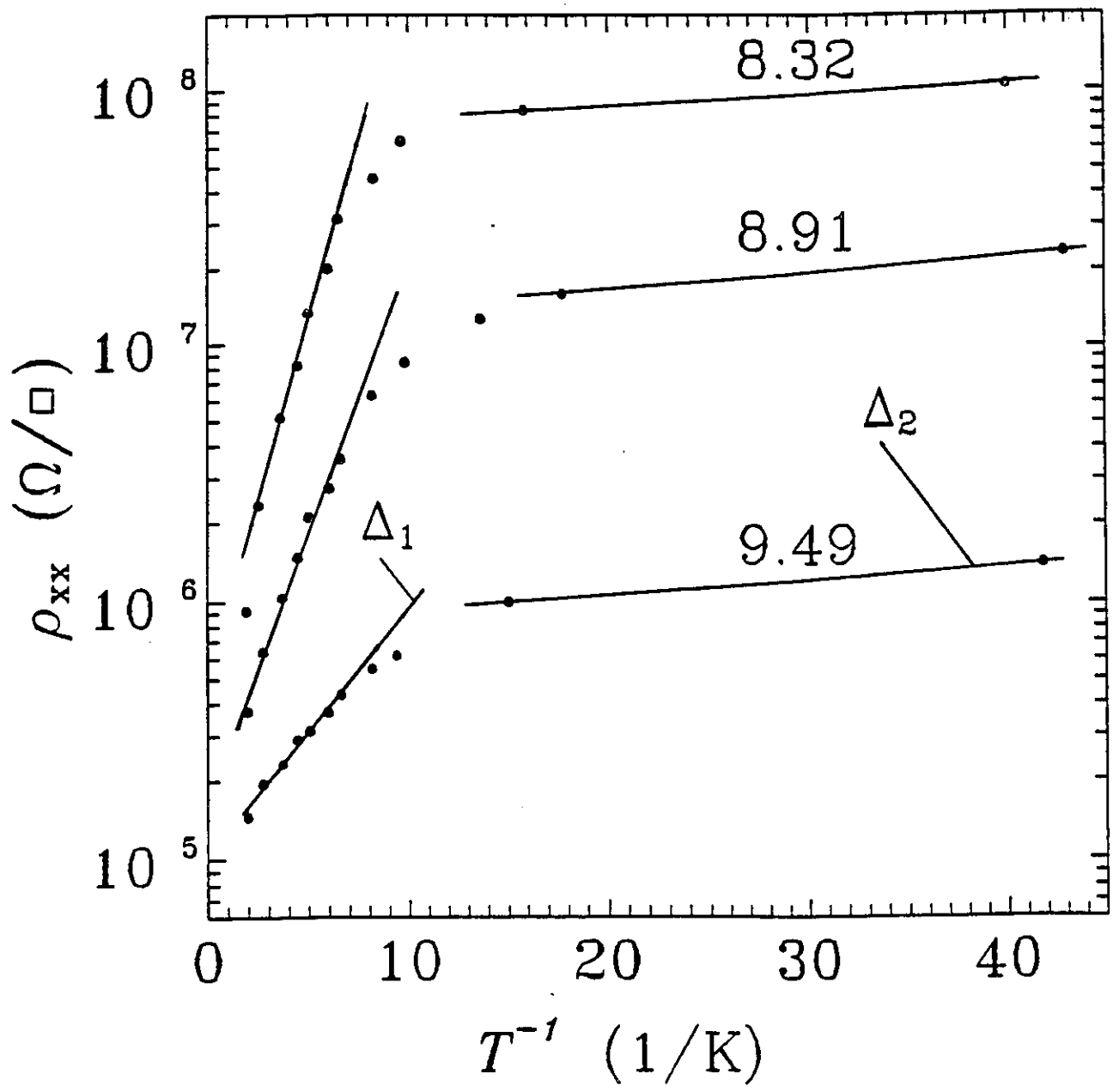

Figure 5.2: Three Arrhenius plots of resistivity versus inverse temperature taken from sample Si-2 at the electron densities indicated next to each curve in units of $10^{10}\,\mathrm{cm}^{-2}$ are shown. The linear fits to the curves indicate the existence of two different temperature regimes, with the activation energies Δ_1 and Δ_2 respectively.

5.3 $I-V$ Characteristics

When the carrier density is below n_{sc}, an important transport signature of the CDW or pinned Wigner crystal can be observed. This is the non-linear $I-V$ characteristic and associated threshold electric field which was discussed in Chapter 2 and observed in Chapter 4 near the half integer filling factors. The $I-V$ curves are measured by setting the gate voltage (and thus the electron density) to a specific value and measuring the voltage drop, V_{xx}, across a pair of longitudinal contacts as a function of the current applied between source and drain. The voltage is measured as the current is swept from negative to positive values (usually with the current at zero voltage subtracted from each point) so that any asymmetry effects or dc current shifts can easily be detected.

Figure 5.3 shows three typical $I-V$ curves measured from sample Si-5 at the electron densities indicated on the graph. These curves have the same form as those observed in Chapter 4. There are two linear regions separated by a distinct threshold voltage, V_t. The differential resistance (slope of the curve) is much higher below V_t than above it, and both V_t and the differential resistance below V_t increase substantially with decreasing carrier density. This figure also shows that above some critical density (which lies between $n_s = 7.79$ and $8.23 \times 10^{10}\,\mathrm{cm}^{-2}$ in this case) the non-linearity in the $I-V$ curves disappears and the threshold voltage becomes zero. Changes in temperature, although not specifically shown here, do not affect V_t until the melting temperature is reached, just as was observed in Chapter 4. The precise value of V_t is calculated for each $I-V$ curve by extrapolating the linear region above the threshold back to $I=0$, or, if sufficient points were measured below the threshold, by plotting differential resistance versus voltage.

If a sufficient number of $I-V$ curves are measured for the same sample at different carrier densities, a graph of V_t (or equivalently E_t, the threshold electric field) versus n_s can be plotted. Figure 5.4 shows such a plot using data from sample Si-2. The open circles are the threshold electric field values and the filled circles are values of the activation energy (actually Δ_1) measured at the same electron densities. Both the linear fit to the activation energy data and the cubic polynomial fit to the threshold electric field data lead to the same value of the critical density, $n_{sc} = 10.05 \times 10^{10}\,\mathrm{cm}^{-2}$. Thus, both the non-linear $I-V$ characteristics and the thermally activated transport disappear above the same critical carrier density and the fact that both of these quantities decrease smoothly to zero as n_s passes above n_{sc} suggests that this phase transition is second order. More recent measurements on the

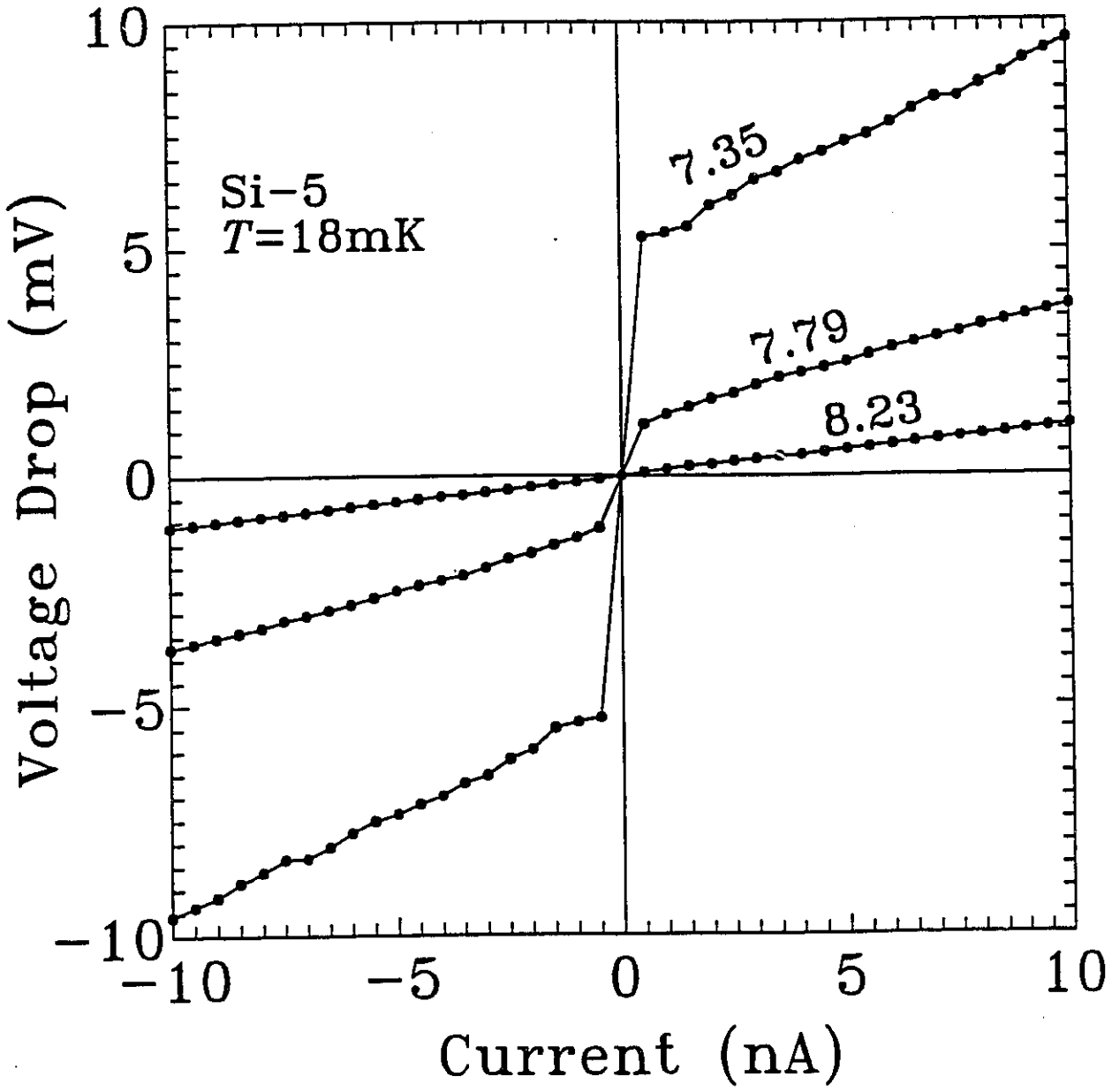

Figure 5.3: The voltage drop across contacts P1 and P3 on sample Si-5 as a function of the current passed from the source to the drain at three different electron densities (indicated by the numbers next to each curve in units of $10^{10}\,\mathrm{cm}^{-2}$).

same samples [Pud94c], however, have used very sensitive techniques to focus on the value of E_t near n_{sc} and have detected a small step from a finite value to zero at the phase transition. This indicates that the transition is weak first order rather than second order.

Repeating this measurement for the other two samples reveals that the higher the mobility of the sample, the lower the critical density at which the unusual transport characteristics appear. Sample Si-5, with a peak mobility of 41000 cm^2/Vs, has a critical density of approximately $7.8 \times 10^{10}\,cm^{-2}$ whereas sample Si-2, with a peak mobility of 23000 cm^2/Vs, has a critical density of approximately $10.0 \times 10^{10}\,cm^{-2}$. Sample Si-11, with an intermediate mobility of 33000 cm^2/Vs, also has an intermediate critical density of $8.5 \times 10^{10}\,cm^{-2}$.

5.4 Noise Measurements

If the electric field threshold represents the point at which a pinned Wigner crystal becomes depinned and starts to slide coherently, then the simple CDW analogy discussed in Chapter 2 implies that above the threshold there should exist a specific voltage noise spectrum associated with the periodic structure of the electric charge in the crystal moving past stationary ionic impurities. For a one dimensional, perfectly coherent CDW, the frequency spectrum is calculated in Equation 2.42 to have a root frequency, $f_0 = j_{CDW}/en_s\lambda$, where λ is the wavelength of the CDW. For a Wigner crystal in a two dimensional Hall bar, j_{CDW} should be replaced by I_x/y, where I_x is the source-drain current and y is the width of the Hall bar, and λ should be replaced by d_n, the length of one of the Wigner crystal lattice vectors. This defines a series of frequency peaks $f_n = I_x/en_s y d_n$. In large samples such as these Si-MOSFETs, however, it is more realistic to expect multiple separately pinned crystallites, each of which will generate its own noise spectrum. Since there will be phase differences between the noise from each crystallite and since not all the crystallites will necessarily move in the direction of macroscopic current flow at any given time, the frequency spectrum is expected to have broad peaks which may not be centred at the precise frequencies f_n.

The lower half of Figure 5.5 shows an $I - V$ curve taken from sample Si-2 using contacts P1 and P3 at an electron density of $8.32 \times 10^{10}\,cm^{-2}$. The upper half of the figure shows a measurement of the root mean square voltage noise in the frequency band from 1 kHz to 300 kHz under the same conditions and over the same current range. There is a marked increase in the voltage noise level as the threshold field is exceeded. This increase

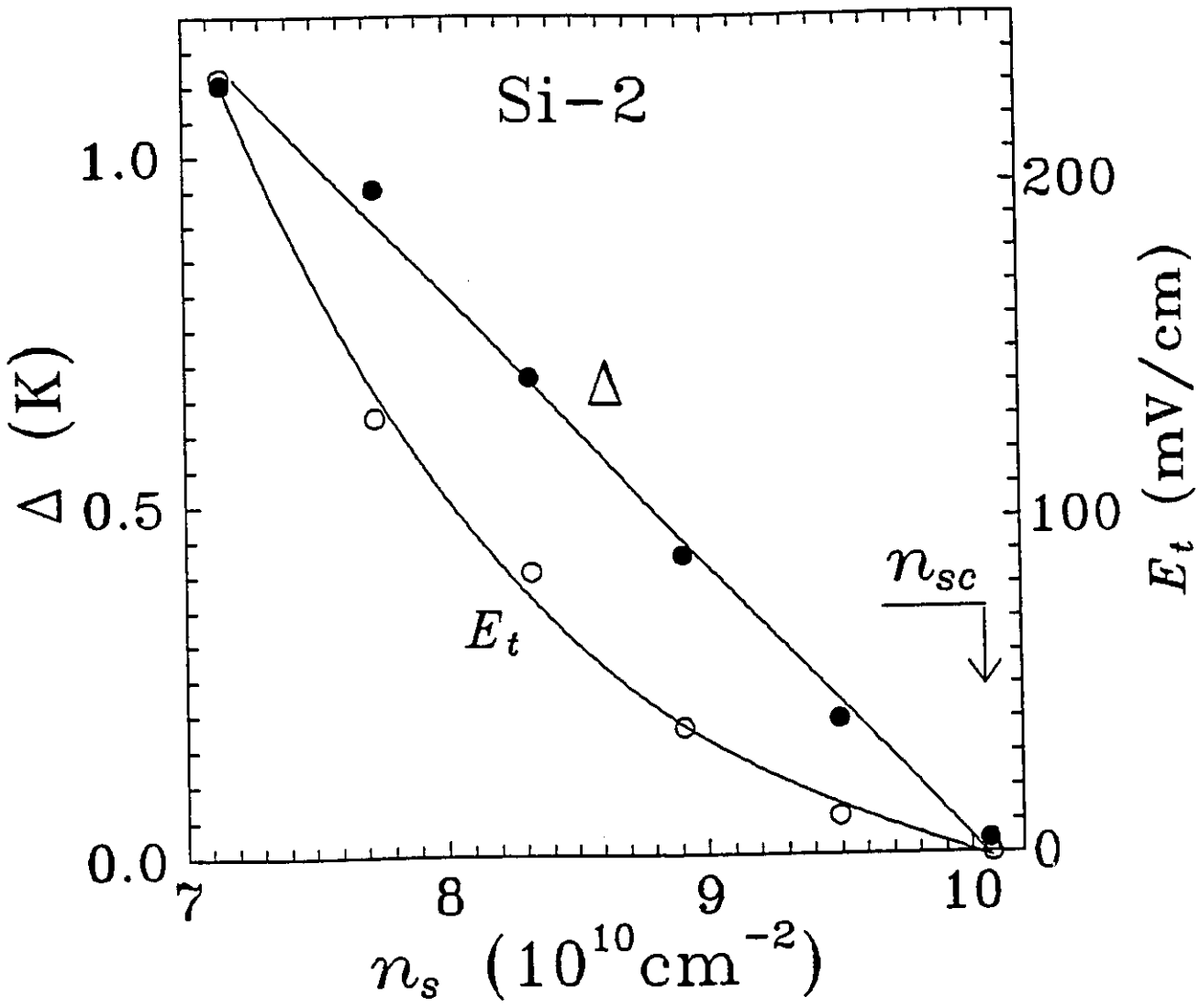

Figure 5.4: This figure shows both the activation energy, Δ, (taken from the temperature regime, $T \geq 120\,\mathrm{mK}$) and threshold field, E_t, measured from sample Si-2 at different values of the electron density. Both the linear fit to Δ and the third degree polynomial fit to E_t lead to the same value for the critical density, $n_{sc} = 10.05 \times 10^{10}\,\mathrm{cm}^{-2}$.

in the broad band noise level just beyond the threshold is observed in all three samples. It indicates an increase in the voltage noise spectrum as the threshold is exceeded, but since it integrates the noise over a wide frequency range it is insensitive to the specific noise frequency spectrum and gives no information about the coherence of the noise generated in different parts of the sample. The asymmetry between the noise generated at negative and at positive currents is interesting. It can not be attributed to any measurement uncertainty in the experiment, since the instruments used for this measurement were very accurate and the error bar on each individual point would be too small to see on the scale of this graph. There were some external noise sources, however, such as pickup noise from the probe, which varied over time and these could be responsible for the small scale differences between the positive and negative current segments of the graph. The fact that the left noise peak is wider than the right peak suggests a hysteresis effect, since this phenomenon depended on the direction in which the current was swept.

Figure 5.6 shows two separate voltage noise frequency spectra obtained from sample Si-2 at an electron density of $8.32 \times 10^{10}\,\mathrm{cm}^{-2}$ using a current of 50 pA (which is above the threshold for this sample under these conditions). A current of 50 pA was selected so that the theoretically predicted noise peaks would fall in the measurable frequency range. This is limited on the low side by the $1/f$ noise which makes measurements below 4 kHz impossible, and on the high side by the RC time constant associated with the MOSFET gate capacitance which cuts off measurements beyond approximately 12 kHz. The similarity of these two independently measured spectra, demonstrates that the general shape of the spectrum is reproducible. The lines drawn through the points are a guide to the eye, to suggest what a continuous spectrum might look like. The error bars (shown only for one set of points to avoid confusion on the graph) indicate that there is flexibility in the precise identification of the peak frequencies (indicated by the arrows on the graph), but two of the peaks are clearly visible in approximately the same positions in both spectra.

These frequency spectrum measurements are difficult to make because the amplitude of the signal is in the $\mathrm{nV/Hz^{1/2}}$ range, and there are other noise sources in the lab which are in the $\mu\mathrm{V/Hz^{1/2}}$ range. Reducing the error bars to the level shown requires each point to be measured separately using a tunable amplifier with the width of the band pass about 1% of the frequency being measured. Comparing the output signal from the tunable amplifier when the 50 pA current is turned on with the output at zero current, allows that part of the noise signal in a narrow frequency band which was specifically attributable to the current to be isolated. Both the $I = 50\,\mathrm{pA}$ signal and

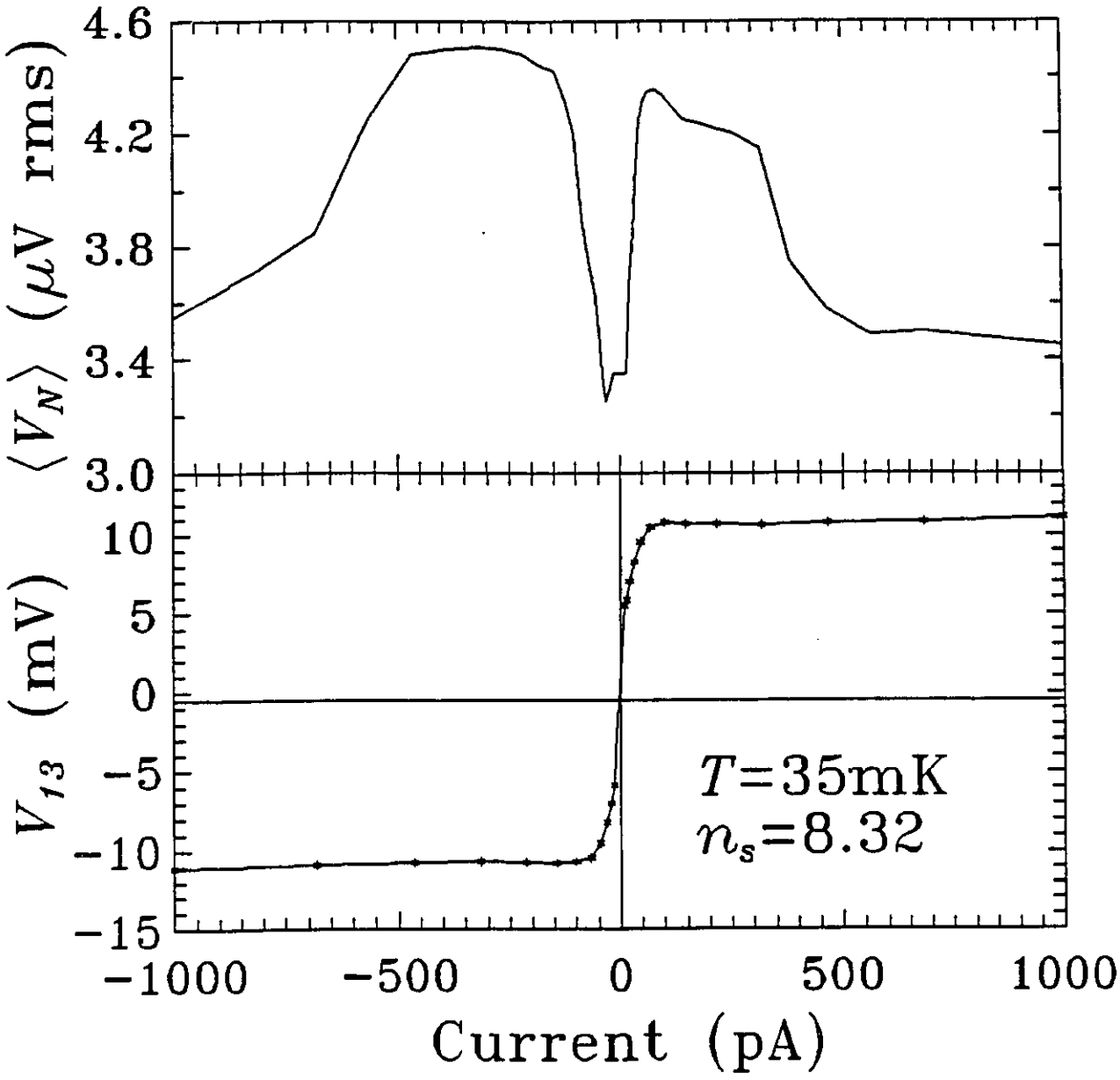

Figure 5.5: The upper section of this figure shows the root mean square broad band noise in the frequency range from $1 - 300$ kHz versus the source-drain current across sample Si-2 at an electron density of $8.32 \times 10^{10}\,\mathrm{cm}^{-2}$. The lower part of the figure shows the corresponding potential drop across contacts P1 and P3. Note the sharp increase in the noise amplitude as the threshold is exceeded.

the $I = 0$ signal are integrated for five to ten minutes before subtraction so that every ten to twenty minutes a new data point is obtained for the current frequency value. When enough data points have been obtained that the standard deviation is low enough to indicate whether that frequency value lies at a maximum, a minimum, or somewhere in between, the tunable amplifier is reconfigured and the noise in a 1% band about a new frequency value is measured.

Figure 5.7 shows two measurements of the voltage noise frequency spectrum taken at different values of the source-drain current. These also used contacts P1 and P3 on sample Si-2 at $T = 30\,\mathrm{mK}$ and $n_s = 8.32 \times 10^{10}\,\mathrm{cm}^{-2}$. The empty circles represent points from a 50 pA spectrum and the filled circles represent points from a spectrum with $I = 75\,\mathrm{pA}$. The three original peaks in the 50 pA spectrum are indicated by the downwards pointing arrows. The middle peak is not readily visible from the positions of the points, but it can be inferred from a knowledge of the peak positions in the previous two spectra shown in Figure 5.6. The absolute frequency positions of these three peaks have shifted slightly upwards, but this is not surprising because several days elapsed between the measurement of the two spectra shown in Figure 5.6 and of the two spectra shown in Figure 5.7. Since the threshold gate voltage of each MOSFET drifted slowly with time after its initial cooldown, the actual electron density may have varied between the two data sets, although the nominal electron density was $8.32 \times 10^{10}\,\mathrm{cm}^{-2}$ in both cases. The most important feature is that in the 75 pA spectrum the divergent $1/f$ noise signal followed by the first two peaks (indicated by the upwards pointing arrows) are shifted to higher frequencies than in the 50 pA spectrum. Since these two spectra were taken on subsequent days, this change can not be explained by a drifting gate voltage. It indicates a frequency spectrum which is current dependent, as predicted for a sliding CDW or Wigner crystal, although the exact linear dependence predicted by Equation 2.42 is not observed. These measurements are still at an early stage, however, and more detailed spectral analysis is required before a definitive statement can be made about the agreement between the prediction of Equation 2.42 and the experimental data.

The experimental results for the zero magnetic field case can be simply summarized. Below some critical electron density, n_{sc}, (which is lower for higher mobility samples) a strongly insulating phase develops which exhibits thermally activated conduction and non-linear $I - V$ characteristics. This phase disappears above some critical temperature, which varies with sample and carrier density, but is typically in the range $400 - 600\,\mathrm{mK}$. The threshold voltage, above which the differential resistance decreases sharply, increases

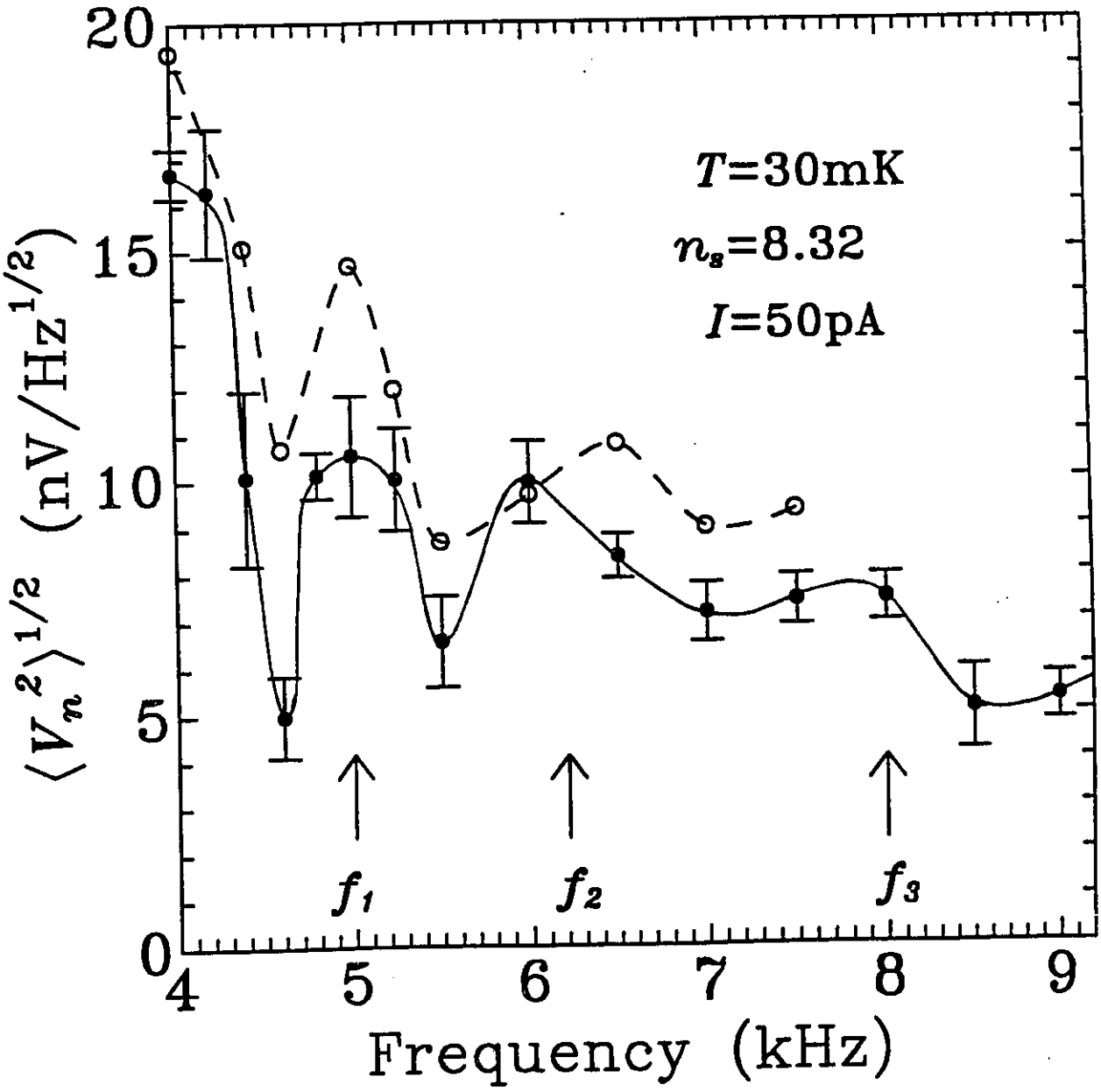

Figure 5.6: This figure shows the amplitude of the noise received in a narrow band of about 1% of the centre frequency for various points between 4 kHz and 9 kHz. The noise potential was measured between contacts P1 and P3 of sample Si-2 at an electron density of 8.32×10^{10} cm^{-2} with a source-drain current of 50 pA. The filled circles and empty circles represent two independent measurements of the same spectrum. The arrows indicate the approximate locations of three peaks in the noise spectra. The lines connecting the points are a guide for the eye.

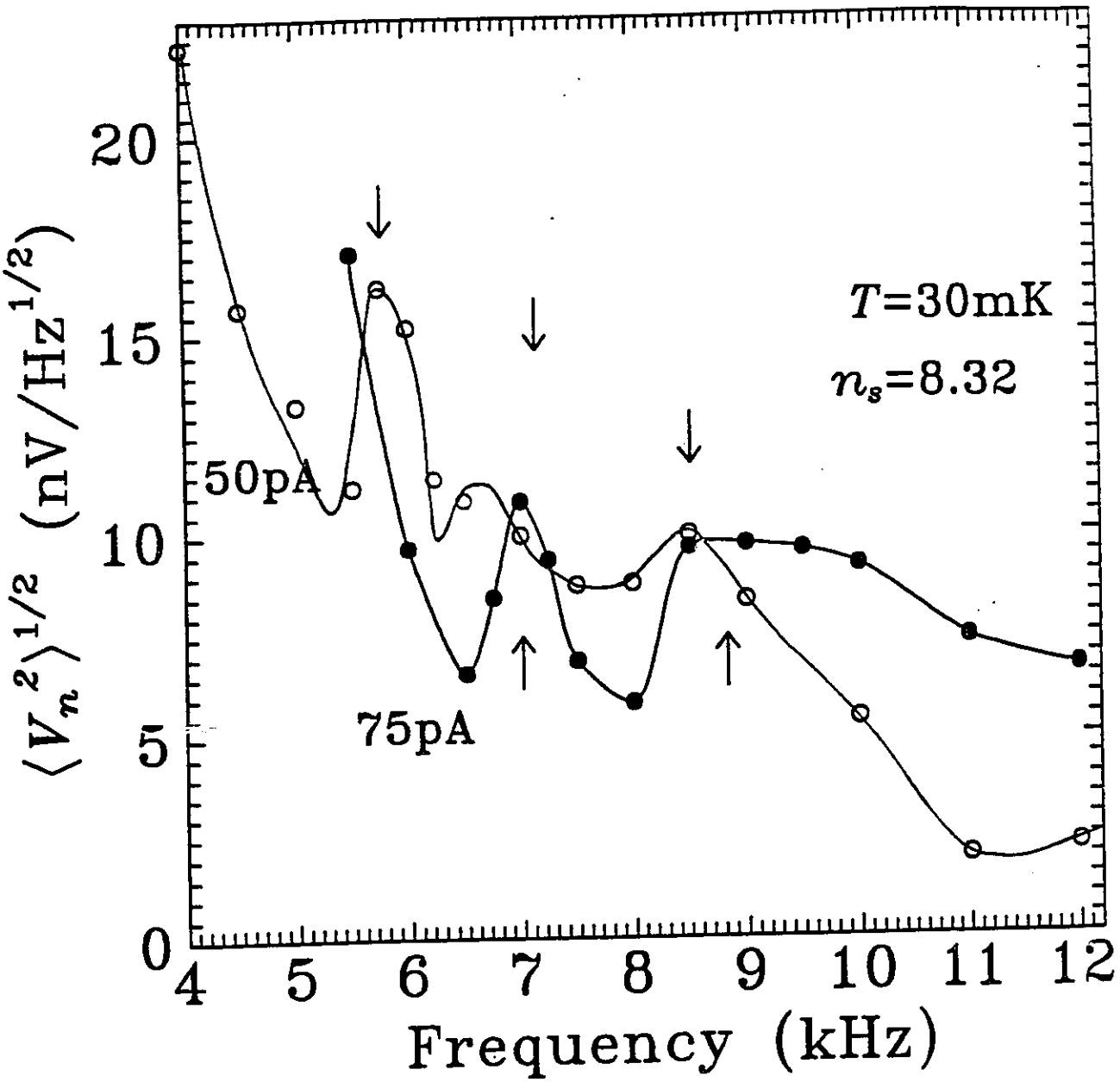

Figure 5.7: This figure shows two measurements of the amplitude of the noise potential between contacts P1 and P3 of sample Si-2 at an electron density of $8.32 \times 10^{10}\,\mathrm{cm}^{-2}$. The empty circles represent the frequency spectrum measured with $I = 50\,\mathrm{pA}$ and the filled circles represent the $I = 75\,\mathrm{pA}$ frequency spectrum.

with decreasing temperature and there is a rise in the broad band voltage noise measured between a pair of longitudinal contacts as that threshold is exceeded. The frequency spectrum of the voltage noise in the region above threshold shows reproducible peaks at specific frequencies in the $4 - 12\,\mathrm{kHz}$ range, and these peaks shift to higher frequencies when the current is increased.

5.5 Discussion and Analysis

The results observed at zero field bear a remarkable resemblance to those observed under a magnetic field. Below a certain critical density an insulating phase with very large R_{xx} develops. The conduction in this insulator is thermally activated with two different activation energies in different temperature regimes and shows non-linear $I - V$ curves with a threshold electric field. Although the values of n_{sc} below which the insulator develops is lower at zero magnetic field than it was at the half integer filling factors, the typical magnitudes of Δ, E_t and R_{xx} below n_{sc} are all very similar to those observed at the half integer filling factors. Even the critical melting temperature of $\sim 500\,\mathrm{mK}$ is the same as the melting temperature for the insulator observed under a magnetic field. The fact that magnetic field sweeps at different values of n_s show a smooth transition from a situation where the sample is only insulating near the half integer filling factors to one where it is insulating everywhere also suggests that the zero field insulator is caused by the same phenomenon as the reentrant insulator observed in the previous chapter. It is therefore likely that the insulating state observed at zero field is not due to single particle localization but is a collective state better described by a pinned Wigner crystal.

5.5.1 Varying Length Scales

In the previous chapter the fact that the threshold field and activation energy both disappear at the same critical density and that the activation energy decreases as the electric field increases were used to suggest that both temperature and applied field were affecting the same transport mechanism. This means that a length scale, ξ, can be defined for this conduction mechanism using the formula:

$$\xi \approx \frac{\Delta_0}{eE_t} \tag{5.1}$$

where Δ_0 is the value of the activation energy for very low currents (i.e. close to zero applied electric field). In the zero field case, this length scale, is typically in the μm range, and is one or two orders of magnitude larger than the typical de Broglie wavelength, λ_{dB}, for electrons at these temperatures (20 mK to 500 mK) and three or four orders of magnitude larger than the typical interelectron spacing, a. This demonstrates that the conduction mechanism is either a multi-electron effect so that $\xi \approx \Delta_0/neE_t$, where n is the number of electrons involved, or a single electron effect which occurs over long distances. Neither of these possibilities is compatible with the model of strong localization at individual impurities. It is possible, however, that the insulator corresponds to a localized state with conduction through some type of percolation or hopping conductivity. The average length of a single hop in the hopping conductivity model, for instance, can easily be much larger than either a or λ_{dB}.

At zero magnetic field, the analysis of length scales was taken a step further. ξ_0, the length scale for the zero field conduction mechanism was calculated at different electron densities. Figure 5.8 shows a logarithmic plot of ξ_0 versus $\delta n_s = (n_{sc} - n_s)$ using data from sample Si-2. This characteristic length scale changes with electron density, getting smaller as the density decreases below the critical density. In fact, the line of best fit reveals that $\xi_0 \propto (\delta n_s)^{-0.8}$. This is important, because of the possibility of percolation or hopping conductivity. Both of these mechanisms have a specific dependence of the localization length on electron density, as defined in Equation 2.38. This is of the form $\xi_0 \propto (\delta n_s)^{-\beta}$, where $\beta = 4/3$ in the classical percolation model and $\beta = 7/3$ if quantum tunnelling is included (i.e. in the hopping regime). Neither of these are close to $\beta = 0.8$ and thus this length scale is definitely not compatible with these types of localization models either.

Figure 5.8 also shows another length scale, L_D, plotted as a function of δn_s. This length scale is calculated by assuming that the conduction above the electric field threshold is caused by a sliding Wigner crystal which can be treated like a sliding charge density wave. This may be too simplistic, but it does give a general idea of a possible length scale for conduction above threshold. The model used was taken from [Nor92], and is described in Section 2.4.2. It assumes that the CDW is weakly pinned and that the amplitude of the CDW, the pinning energy and the elastic modulus of the CDW are all constant. This may not be true in the immediate vicinity of the phase transition near n_{sc}, however, so this model may not be valid for small

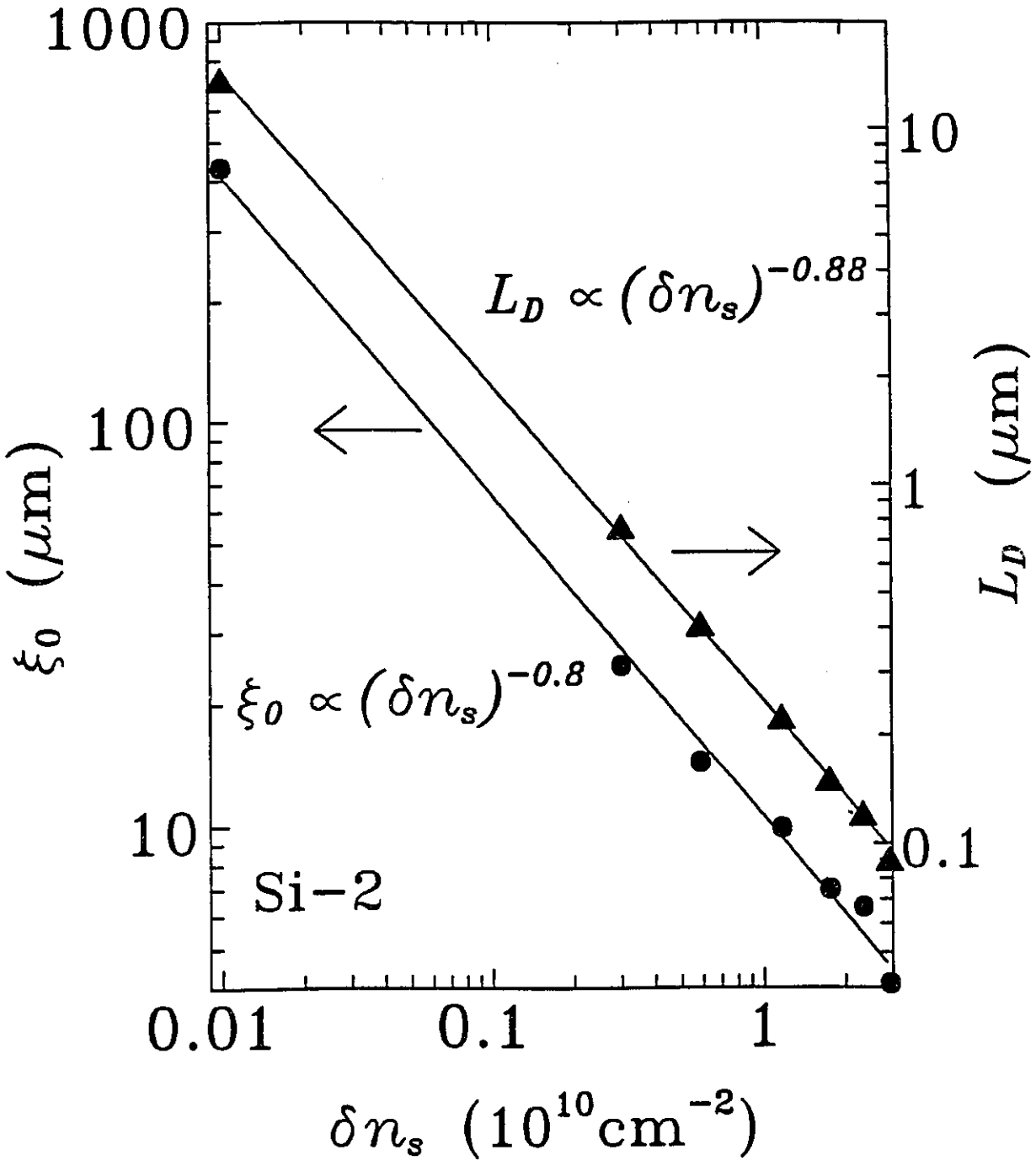

Figure 5.8: This figure shows the length scales for sample Si-2 calculated from two different models as a function of $\delta n_s = (n_{sc} - n_s)$.

values of δn_s. In the region where the model is valid, it predicts that:

$$E_t = \frac{K_T a}{2\pi e n_s L_D^2} \approx \frac{0.19 \times 10^{-10}}{L_D^2} \frac{\text{V}}{\text{cm}} \tag{5.2}$$

where K_T is the transverse shear modulus of the CDW, a is the interelectron spacing, and L_D is the correlation length (or domain length) in the CDW. It is therefore possible to use the experimental data for E_t at different values of n_s to calculate L_D as a function of n_s (and thus of δn_s). The plot in Figure 5.8 reveals that L_D can also be fitted with a power law dependence on δn_s, but in this case the exponent is $\beta = 0.88$. It is interesting to note that the values of L_D were much lower than the corresponding values of ξ_0 for the same values of δn_s. The lowest value of L_D is only about four times larger than a at that electron density, which indicates that the domains can become very small if the density is decreased far below n_{sc}. Eventually L_D will diminish sufficiently that it is of the same order as a, at which point the model of a CDW with coherent domains no longer makes sense since each domain will only include one electron. At that point the model of single particle localization would be more appropriate and L_D would no longer be a valid length scale.

5.5.2 Single Particle Localization

In seeking an explanation for the insulating phase, several types of localization can be quickly dismissed. Magnetic freeze out or a lifting of the extended states above the Fermi level due to broadened Landau levels are both impossible, since there is no magnetic field present. Mott localization is also extremely unlikely due to the random nature of the impurity potentials in a Si-MOSFET. Inhomogeneities in the sample leading to an effective narrowing of the sample width as the electron density is reduced are also ruled out because of the capacitance measurements which were described in Chapter 4.

The only single particle localization model which is reasonable is that of Anderson localization. There are several specific pieces of experimental evidence, however, which indicate that Anderson localization is not responsible for the insulating phase. The localization length, ξ_0, calculated from the threshold field is much larger than the de Broglie wavelength and the interelectron spacing, so that the electrons do not appear to be localized at individual impurity sites. The electron density dependence of ξ_0 is incompatible with the predictions for percolation or hopping conductivity, so that localization at clusters of resonant sites where several electrons are shared by a group of impurities (and thus ξ_0 can be as large as the

cluster size) is also ruled out. The fact that the conductivity is thermally activated with $\rho_{xx}(T) \propto \exp(\Delta/T)$ rather than $\rho_{xx}(T) \propto \exp(T_0/T)^{1/3}$ or $\rho_{xx}(T) \propto \exp(T_1/T)^{1/2}$ also discriminates against variable range hopping, which is the expected conduction mechanism in an Anderson localized system at low temperatures. Finally, the differential resistance in the above threshold regime is typically larger than $10^6\,\Omega$ and the conductivity is still thermally activated, indicating that the threshold does not represent a depinning of electrons from impurities back into a normal two dimensional electron gas state.

Since single particle localization can not explain many of the experimental observations and is actually in direct contradiction with some of them, it is necessary to seek some other explanation for the insulating state. The most likely candidate for a two dimensional collective insulator is a pinned Wigner crystal or charge density wave, so a detailed study of how the predictions of these models compare with the experimental data must be carried out.

5.5.3 CDW - Phase Slip Model

The qualitative predictions about the transport characteristics of a pinned CDW were listed in Section 2.4.2. These include thermally activated transport, non-linear $I-V$ curves, and a current dependent noise frequency spectrum in the above threshold regime. Above a melting temperature, T_M, all of these characteristics disappear. These predictions are supported by the data. The $I-V$ curves (See Figure 5.3) are definitely non-linear with a threshold field above which the differential resistance decreases and the conduction both below and above the threshold is thermally activated (See Figure 5.2). There is also a definite melting temperature ($T_M \sim 500\,\mathrm{mK}$) beyond which all these characteristics disappear. Figure 5.5 shows the increase in the background noise over a wide frequency range which is expected as the threshold field is exceeded, and Figure 5.6 demonstrates that this noise exhibits distinct peaks in its frequency spectrum in the $4-9\,\mathrm{kHz}$ range. This is compatible with Equation 2.42, which describes the fundamental frequency of the voltage noise spectrum for a sliding CDW as $f_0 = j_{CDW}/en_s\lambda_{CDW}$, where λ_{CDW} is the wavelength of the charge density wave. For a two dimensional system with Hall bar geometry, j_{CDW} can be replaced by I_x/y, where I_x is the source-drain current and y is the width of the Hall bar, so that $f_0 = I_x/en_s y\lambda_{CDW}$. For a fundamental frequency of $f_0 \approx 5\,\mathrm{kHz}$ with $n_s = 8.32 \times 10^{10}\,\mathrm{cm}^{-2}$, $I_x = 50\,\mathrm{pA}$ and $y = 0.8\,\mathrm{mm}$, this gives $\lambda_{CDW} \approx 90\,\mathrm{nm}$. This is a reasonable value when compared with the interelectron spacing at this density of $a \approx 20\mathrm{nm}$. Finally, Figure 5.7 shows that the spectrum moves upwards in

frequency as the current is increased, also in accordance with Equation 2.42, although the exact linear dependence of fundamental frequency on current is not experimentally observed in this preliminary measurement.

An even more specific prediction of the pinned CDW model is related to the phase slip model for conduction which has been extensively developed with respect to one dimensional charge density waves [Bor86, Gil86, Nad92]. This model assumes that the CDW is broken into many small phase coherent segments (equivalent to individual crystallites in the Wigner crystal model) and that charge transport occurs by a series of phase slips in which the phase of the CDW within a single segment changes by 2π, shifting the electrons in that segment by one CDW wavelength with respect to the stationary impurities. Phase slips are propogated through the sample as 2π phase solitons, stable travelling waves of phase displacement. If many phase solitons are present, such as in the region beyond the threshold, they have been predicted to aggregate into multi-soliton complexes which are equivalent to dislocations [Bra91, Nad92]. The soliton complexes are then equivalent to the dislocations which are responsible for charge transport in a pinned Wigner crystal framework. This model was discussed in Section 2.5.4, and it was shown that it could be related to a two dimensional system, resulting in an equation which expressed the relationship between conductivity, σ_{xx}, applied field, E_{app}, and temperature, T. Repeating the final result previously shown in Equation 2.58 gives:

$$\sigma_{xx} = \frac{I_0 l}{V_{xx} W} \exp\left[\frac{-\Delta}{T}\right] \sinh\left[\frac{neaV_{xx}}{lk_b}\frac{1}{T}\right] \tag{5.3}$$

In this equation, l is the distance between the contacts used to measure V_{xx} so that $E_{app} = V_{xx}/l$ and W is the width of the Hall bar sample. Δ is an activation energy, n is the number of electrons involved in a single phase slip event, I_0 is a magnitude constant related to the size of the phase slip current over the entire sample and a is the interelectron spacing, which is assumed to be the distance over which the phase of the CDW changes by 2π. It is difficult to relate I_0 and n to specific experimental parameters, but Δ should be similar to the activation energy, Δ_1, which is found from standard Arrhenius plots and all the other values can be found experimentally. This gives a precise form for the temperature and applied potential dependence of the sample conductivity if the conduction in the insulating regime is indeed caused by phase slippage in a pinned CDW or Wigner crystal.

The experimental points contained in any $I - V$ curve can be converted to a $\sigma_{xx} - E_{app}$ curve for one value of T. Using different $I - V$ curves measured at different temperatures but at the same value of n_s, a picture of

σ_{xx} as a function of E_{app} and T can be developed. If the phase slip model accurately describes the charge transport mechanism in the Si-MOSFETs then a single fit to Equation 5.3 with its three unknown parameters, I_0, Δ, and n, should yield a theoretical curve for each temperature which matches the experimental $\sigma_{xx} - E_{app}$ curve measured at that temperature. Figure 5.9 gives one example of this type of fit. A pair of σ_{xx} versus E_{app} curves taken from sample Si-5 at an electron density of $6.69 \times 10^{10}\,\mathrm{cm}^{-2}$ and at two different temperatures (indicated by the values of T^{-1} on the graph) are shown. The starred points with error bars are the experimental values. The two solid lines were produced using a single least squares best fit to Equation 5.3. There is excellent agreement between both sets of experimental data and the fitted curve. This process was repeated for many other data sets at different values of n_s from sample Si-5 and for some data sets from the other two samples [Cam94]. In almost all cases the phase slip model provides fits which are within the experimental uncertainties. Data sets taken at very low temperatures, however, are not compatible with the phase slip model. This is expected, since a change in the activation energy to a much lower value, Δ_2, is observed as the temperature is lowered below approximately 120 mK, so a change in the dominant conduction mechanism at very low temperatures is compatible with the experimental data.

The fact that the phase slip model fits the experimental data so well suggests that the transport data in the temperature regime from $\approx 120\,\mathrm{mK}$ to $\approx 500\,\mathrm{mK}$ is compatible with the phase slip model. This is further validated when the values of Δ, one of the three fitting parameters in the phase slip model, are compared with the values of Δ_1 obtained from Arrhenius plots of ρ_{xx} versus T^{-1}. Figure 5.10 shows a set of five values of Δ obtained from the phase slip model using sets of $I - V$ curves measured on sample Si-5 at different values of n_s. The solid line is a line of best fit to the three middle points. The point with the lowest electron density is ignored because it does not fit with the others and is probably influenced by the beginning of the transition to single particle localization which can occur at very low densities (as mentioned in the previous section). The point with the highest density is also ignored in the linear fit because it is slightly above the critical density for sample Si-5. This is verified by the fact that the solid line of best fit to the three valid points indicates that $\Delta = 0\,\mathrm{K}$ at $n_{sc} \approx 7.7 \times 10^{10}\,\mathrm{cm}^{-2}$ which occurs before the last point. This value of n_{sc} is in excellent agreement with the critical density of $n_{sc} \approx 7.8 \times 10^{10}\,\mathrm{cm}^{-2}$ which was calculated for sample Si-5 at zero field using experimental values of Δ_1 and E_t. Thus, the phase slip model produces conductivity versus temperature and applied field curves which agree with the experimental data and activation energies which

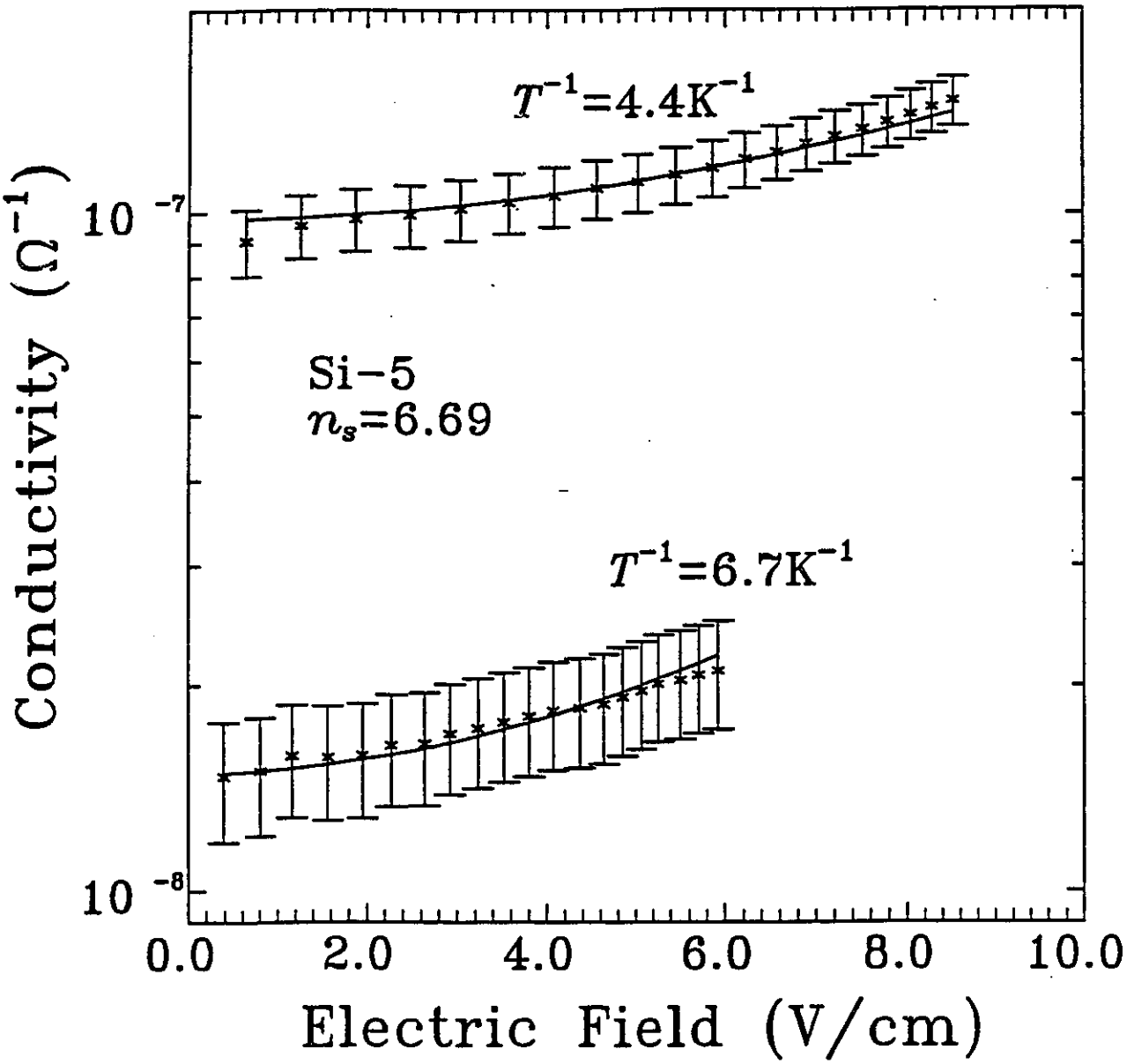

Figure 5.9: Two σ_{xx} versus E_{app} curves are plotted in this figure. The data points come from two $I-V$ curves from sample Si-5 at an electron density of $6.69 \times 10^{10}\,\mathrm{cm}^{-2}$ and at the temperatures indicated by the values of T^{-1} on the graph. The solid lines represent the results of a single fit to the data at both temperatures using the phase slip model.

are similar in size to the values of Δ_1 obtained from standard Arrhenius plots. These activation energies are linear with δn_s, just as the values of Δ_1 are (See Figure 5.4) and they lead to the same value of n_{sc}. All of this evidence suggests that the pinned CDW with charge transport through phase slippage is an excellent model for describing the zero field insulating state observed in the Si-MOSFETs.

5.5.4 Wigner Crystal - Dislocation Pair Model

If the insulating state can be explained as a charge density wave then in this very low density regime where the interelectron spacing and the Bohr radius are both of the same order of magnitude, the overlap of the electron wavefunctions will be quite small and the system can also be regarded as a Wigner crystal. Quantitatively, the pinned Wigner crystal shares the same transport characteristics as a pinned CDW. Thus all of the previous discussion with respect to activated transport, non-linear $I - V$ curves, melting temperature and noise characteristics also holds true in the Wigner crystal framework. Some more quantitative predictions are required to determine if the zero field insulating state really is best described as a pinned Wigner crystal.

Sample disorder has a strong influence on the results at zero magnetic field. All three high mobility samples (Si-2, Si-5 and Si-11) showed the same transport characteristics, but the critical density above which these characteristics disappeared was higher in the lower mobility samples and for any given value of n_s, the insulator was stronger the lower the sample mobility. It is therefore important to use a model of the Wigner crystal which takes into account the effects of impurities in the sample. In Section 2.5.4, a model developed by S. T. Chui and coworkers [Chu91a, Chu91b, Chu93b, Chu93c] was discussed. This model suggests that the transport below threshold occurs principally through the mechanism of mobile bound dislocation pairs and that above the threshold the electric field provides enough energy to unbind the dislocation pairs so that a large current is carried by a steady stream of dislocation pairs being formed and then breaking apart. The entire crystal is not sliding coherently above the threshold, but the unbound dislocations can be considered as the edges of coherent crystallites so that this model considers the motion of the crystallite boundaries rather than the crystallites themselves. It can also be related to a phase slip model, where the dislocations are equivalent to aggregated groups of phase solitons [Bra91, Nad92, Chu93b]. There are three specific predictions which arise from this model and which can be compared with the experimental results

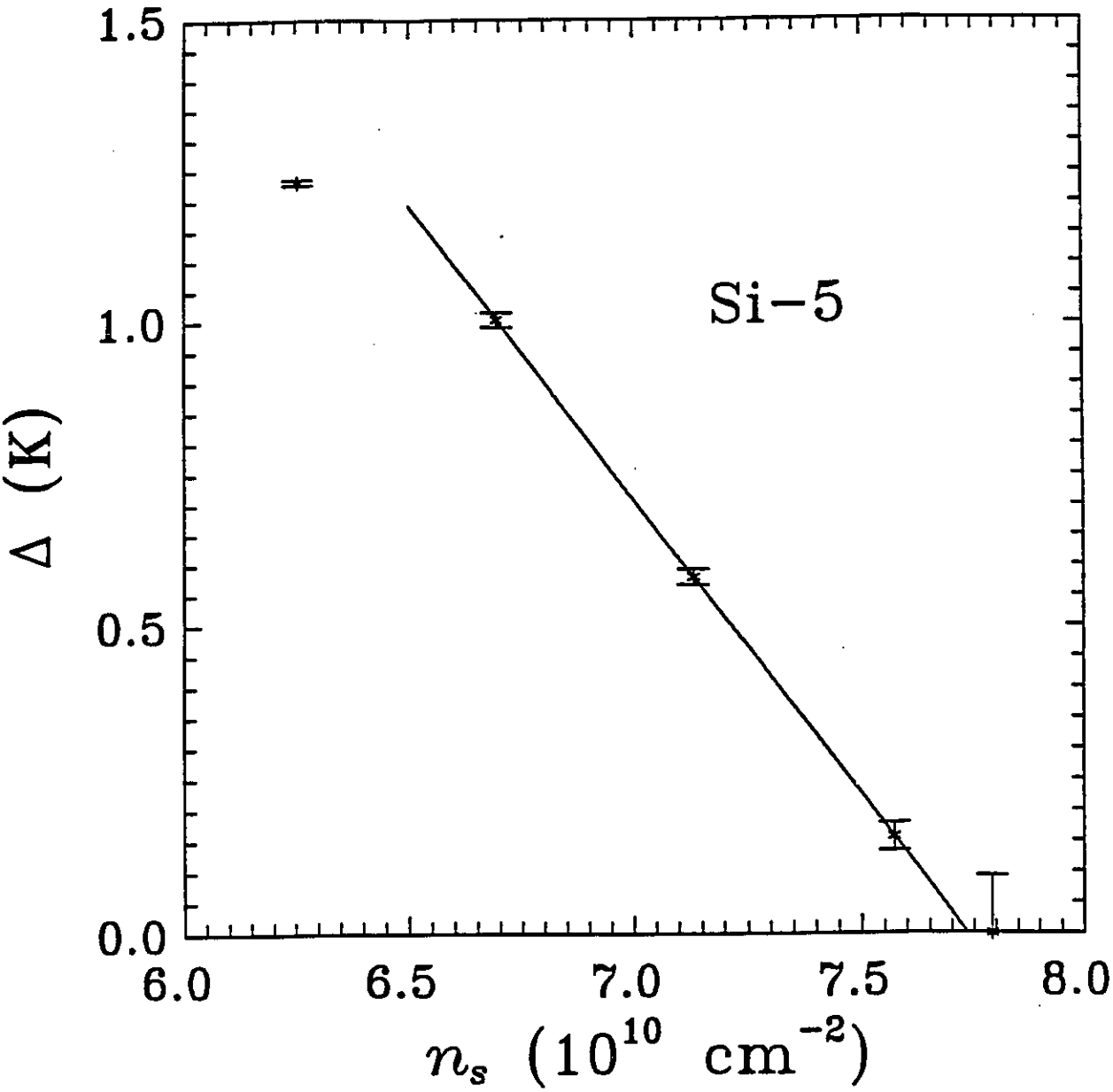

Figure 5.10: The values of the activation energy parameter from the phase slip fits to sample Si-5 at several different values of the electron density are plotted to determine the critical density. The solid line is a fit to the middle three points. It reveals that $n_{sc} \approx 7.7 \times 10^{10}\,\mathrm{cm}^{-2}$.

at zero magnetic field.

The first prediction comes from quantum Monte-Carlo calculations of the ground state energy of the quantum fluid versus the electron solid [Chu93c]. Using parameters which took into account the valley degeneracy and specific scattering mechanisms present in high mobility Si-MOSFETs such as Si-2, Si-5 and Si-11, the cold melting density at zero magnetic field was predicted to be $n_{CM} = 6.9 \times 10^{10}\,\mathrm{cm}^{-2}$. This is very close to the experimentally determined range of critical densities for the metal-insulator transition at zero magnetic field of $n_{sc} = 8-10\times10^{10}\,\mathrm{cm}^{-2}$. Thus the observed metal-insulator transition is occuring at the same electron density as the theoretically predicted phase transition from a quantum fluid to a Wigner crystal.

The second prediction is that the thermal activation energy should depend linearly on $\delta n_s = (n_{sc} - n_s)$, as originally noted in Equation 2.60. This prediction is clearly satisfied by the data (See Figure 5.4 and Figure 5.10). The third prediction relates to the specific tunneling mechanism by which the dislocation pairs are broken apart in the regime above the threshold field, E_t. It was originally stated in Equation 2.63, which gives the criterion that $E_t \propto (n_{sc} - n_s)^{1.5}$. This criterion is also satisfied by the experimental data. Figure 5.11, for instance, shows a logarithmic plot of E_t versus δn_s using data taken from sample Si-2. The solid line is the line of best fit and it gives a power law dependence of $E_t \propto \delta n_s^{1.8\pm0.2}$. This is very close to the predicted dependence which is indicated on the graph by the dashed line, and other similar plots have shown even better agreement with exponents such as 1.7 and 1.5 [Pud94a]. Thus, the dislocation pair model succesfully predicts the dependences of both Δ and E_t on δn_s.

This model of the insulator as a pinned polycrystalline Wigner solid, with transport through the motion of dislocation pairs (which are equivalent to complexes of phase solitons in a CDW) seems to explain all of the experimental data. There should be a metal-insulator transition when the electron density falls below the cold melting density, n_{CM}, and the electrons freeze into a pinned crystal, so the large values of R_{xx} associated with densities below n_{sc} are expected. The theoretical value of n_{CM} using this model is also very close to the experimental range of n_{sc}. This also explains the observed melting temperature, since a zero field Wigner crystal will always melt above some critical temperature T_M. Dislocation pair creation is a thermally activated process and beyond a certain threshold field, E_t, the electric field will break the pairs apart, allowing free dislocations to carry a large current. This explains the activated transport and the non-linear $I-V$ curves. Indeed, the electron density dependence of both E_t and Δ which is predicted from this theory agrees with their experimental dependences on δn_s. The facts

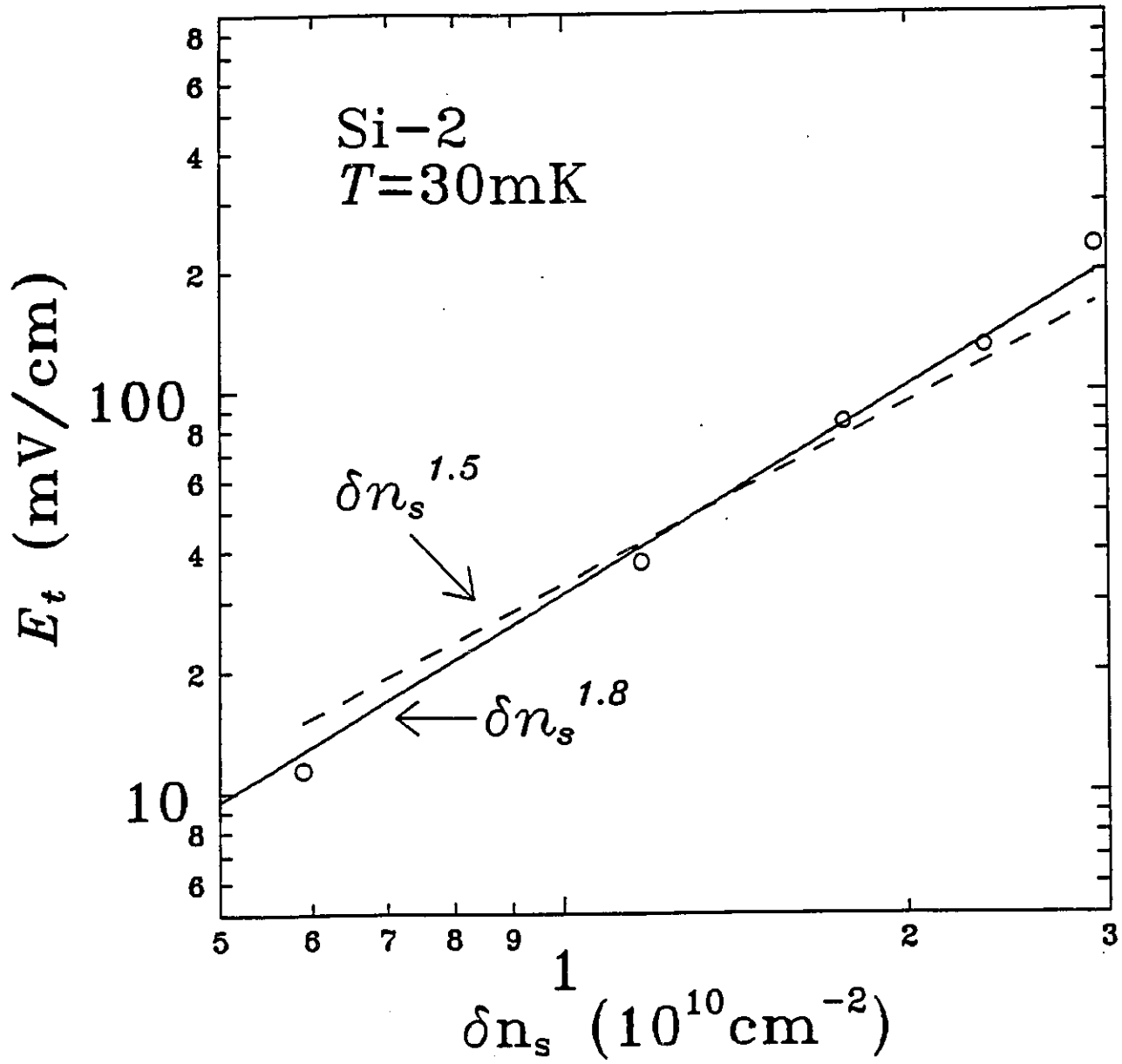

Figure 5.11: This figure shows E_t versus $\delta n_s = (n_{sc} - n_s)$ on a logarithmic scale. The data were measured from sample Si-2 at a temperature of 30 mK, and can be fit by $E_t = \delta n_s^{1.8}$, as indicated by the solid line on the graph. The dashed line shows the theoretical dependence of $E_t = \delta n_s^{1.5}$, predicted by S. T. Chui [Chu91b, Chu93b].

that the conduction mechanism is still activated above threshold and that the activation energy is observed to decrease as the electric field increases also fit in the framework of this model, since conduction both above and below the threshold occurs through the same mechanism. The increased noise above the threshold field is expected to occur as unbound dislocations move past stationary impurities. These dislocations involve the displacement of electrons from their equilibrium positions, so that although they have an underlying periodic structure, that period gets warped by the presence of the dislocation. This helps to explain why the peaks in the noise voltage spectrum are so small in comparison to the background noise and why their dependence on current does not follow the ideal model for a perfectly coherent crystal or CDW. The long length scales, ξ_0, calculated from Δ and E_t also fit within this framework, since the dislocation pair is a multielectron excitation with a finite size which is many times the interelectron spacing. Finally, when conduction is considered in terms of the equivalent model of phase slip in a CDW, the temperature and electric field dependence of the conductivity can be fit by a single phase slip equation which produces values of Δ and n_{sc} that agree with those from the more direct experimental results.

5.5.5 Zero Field Phase Diagram

There is one final calculation which ties many of the separate concepts previously mentioned together. Both of the length scales shown in Figure 5.9 are dependent on δn_s, so that if n_s is decreased sufficiently, the interelectron spacing, a, which increases as n_s decreases, will eventually become the largest length scale in the system and the single particle localization (SPL) model will become appropriate. Figure 5.12 is a plot of electron density versus peak mobility. The experimentally measured zero magnetic field values of n_{sc} for the three samples discussed in this chapter as well as for an additional low mobility sample (Si-39) and a new high mobility sample (Si-15) are plotted as filled circles. The densities coresponding to the transition to SPL were calculated for the three samples discussed in this chapter by extrapolating L_D versus δn_s from Figure 5.8 (and similar data sets for samples Si-5 and Si-11) down to $L_D = 2a$. These were plotted as the filled triangles. The dashed line is a theoretical estimate for the critical density at which the transition to SPL occurs in a Si-MOSFET. This estimate was obtained by deriving the density of impurities from the peak mobility using the techniques of A. Gold described in Section 2.3 [Gol85, Gol86a]. Another calculation by A. Gold and W. Gotze [Gol86b] of the critical density for the transition to SPL as a function of the impurity density in a Si-MOSFET was then extrapolated to

lower densities in order to obtain the final result shown. The empty squares plotted in the figure are experimental measurements of the critical density corresponding to the onset of thermally activated transport in a series of low mobility Si-MOSFETs [Yag80].

The solid line on the graph is a theoretical estimate of the cold melting density, n_{CM}, as a function of mobility in Si-MOSFETs. It was calculated in the following manner. An initial value of $n_{CM} = 4.5 \times 10^{10}\,\mathrm{cm}^{-2}$ was assumed to be reasonable for a Si-MOSFET with no impurities at zero temperature and magnetic field. This value was previously calculated in Section 2.5 using $r_{sc} = 37 \pm 5$ (as calculated by B. Tanatar and D. M. Ceperley [Tan89] for a perfectly coherent Wigner crystal) and an additional factor of 6.8 obtained by treating the two valleys in (100)-Si as separate coupled electron lattices (as described by L. Swierkowski et. al. [Swi91] for a pair of lattices in adjacent quantum wells). The effects of impurities were then introduced using the calculations of the phonon spectral density and the mean squared displacement of electrons from their equilibrium lattice positions, $\langle \delta r^2 \rangle$, developed by A. G. Eguiluz et. al. [Egu83]. The authors of this paper had explicitly calculated the critical value of $\langle \delta r^2 \rangle / a^2$ in a perfect lattice corresponding to a cold melting density of $4.5 \times 10^{10}\,\mathrm{cm}^{-2}$ and so this was taken as the melting criterion for a Wigner crystal with any level of impurities. The paper also provided the results of calculations of $\langle \delta r^2 \rangle / a^2$, as a function of the electron density, n_s, at various values of the impurity density ratio $c = n_i/n_s$ using a phonon spectral density mechanism. It was then possible for a specific value of c to determine the precise value of n_{CM} which would give the critical melting value of $\langle \delta r^2 \rangle / a^2$. Finally, since each value of peak mobility in Figure 5.12 corresponds to a specific value of n_i, each specific value of $c = n_i/n_s$ gave a curve on the density versus mobility graph. At some point this curve would intersect the horizontal line given by $n_s = n_{CM}$ for the corresponding value of c, and this gave a single point on the solid line shown in the figure. After calculating several points explicitly the rest of the line was interpolated. This means that the solid line is a purely theoretical derivation of the phase transition to a Wigner crystal and contains no fitting parameters taken from the experimental data.

The extrapolation of the solid line to lower mobilities was stopped where it intersected the dashed line corresponding to the transition to single particle localization. Any sample with peak mobility lower than approximately $12000\,\mathrm{cm}^2/\mathrm{Vs}$ will not show any evidence of a transition to a Wigner crystal with decreasing density because it will enter the SPL regime first. This results in a division of the phase diagram into three sections. At high densities the 2DES in the Si-MOSFET is metallic (denoted by *metal* on the graph) and

at low densities it is in a single particle localization state (denoted by SPL), where every electron is localized at a separate impurity site. For high mobility MOSFETs (peak mobility beyond $12000\,\mathrm{cm^2/Vs}$) there is also a narrow range of electron densities where the 2DES forms an electron solid (denoted by ES). This electron solid is not a perfectly coherent Wigner crystal, but a multi-crystallite or polycrystalline state in which the coherence length decreases with decreasing electron density until eventually the crystallites are so small that SPL becomes dominant. There is excellent agreement between the experimentally measured critical densities and the theoretically calculated phase boundaries for both the transition to an electron solid and the transition to SPL. The fact that sample Si-39 did not show the characteristics of a Wigner crystal at any density, in marked contrast to the high mobility Si-15 provides further verification of the validity of this phase diagram.

In concluding this chapter, the characteristics of the insulator observed at zero magnetic field must be summarized. This insulating phase shows a critical cold melting density and a melting temperature, non-linear $I-V$ curves with a distinct threshold electric field, thermally activated conduction, a broad band noise background which increases as the threshold is exceeded, and noise frequency peaks which shift upwards in frequency as the current is increased. The temperature dependence of the conductivity and the electron density dependence of the length scale ξ calculated from Δ_0 and E_t are both incompatible with the variable range hopping conduction mechanism which is associated with Anderson localization in this temperature regime. Sample inhomogeneity effects were ruled out in Chapter 4 and Mott localization is highly unlikely in a Si-MOSFET where the scatterers are close to the 2DES and randomly spaced. Thus this insulating phase is not compatible with traditional localization models, and an alternative explanation is required. One explanation which does seem to fit all of the data is that of a pinned Wigner crystal. In addition to the qualitative agreement between the observed characteristics and those generally predicted for a pinned Wigner crystal, there is also quantitative agreement between the data and the specific models of charge transport by dislocation pairs in a Wigner crystal or phase slip in a CDW. This comes in the form of the dependence of Δ and E_t on δn_s agreeing with the dislocation pair transport model and the dependence of σ_{xx} on E_{app} (or V_{xx}) and T agreeing with the phase slip model. Finally, theoretical calculations of the phase transition from a metallic state to an electron solid and then from the electron solid to a single particle localization state agree with the observed transitions in both the samples examined in this chapter and an additional pair of samples which were tested later by Dr. V. M. Pudalov.

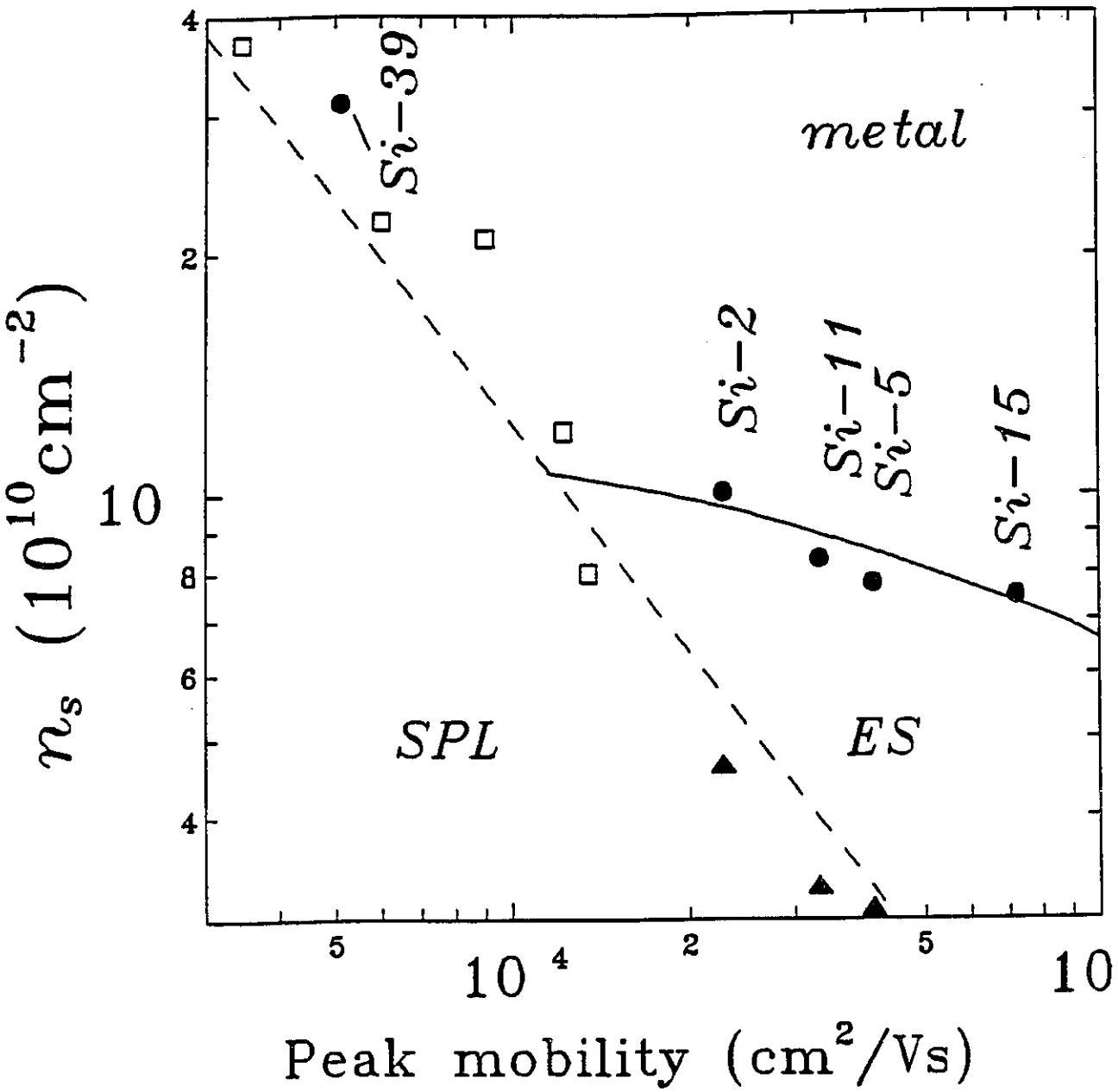

Figure 5.12: This graph shows the critical density for a phase transition versus the peak electronic mobility. The filled circles are experimental n_{sc} values and the filled triangles are experimental values for the transition to single particle localization (SPL) calculated using $L_D = 2a$. The empty squares also represent experimental points at which an Anderson transition was observed [Yag80]. The dashed line is a calculation of the critical density for a transition to SPL and the solid curve is a calculation of the cold melting density at which a Wigner crystal will form.

Chapter 6. Conclusions

Magneto-transport measurements on the two-dimensional electron system in extremely high mobility Si-MOSFETs have been described in this thesis. The experiments were carried out at very low temperatures ($20\,\mathrm{mK} < T < 4.2\,\mathrm{K}$) and magnetic fields in the range $0 - 5$ Tesla. The appearance below a critical temperature, $T_M \sim 500\,\mathrm{mK}$, and carrier density, $n_{sc} \sim 9 \times 10^{10}$ cm^{-2}, of reentrant metal-insulator transitions in the diagonal resistivity at half-filled Landau levels led to a study of the temperature and carrier density dependence of the insulating state. At even lower electron densities the samples also became insulating at zero magnetic field, although there were still well defined integer quantum Hall effect states at filling factors $\nu = 1$ and $\nu = 2$. The transport characteristics of both the zero field insulator and the half filled Landau level insulators were found to be very similar and the roles of both electron density and sample mobility in terms of relative sample disorder were studied to investigate how disorder related to the onset of the insulator.

The insulating state exhibited a thermally activated transport mechanism with the activation energy depending linearly on $\delta n_s = (n_{sc} - n_s)$. It also exhibited a threshold field behaviour in which the differential resistivity declined sharply as the applied electric field exceeded a certain critical value which depended on electron density but not significantly on temperature. Both the threshold field and the activation energy declined to zero at the same value of the critical melting density as was determined from the initial plots of R_{xx} versus n_s. The length scales calculated for the phenomenon were much larger than those associated with single particle localization models, and this, combined with the relatively small threshold fields suggested that the insulating state was more likely the result of a collective rather than a single particle phenomenon. It was observed that if n_s was decreased sufficiently, then E_t continued to increase and eventually the localization length, ξ, could approach the SPL limit (just as n_s would eventually approach n_i), but this was clearly not the case near n_{sc}. The fact that the metal-insulator transition was multiply reentrant as the magnetic field changed was unusual, but it could be explained in terms of a Wigner crystal in equilibrium with a quantum fluid. A pinned Wigner crystal also explains the activated transport, non-linear $I - V$ curves, critical melting temperature, cold melting

density and the mobility dependence of n_{sc}, in which samples with higher mobility experienced the onset of the insulating phase at lower electron densities. This possibility was further strengthened by the fact that a theoretical calculation of n_{CM} for a pinned Wigner crystal using parameters appropriate for the Si-MOSFETs studied in this thesis agreed with the experimentally determined values of n_{sc}.

The zero magnetic field insulator showed additional evidence in favour of a pinned Wigner crystal. A sharp increase in the broad band noise background measured from the sample was observed as the threshold field was exceeded. A spectral analysis of this noise revealed a series of distinct frequency peaks which shifted upwards in frequency as the current was increased. This was consistent with a pinned Wigner crystal which was depinned by the applied field and began to slide, reducing the differential resistivity beyond the depinning field and creating a voltage noise spectrum as the regularly spaced electrons in the crystal matrix moved past sharp impurity potentials in the sample. There was even more specific agreement between the experimental data and the equivalent models of charge transport by dislocation pairs in a Wigner crystal or phase slip in a CDW. This came in the form of the dependence of Δ and E_t on δn_s agreeing with the dislocation pair transport model and the dependence of σ_{xx} on E_{app} (or V_{xx}) and T agreeing with the phase slip model. Finally, theoretical calculations of the electron densities corresponding to the phase transition from a metallic state to an electron solid and then from the electron solid to a single particle localization state agreed with the observed transition densities for all the samples studies in this thesis.

Thus the best explanation for the anomalous R_{xx} and all the other characteristics that were subsequently found is that a pinned Wigner crystal is formed near the half integer filling factors in these very high mobility Si-MOSFETs when the temperature and electron density are sufficiently low, and that at even lower electron densities this Wigner crystal also appears at zero magnetic field. This means that the experiments described in this thesis were the first to show evidence for the existence of a quantum Wigner crystal in any system.

This work could be extended to better reveal the underlying periodicity of the Wigner electron lattice. This could be achieved by noise measurements using low temperature amplification of the signal or possibly an optical probe of the local order such as photoluminescence spectroscopy. It would be interesting to study other silicon based heterostructures such as Si/SiGe to confirm the role of the Coulomb interaction energy on the onset of the insulating state. This would also allow an extension of the present phase diagram

for the Wigner solid due to the higher mobilities which can be achieved in such systems.

Bibliography

[Abr79] E. Abrahams, P. W. Anderson, D. C. Licciardello and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979)

[Ale94] I. L. Aleiner and I. M. Ruzin, Phys. Rev. Lett. **72**, 1056 (1994)

[And82] T. Ando, A. B. Fowler and F. Stern, Rev. Mod. Phys. **54**, 437 (1982)

[And88] E. Andrei, G. Deville, D. C. Glattli and F. I. B. Williams, Phys. Rev. Lett. **60**, 2765 (1988)

[Ash76] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College, Philadelphia (1976)

[Azb92] M. Y. Azbel', Phys. Rev. B **45**, 4208 (1992)

[Bal70] A. Baldereschi, Phys. Rev. B **1**, 4673 (1970)

[Bar79] J. Bardeen, Phys. Rev. Lett. **42**, 1498 (1979)

[Bar80] J. Bardeen, Phys. Rev. Lett. **45**, 1978 (1980)

[Bas88] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, John Wiley and Sons, New York (1988)

[Bes92] M. Besson, E. Gornik, C. M. Engelhardt and G. Weimann, Semicond. Sci. Technol. **7**, 1274 (1992)

[Bon76] L. Bonsall and A. A. Maradudin, Surf. Sci. **58**, 312 (1976)

[Bon77] L. Bonsall and A. A. Maradudin, Phys. Rev. B **15**, 1959 (1977)

[Bor86] D. V. Borodin, S. V. Zaitsev-Zotov and F. Y. Nad, Zh. Eksp. Teor. Fiz. **90**, 318 [JETP **63**, 184] (1986)

[Bra91] N. Brazovskii and V. Matveenko, J. Phys. **1**, 269 (1991)

[Buh91] H. Buhmann, W. Joss, K. von Klitzing, I. V. Kukushkin, A. S. Plaut, G. Martinez, K. Ploog and V. B. Timofeev, Phys. Rev. Lett. **66**, 926 (1991)

[Cam94] J. W. Campbell, M. D'Iorio and V. M. Pudalov, Physica B **194-196**, 1241 (1994)

[Cep78] D. Ceperley, Phys. Rev. B **18**, 3126 (1978)

[Cha88] T. Chakraborty and P. Pietilainen, *The Fractional Quantum Hall Effect*, Springer-Verlag, Berlin, (1988)

[Cha80] K.M. Cham and R. G. Wheeler, Phys. Rev. Lett. **44**, 1472 (1980)

[Cha72] A. V. Chaplik, Zh. Eksp. Teor. Fiz. **62**, 746 [JETP **35**, 395] (1972)

[Chu91a] S. T. Chui and K. Esfarjani, Phys. Rev. B **44**, 11498 (1991)

[Chu91b] S. T. Chui and K. Esfarjani, Phys. Rev. Lett. **66**, 652 (1991)

[Chu93a] S. T. Chui, Solid. State. Comm. **86**, 605 (1993)

[Chu93b] S. T. Chui, Phys. Lett. A **180**, 149 (1993)

[Chu93c] S. T. Chui and B. Tanatar, Reports of the Bartol Research Institute, BA-93-02 (1993)

[Chu94] S. T. Chui, *Physics of Quantum Solids of Electrons*, pp. 88-139, International Press, Boston (1994)

[Col91] P. T. Coleridge, Phys. Rev. B **44**, 3793 (1991)

[Cra71] R. S. Crandall and R. W. Williams, Phys. Lett. A **34**, 404 (1971)

[Dev84] G. Deville, A. Valdes, E. Y. Andrei and F. I. B. Williams, Phys. Rev. Lett. **53**, 588 (1984)

[DIo90] M. D'Iorio, V. M. Pudalov and S. G. Semenchinsky, Phys. Lett. A **150**, 422 (1990)

[DIo92a] M. D'Iorio, V. M. Pudalov and S. G. Semenchinsky, Phys. Rev. B **46**, 992 (1992)

[DIo92b] M. D'Iorio, J. W. Campbell, V. M. Pudalov and S. G. Semenchinsky, Surf. Sci. **263**, 49 (1992)

[Dol94] V. T. Dolgopolov, G. V. Kravchenko, S. V. Kravchenko and A. A. Shashkin, Surf. Sci. **305**, 96 (1994)

[Efr75] A. L. Efros and B. I. Shklovskii, J. Phys. C **8**, L49 (1975)

[Egu83] A. G. Eguiluz, A. A. Maradudin and R. J. Elliott, Phys. Rev. B **27**, 4933 (1983)

[Eis92] J. P. Eisenstein, G. S. Boebinger, L. N. Pfeiffer, K. W. West and S. He, Phys. Rev. Lett. **68**, 1383 (1992)

[Fan70] F. F. Fang and A. B. Fowler, J. Appl. Phys. **41**, 1825 (1970)

[Fer93] H. A. Fertig, D. Z. Liu and S. D. Sarma, Phys. Rev. Lett. **70**, 1545 (1993)

[Fle79] R. M. Fleming and C. C. Grimes, Phys. Rev. Lett. **42**, 1423 (1979)

[Fuk78] H. Fukuyama and P. A. Lee, Phys. Rev. B **18**, 6245 (1978)

[Fur86] J. E. Furneaux, D. A. Syphers, J. S. Brooks, G. M. Schmiedeshoff, R. G. Wheeler and P. J. Stiles, Surf. Sci. **179**, 154 (1986)

[Fur87] J. E. Furneaux, D. A. Syphers, J. S. Brooks, G. M. Schmiedeshoff, R. G. Wheeler, G. Dorda, R. P. Smith and P. J. Stiles, *Proceedings of 18th Conf. on the Physics of Semiconductors*, World Scientific, Singapore, 421 (1987)

[Gil86] J. C. Gill, J. Phys. C **19**, 6589 (1986)

[Gla90] D. C. Glattli, G. Deville, V. Duburcq, F. I. B. Williams, E. Paris, B. Etienne and E. Y. Andrei, Surf. Sci. **229**, 344 (1990)

[Gol85] A. Gold, Phys. Rev. Lett. **54**, 1079 (1985)

[Gol86a] A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986)

[Gol86b] A. Gold and W. Gotze, Phys. Rev. B **33**, 2495 (1986)

[Gol92] B. B. Goldberg, D. Heiman, A. Pinczuk, L. Pfeiffer and K. W. West, Surf. Sci. **263**, 9 (1992)

[Gol90] V. J. Goldman, M. Santos, M. Shayegan and J. E. Cunningham, Phys. Rev. Lett. **65**, 2189 (1990)

[Gol93] V. J. Goldman, J. K. Wang, B. Su and M. Shayegan, Phys. Rev. Lett. **70**, 647 (1993)

[Gor73] L. P. Gor'kov and D. M. Chernikova, Zh. Eksp. Teor. Fiz. Pis'ma Red. **18**, 119 [JETP Lett. **18**, 68] (1973)

[Gri79] C. C. Grimes and G. Adams, Phys. Rev. Lett. **42**, 795 (1979)

[Gru88] G. Gruner, Reviews of Modern Physics **60**, 1129 (1988)

[Hal94] B. I. Halperin, Surf. Sci. **305**, 1 (1994)

[Har85] J. P. Harrang, R. J. Higgins, R. K. Goodall, P. R. Jay, M. Laviron and P. Delescluse, Phys. Rev. B **32**, 8126 (1985)

[Hoc75] R. W. Hockney and T. R. Brown, J. Phys. C **8**, 1813 (1975)

[Ima84] M. Imada and M. Takahashi, J. Phys. Soc. Jpn. **53**, 3770 (1984)

[Iof60] A. F. Ioffe and A. R. Regel, Prog. Semicond. **4**, 237 (1960)

[Jai90] J. K. Jain, S. A. Kivelson and N. Trivedi, Phys. Rev. Lett. **64**, 1297 (1990)

[Jia89] H. W. Jiang and A. J. Dahm, Phys. Rev. Lett. **62**, 1396 (1989)

[Jia90a] H. W. Jiang and A. J. Dahm, Surf. Sci. **229**, 352 (1990)

[Jia90b] H. W. Jiang, R. L. Willet, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer and K. W. West, Phys. Rev. Lett. **65**, 633 (1990)

[Jia91] H. W. Jiang, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer and K. W. West, Phys. Rev. B **44**, 8107 (1991)

[Jia94] H. W. Jiang, I. Glozman, C. E. Johnson and S. T. Hannahs, Surf. Sci. **305**, 120 (1994)

[Kas87] M. A. Kastner, G. A. Thomas and S. R. Ovshinsky, *Disordered Semiconductors*, Plenum Press, New York (1987)

[Khm83] D. E. Khmel'nitzkii, Pis'ma Zh. Eksp. Teor. Fiz. **38**, 454 [JETP Lett. **38**, 552] (1983)

[Kit86] C. Kittel, *Introduction to Solid State Physics*, John Wiley And Sons, New York, 6^{th} Ed. (1986)

[Kiv92] S. Kivelson, D. H. Lee, and S. C. Zhang, Phys. Rev. B **46**, 2223 (1992)

[Kli80] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980)

[Koh55] W. Kohn and J. M. Luttinger, Phys. Rev. **98**, 915 (1955)

[Kra91] S. V. Kravchenko, Jos A. A. J. Perenboom and V. M. Pudalov, Phys. Rev. B **44**, 13513 (1991)

[Kra94] S. V. Kravchenko, J. E. Furneaux, and V. M. Pudalov, Phys. Rev. B **49**, 2250 (1994)

[Kru91a] G. H. Kruithof, *High Mobility MOSFETs With Normal And Superconducting Gates*, Ph.D. Thesis, Rijksuniversiteit Groningen (1991)

[Kru91b] G. H. Kruithof, T. M. Klapwijk and S. Bakker, Phys. Rev. B **43**, 6642 (1991)

[Kuk94] I. V. Kukushkin, V. I. Fal'ko, R. J. Haug, K. von Klitzing, K. Eberl and K. Totemayer, Phys. Rev. Lett. **72**, 3594 (1994)

[Lan92] P. T. Landsberg, *Handbook On Semiconductors - Vol. 1: Basic Properties of Semiconductors*, Chap. 19, North-Holland, Amsterdam 2^{nd} Ed. (1992)

[Lau83a] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983)

[Lau83b] R. B. Laughlin, Phys. Rev. B **27**, 3383 (1983)

[Lau84] R. B. Laughlin, Phys. Rev. Lett. **52**, 2304 (1984)

[Lut55] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955)

[Mad91] O. Madelung, *Semiconductors: Group IV Elements and III-V Compounds*, Springer-Verlag, Berlin (1991)

[Mel90] C. J. Mellor and W. F. Vinen, Surf. Sci. **229**, 368 (1990)

[Mot75] N. F. Mott, M. Pepper, S. Pollit, R. H. Wallis and C. J. Adkins, Proc. R. Soc. London Ser. A **345**, 169 (1975)

[Nad92] F. Y. Nad and P. Monceau, Phys. Rev. B **46**, 7413 (1992)

[Nor92] B. G. A. Normand, P. B. Littlewood and A. J. Millis, Phys. Rev. B **46**, 3920 (1992)

[Paa92] M. A. Paalanen, R. L. Willet, R. R. Ruel, P. B. Littlewood, K. W. West and L. N. Pfeiffer, Phys. Rev. B **45**, 13784 (1992)

[Pep74] M. Pepper, S. Pollitt, C. J. Adkins and R. E. Oakley, Phys. Lett. A **47**, 71 (1974)

[Pep75] M. Pepper, S. Pollitt, C. J. Adkins and R. A. Stradling, Crit. Rev. Solid State Sci. **5**, 375 (1975)

[Pla74] P. M. Platzmann and H. Fukuyama, Phys. Rev. B **10**, 3150 (1974)

[Pra90] R. E. Prange and S. M. Girvin, *The Quantum Hall Effect*, Springer-Verlag, New York, 2nd Ed. (1990)

[Pru85] A. M. M. Pruisken, Phys. Rev. B **32**, 2636 (1985)

[Pud84] V. M. Pudalov and S. G. Semenchinsky, Pis'ma Zh. Eksp. Teor. Fiz. **39**, 143 [JETP Lett. **39**, 170] (1984)

[Pud93a] V. M. Pudalov, M. D'Iorio, S. V. Kravchenko and J. W. Campbell, Phys. Rev. Lett. **70**, 1866 (1993)

[Pud93b] V. M. Pudalov, M. D'Iorio and J. W. Campbell, Pis'ma Zh. Eksp. Teor. Fiz. **57**, 592 [JETP Lett. **57**, 608] (1993)

[Pud94a] V. M. Pudalov, M. D'Iorio and J. W. Campbell, Physica B **194-196**, 1291 (1994)

[Pud94b] V. M. Pudalov, Bull. Amer. Phys. Soc. **39**, 805 (1994)

[Pud94c] V. M. Pudalov and S. T. Chui, Phys. Rev. B **49**, 14062 (1994)

[Ros85] T. F. Rosenbaum, S. B. Field, D. A. Nelson and P. B. Littlewood, Phys. Rev. Lett. **54**, 241 (1985)

[Ruz92] I. M. Ruzin, S. Marianer and B. I. Shklovskii, Phys. Rev. B **46**, 3999 (1992)

[Saj90] T. Sajoto, Y. W. Suen, L. W. Engel, M. B. Santos, and M. Shayegan, Phys. Rev. B **41**, 8449 (1990)

[Sar85] S. Das Sarma and F. Stern, Phys. Rev. B **32**, 8442 (1985)

[Sha81] B. Shapiro and E. Abrahams, Phys Rev B **24**, 4025 (1981)

[Shk84] B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Smiconductors*, Springer-Verlag, Berlin (1984)

[Son90] P. C. van Son, G. H. Kruithof and T. M. Klapwijk, Phys. Rev. B **42**, 11267 (1990)

[Ste67a] F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967)

[Ste67b] F. Stern, Phys. Rev. Lett. **18**, 546 (1967)

[Sue92] Y. W. Suen, L. W. Engel, M. B. Santos, M. Shayegan and D. C. Tsui, Phys. Rev. Lett. **68**, 1379 (1992)

[Sum93] G. M. Summers, R. J. Warburton, J. G. Michels, R. J. Nocholas, J. J. Harris and C. T. Foxon, Phys. Rev. Lett. **70**, 2150 (1993)

[Swi91] L. Swierkowski, D. Neilson and J. Szymanski, Phys. Rev. Lett. **67**, 240 (1991)

[Tan89] B. Tanatar and D. M. Ceperley, Phys. Rev. B **39**, 5005 (1989)

[Tho78] D. J. Thouless, J. Phys. C **11**, L189 (1978)

[Tsu82] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982)

[Tur92] A. J. Turberfield, S. R. Haynes, P. A. Wright, R. A. Ford, R. G. Clark, J. F. Ryan, J. J. Harris and C. T. Foxon, Surf. Sci. **263**, 1 (1992)

[Vie91] O. Viehweger and K. B. Efetov, Phys. Rev. B **44**, 1168 (1991)

[Wei85] H. P. Wei, D. C. Tsui and A. M. M. Pruisken, Phys. Rev. B **33**, 1488 (1985)

[Wei88] H. P. Wei, D. C. Tsui, M. A. Paalanen and A. M. M. Pruisken, Phys. Rev. Lett. **61**, 1294 (1988)

[Wei91] C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures : Fundamentals and Applications*, Academic Press Inc., San Diego (1991)

[Wig34] E. Wigner, Phys. Rev. **46**, 1002 (1934)

[Wig38] E. Wigner, Trans. Faraday Soc. **34**, 678 (1938)

[Wil87] R. L. Willett, J. P. Eisenstein, H. L. Stormer, D. C. Tsui, A. C. Gossard and J. H. English, Phys. Rev. Lett. **59**, 1776 (1987)

[Wil88] R. L. Willett, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer, K. W. West and K. W. Baldwin, Phys. Rev. B **38**, 7881 (1988)

[Wil71] R. Williams and R. S. Crandall, Phys. Lett. A **36**, 35 (1971)

[Wil91] F. I. B. Williams, P. A. Wright, R. G. Clark, E. Y. Andrei, G. Deville, D. C. Glattli, O. Probst, B. Etienne, C. Dorin, C. T. Foxon and J. J. Harris, Phys. Rev. Lett. **66**, 3285 (1991)

[Yag80] A. Yagi and M. Nakai, Surf. Sci. **98**, 174 (1980)

[Zet82] A. Zettl, G. Gruner and A. H. Thompson, Phys. Rev. B **26**, 5760 (1982)

[Zha92] S. C. Zhang, S. Kivelson and D. H. Lee, Phys. Rev. Lett. **69**, 1252 (1992)

[Zhe94] L. Zheng and H. A. Fertig, Phys. Rev. B **50**, 4984 (1994)