On The Thermodynamics at Quantum Phase Transitions in Two Dimensional Electron Systems

Babak Hosseinkhani

On The Thermodynamics at Quantum Phase Transitions in Two Dimensional Electron Systems

Babak Hosseinkhani

8 September 2004

On The Thermodynamics at Quantum Phase Transitions in Two Dimensional Electron Systems

PROEFSCHRIFT

TER VERKRIJGING VAN
DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT LEIDEN,
OP GEZAG VAN DE RECTOR MAGNIFICUS DR. D.D. BREIMER,
HOOGLERAAR IN DE FACULTEIT DER WISKUNDE EN
NATUURWETENSCHAPPEN EN DIE DER GENEESKUNDE,
VOLGENS BESLUIT VAN HET COLLEGE VOOR PROMOTIES
TE VERDEDIGEN OP WOENSDAG 8 SEPTEMBER 2004
KLOKKE 14.15 UUR

DOOR

BABAK HOSSEINKHANI

geboren te Kermanshah in 1973

Promotiecommissie:

Promotor:	Prof. dr. J. Zaanen
Referent:	Prof. dr. J.M.J. van Leeuwen
Overige Leden:	Prof.dr. C.W.J. Beenakker
	Dr. C. Filippi
	Dr. A. Hoekstra
	Prof. dr. P.H. Kes
	Prof.dr. T.M. Klapwijk (T.U. Delft)
	Prof.dr. T. Odijk (T.U. Delft)

Het onderzoek beschreven in dit proefschrift is uitgevoerd als onderdeel van het wetenschappelijke programma van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en de Stichting voor Fundamenteel Onderzoek der Materie (FOM).

The research described in this thesis has been carried out as part of the scientific program of the Netherlands Organization for Scientific Research (NWO) and the Foundation for Fundamental Research on Matter (FOM).

Contents

Chapter 1

Introduction

Understanding the basic physical processes determining whether a material is a conductor or an insulator continues to be a central theme of Condensed Matter Physics. In recent years, much progress has been achieved in understanding weakly disorder metals [1], but the physics of the metal-insulator transition (MIT) in presence of disorder remains largely an open problem.

According to the scaling theory of localization [2] there can be no metallic state in two dimensions in zero magnetic field ($B = 0$). Within this two decades-old theory, all carriers are localized in an infinitely large two-dimensional (2D) system at zero temperature. With decreasing temperature the resistance is expected to grow logarithmically (weak localization) or exponentially (strong localization), becoming infinite as $T \rightarrow 0$. Although this prediction was made for 2D systems of non-interacting particles, subsequent theoretical work showed that weak interactions between the electrons increase the localization even further [3]. In the opposite limit of very strong interactions between particles, a 2D electron system is expected to become a Wigner crystal [4]; in the presence of even a small amount of disorder, such a crystal is expected to be pinned so that the system of crystallized electrons would not conduct at zero temperature. Therefore, 2D systems were not expected to be conducting in either limit: weak (or absent), or very strong interactions between carriers. Experiments performed in the early 1980's on various 2D system confirmed these predictions. Thin metallic films and silicon metal-oxide-semiconductor field-effect transistors (MOSFETs) displayed the expected logarithmic increase in resistivity [5]. At (relatively) low electron densities, an exponential increase of the resistivity of silicon MOSFETs as a function of inverse temperature was reported [5][1]. The agreement between theoretical expectations and experimental results was convincing, and for nearly two decades, the question of whether a conducting state is possible in 2D was considered resolved.

However, from time to time, indications appeared that the accepted view may not always be correct. A number of experimental results also suggested that metallic behavior is possible in two dimensions. From an analysis of experimental data

[1] These experiments were performed in the limited range of electron densities.

obtained in GaAs/AlGaAs heterostructures, Gold and Dolgopolov [6, 62] concluded that a metal-insulator transition existed in clean samples. These observations were not well-recognized and the conventional wisdom persisted that there can be no metallic state in two dimensions in the absence of magnetic field.

In the view of accumulated theoretical and empirical wisdom acquired over all these years, the experimental observation made in 1995 [7] of a MIT in two-dimensions was a major surprise and is a subject of great current controversy. Systematic studies of the temperature dependence of the resistance in zero magnetic field in a variety of dilute, low-disordered 2D systems have suggested that this earlier point of view may be incorrect. Metallic behavior (resistivity that decreases with decreasing temperature) has been observed down to the lowest accessible temperatures at electron (n) or hole (p) densities above some critical density n_c (or p_c). Below this critical density, the behavior of the resistance is insulating, thus suggesting that a metal-to-insulator transition in two dimensions occurs as the density is varied. At the critical density, the resistivity is found to be nearly independent of temperature and of the order of the quantum unit of resistance, $h/e^2 \approx 25.6\ k\Omega$.[2]

Application of an external magnetic field of the order of a few Tesla, either parallel, tilted, or perpendicular to the 2D plane, suppresses the metallic behavior and gives rise to an enormous positive magneto-resistance on both sides of the transition. Neither the metallic behavior nor its suppression by a magnetic field is currently understood.

1.1 MIT experiments

As mentioned earlier, soon after the publication of the theory of weak localization its prediction were seemingly verified in experiments on thin metallic films and silicon metal-oxide-semiconductor field-effect transistors (MOSFET)[3]. These experiments displayed the expected logarithmic increase (weak localization) in resistivity as well as an exponential behavior (strong localization) of that as a function of temperature at relatively low electron densities [5].

The more recent experiments on a variety of samples in a wide range of density and higher purity than earlier have re-focused attention on the problem of disorder and interaction in two dimensions. Here we divide these experiments in two different groups. The first group focuses on transport properties: experiments involving resistivity dependence on temperature, electron density, electric and magnetic fields [8]. And the other one includes non-transport behavior: experiments like measurement of compressibility and thermopower versus density of charge carriers in the sample as well as measurements on thermodynamic spin susceptibility and effective mass [9]. We give a very brief review of the most important ones among these experiments and pay more attention to the second group (ther-

[2]Recall that in two dimensions, resistivity and resistance per square are the same quantity.

[3]These samples were not very clean with a high density of charge carries and resistivity of order of $10^{-2}\ h/e^2$.

modynamics) in the following chapters, since this is in fact the main focus of the thesis.

1.1.1 Transport

Transport experiments in the absence and presence of a magnetic field B have been performed in variety of different 2D electron (hole) systems like Si-MOSFET's and p-GaAs/AlGaAs. The essential messages of these experimenters is summarized as following:

- As can be seen from Fig.1.1 and Fig.1.2, strong metallic temperature dependence of the resistivity ($d\rho/dT > 0$) is observed in clean dilute 2D systems at carrier densities above some critical value, while insulating ($d\rho/dT < 0$) behavior is seen at densities below the critical value. The metallic behavior starts at temperatures below some fraction of the Fermi temperature and continues down to the lowest accessed temperatures, $T/T_F < 10^{-2}$. At the critical density, there appears to be a transition from a metallic-like phase to a strongly localized one. The latter phase is what is expected for 2D systems under the conditions where metallic behavior is seen in the present experiments .

- Metallic behavior persists to rather high carrier densities (smaller r_s) but its relative strength decreases with density. A weak insulating temperature dependence reminiscent of Anderson localization is observed at higher densities of the order of those used in the experiments in the 1980s.

- Another unusual property of dilute two-dimensional systems is their enormous response to an external magnetic field. Although we are not going to address this issue further, it is also one of the important features of these systems. An external magnetic field applied at arbitrary angle with respect to the 2D plane suppresses the metallic behavior and eliminates it completely at $n < 1.5n_c$. The temperature dependence of resistivity also changes drastically toward insulating behavior by turning on a magnetic field of a few Tesla, Fig. 1.3 . The suppression of metallic temperature dependence is rather independent of the orientation of the magnetic field relative to 2D plane. Note that his is not due to a change in the level of disorder or in carrier density. The suppression of the metallic behavior appears to be correlated with the degree of spin polarization. Similar qualitative behavior has been reported on p-GaAs/AlGaAs heterostructure.

- These effects have been observed in different electron and hole systems. In some 2D systems, the resistivity was found to scale with temperature and/or electric field on both sides of a critical density, and a conductivity-resistivity symmetry was observed around the transition, consistent with a quantum

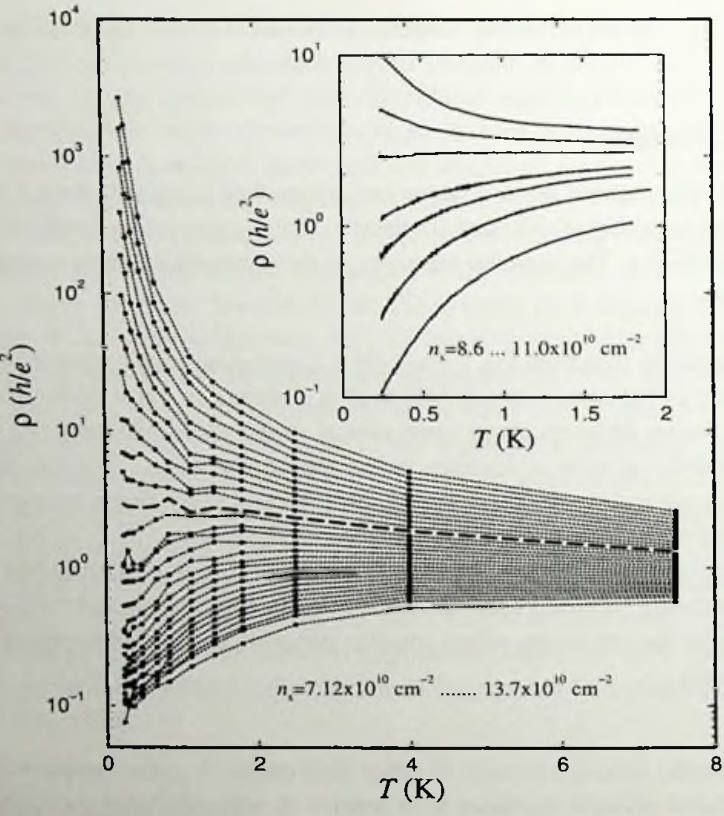


Figure 1.1: Temperature dependence of the resistivity in dilute low-disordered Si-MOSFET for 30 different electron densities (from Kravchenko *et al.* [7]). The insert shows high precision measurement of $\rho(T)$ close to the critical density for another sample; the electron densities are between 8.6 and11.0 $\times 10^{10}$ cm^{-2} (from Sarachik *et al.*[10]).

critical point describing the zero temperature metal-insulator transition. By choosing proper values for the scaling parameter $T_0(n)$ for each density n, the data collapses onto two curves: and insulating and a metallic branches. At the critical density, the resistivity below some temperature was found to be practically independent of temperature in the best samples [7].

1.1.2 Thermodynamics and Non-transport Properties

Almost all measurements performed on these systems to date concern transport: the diagonal resistivity, the Hall resistivity and the magnetoresistance. Very little is known regarding the thermodynamic behavior of 2D electron and hole systems. Clearly, if a true zero-temperature phase transition exists, it may reveal itself also in the thermodynamic properties of the 2D electron systems. This is indeed very important because more often than not non-transport behavior is easier to understand

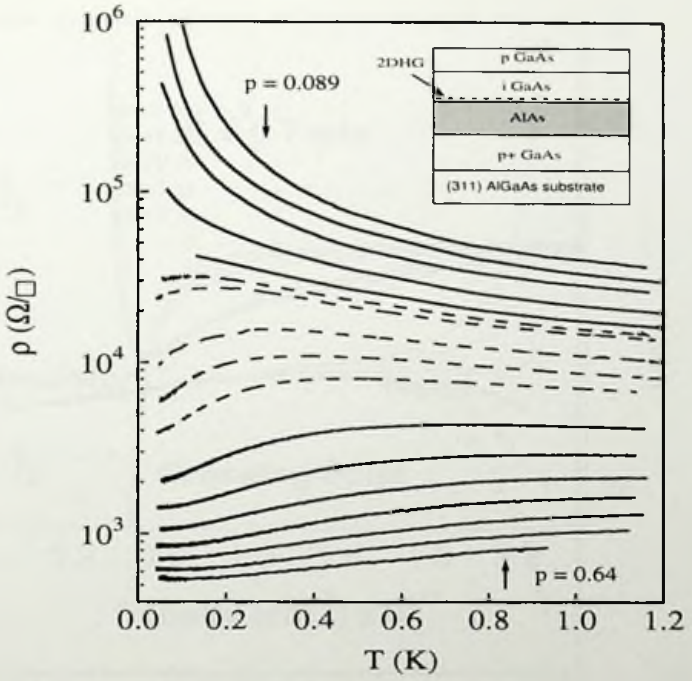


Figure 1.2: Resistivity per square as a function of temperature (for the 2D hole gas (2DHG) in p-GaAs/AlGaAs) obtained at $B = 0$ at various fixed hole densities, p, between 8.9×10^9 and 6.4×10^{10} cm^{-2}. Three different regime can be recognized quite clearly: insulating, mixed (dashed line) and metallic-like regime at low, intermediate and high density, respectively. The inset shows schematic view of a p-type ISIS structure used in the experiment (from Hanein *et al.* [11]).

(for example one does not need to worry about complications such as the emergence of quantum coherence, at the heart of transport measurements). It should be that, the nature of metal-insulator phase transition (whether classical or quantum) and the generic properties of 2D electron/hole systems near MIT can be addressed more clearly through the understanding of the thermodynamic properties of such systems. In fact recent novel and important developments in measurements of thermodynamic properties of 2D systems near the MIT have made it possible to look into this phenomena of the 2D electron/hole systems from different perspective; and understanding this important aspect of the MIT is the main focus of the research presented in this thesis. Here we briefly mention some of the measurements involving thermodynamic properties and later we will address the most crucial findings in more detail:

- *Compressibility*- Measurements of compressibility have recently been reported by two experimental groups. Using a technique based on measurement

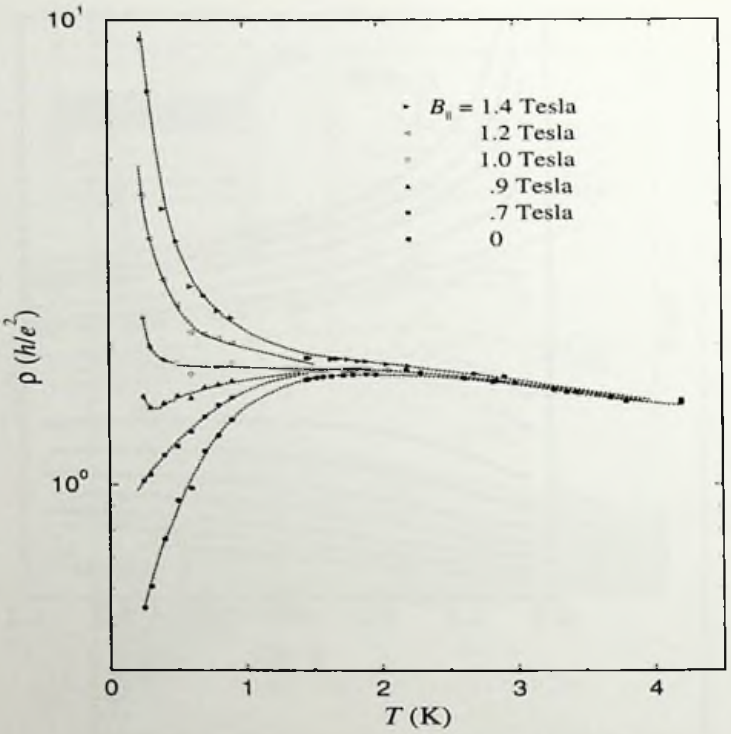


Figure 1.3: Resistivity versus temperature for five different fixed magnetic fields applied parallel to the plane of a Si-MOSFET. The electron density is $8.83 \times 10^{10}\ cm^{-2}$ (from Simonian *et al.* [12]).

of the capacitance, Dultz and Jiang [13] found that the compressibility of 2D holes in p-GaAs/AlGaAs heterostructure changes sign at the critical density for the metal-insulator transition. Ilani *et al.* [14, 15] determined the compressibility from measurements of the local chemical potential using single electron transistors. They found qualitatively different behavior of the compressibility at low and high electron densities, with a crossover density that again agrees quantitatively with the transport critical density. Both experiments suggest that the system undergoes a thermodynamic change at the transition. The behavior of the compressibility is a key signature of the nature of the metal-insulator transition. The freezing of the electron liquid into a disordered Wigner solid, for example, should be accompanied by a change of the compressibility from negative to positive. Fig. 1.4 shows this typical behavior near the transition point. This will be discussed in more detail in the chapter 4.

- *Thermopower-* Fletcher *et al.* [16] performed thermopower measurements in high-mobility silicon MOSFETs and found that the diffusion thermopower diverges at $n = n_c$ in a way similar to the divergence expected [17] for a

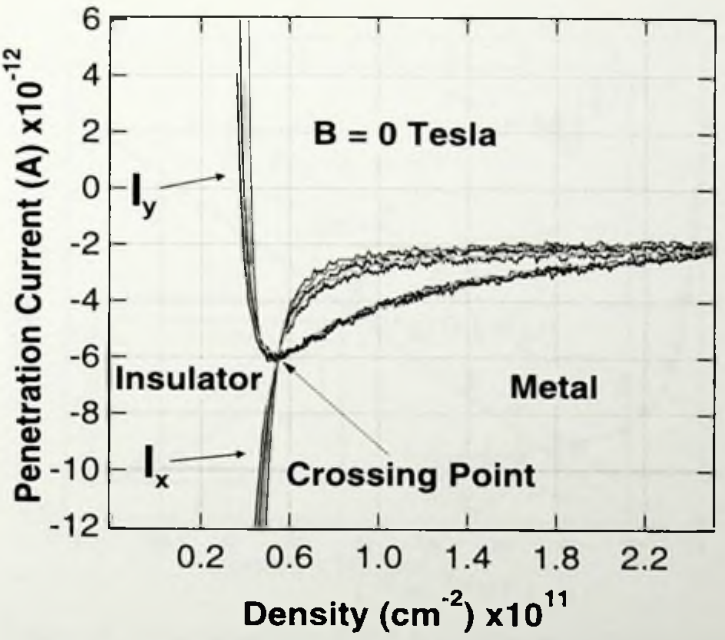


Figure 1.4: $I_x \propto -R_s$ (resistance (dissipation) of the 2DHS) and $I_y \propto \kappa^{-1}$(inverse compressibility of 2DHS) vs density for five temperatures at an excitation frequency of 100 Hz: blue- 0.33 K, green- 0.56 K, black- 0.82 K, orange- 1.02 K, red- 1.28 K. The crossing point of the five dissipation channel curves corresponds to the metal-insulator phase transition at $B = 0$. The minimum in the inverse compressibility channel occurs at the same density of $p = 5.5 \times 10^{10}\text{cm}^{-2}$ (from Dultz *et al.* [13]).

3D Anderson metal-insulator transition, and consistent with the existence of a mobility edge in two dimensions. We are not going to address this issue further in this thesis.

- *Spin Susceptibility-* In Fermi-liquid theory, the electron effective mass, m^*, and the g-factor (and, therefore, the spin susceptibility $\chi \propto g^* m^*$) are renormalized due to electron-electron interactions. In silicon MOSFETs, various experimental methods provide evidence for a sharp increase and possible divergence of the spin susceptibility at some finite sample-independent electron density, n_χ, which is at or very near the critical density for the MIT in high mobility samples [18, 19, 20]. A novel and very promising method has recently been used by Prus *et al.* to obtain direct measurements of the thermodynamic spin susceptibility. The method entails modulating the (parallel) magnetic field by an auxiliary coil and measuring the AC current induced between the gate and the 2D electron gas, which is proportional to $\partial\mu/\partial B$

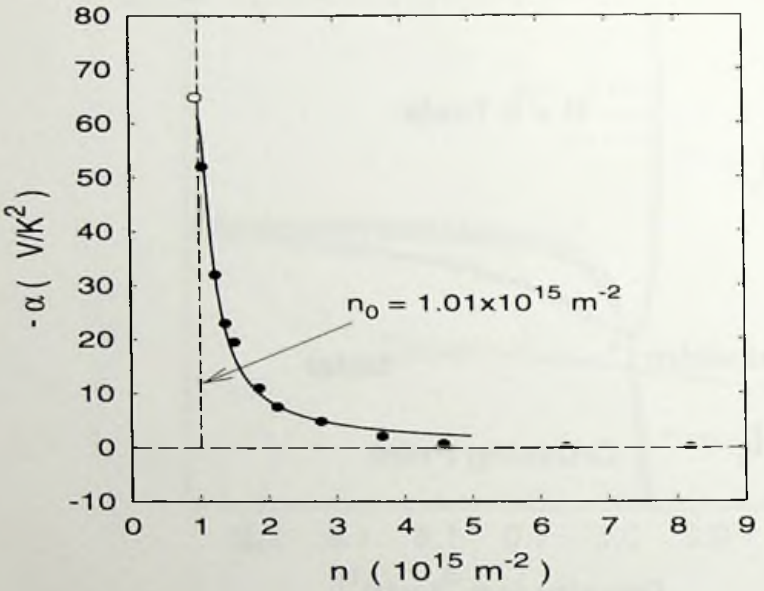


Figure 1.5: Thermopower at the MIT. Density dependence of S^d/T for a high mobility Si-MOSFET sample. The single open symbol is for $n = 0.97 \times 10^{15} m^{-2}$ which is just below n_c but $k_B T$ broadening should make this indistinguishable from n_c. The line is the best fit data based on the theory of 3D Anderson localization with $n < 4 \times 10^{15} m^{-2}$(from Fletcher *et al.* [16]).

(where μ is the chemical potential). Using Maxwell's relation,

$$\frac{\partial \mu}{\partial B} = -\frac{\partial M}{\partial n}, \tag{1.1}$$

one can obtain the magnetization M by integrating the induced current over n. The magnetic susceptibility can then be determined from the slope of the $M(B)$ versus B dependence at small fields. To date, Prus *et al.* have reported data for one sample which becomes insulating at a relatively high electron density ($1.3 \times 10^{11}\,\mathrm{cm}^{-2}$). In GaAs/AlGaAs heterostructures, a similar strong increase of the spin susceptibility at ultra-low carrier densities has now been established based on an analysis of the Shubnikov-de Haas oscillations [21]. The results are shown in Fig. 1.6 by solid symbols; χ increases by more than a factor of two as the density decreases. To describe the experimental data, an empirical equation $\chi \propto n^{-0.4}$ has been suggested [21], which works well in the entire range of electron densities spanned. Although χ tends to diverge, it is not clear from these experiments whether it does so at a finite density. Unlike the Stoner instability which entails an increase in the g-factor, the increase in the susceptibility in these systems is due to an increase of the effective mass. The effective mass is, in turn, found to be independent of the degree of spin polarization, implying that the increase is

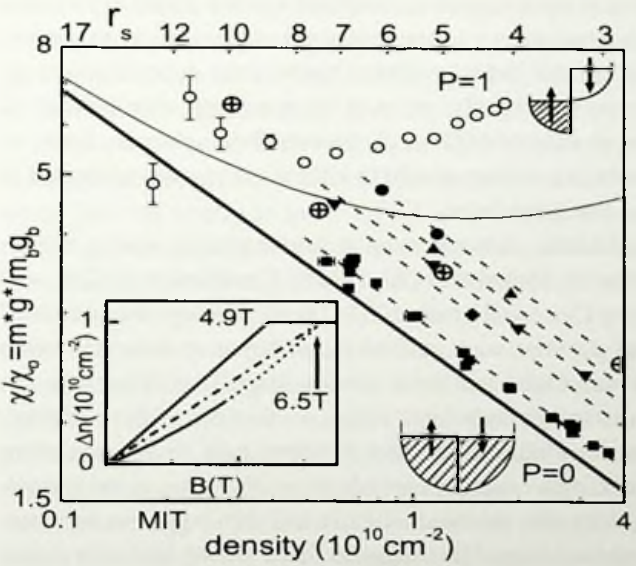


Figure 1.6: Density-dependence of m^*g^* in an ultra-clean 2D electron system in GaAs/AlGaAs determined by two different methods. The solid data points are obtained from tilted-field Shubnikov-de~Haas measurements with different configurations of Landau levels. The parallel dashed lines indicate a power law dependence of m^*g^* with a single exponent for all level configurations. The open circles show nonmonotonic density-dependence of m^*g^*derived from the full polarization field, B_c, using the parallel field method. The inset shows the net spin for $n_s = 1 \cdot 10^{10}$ with interpolated regime (solid line) and extrapolated regime (dotted line). B_c=4.9T from the in-plane field method and B_{ext}=6.5~T from extrapolation of the tilted-field method. The thick solid line represents extrapolation of m^*g^* to the P=0 limit.(From Kravchenko and Sarachik [9])

not due to spin exchange, in disagreement with the Fermi liquid model.

1.2 Theoretical Understanding

At present, there is no consensus about the nature of any of these effects. One feature that distinguishes the systems now under study from those examined in the past is that the interactions are enormous. As explained before, the dimensionless measure of the interaction strength, r_s, is of order 10 or higher. Thus we have an unambiguous example of the *strong-coupling many-body problem* for which theoretical methods are still poorly developed; it is still a frontier of theoretical condensed matter physics. The old problem of the interplay of disorder (Anderson localization) and electron-electron interaction (Mott localization) is presented here in an extreme limit. In spite of this situation, whether the electron-electron interaction plays the dominant role in this dynamics is controversial, as discussed below.

Various explanations have been suggested, ranging from non-Fermi liquid states and different kinds of superconductivity to single-particle physics based on temperature dependent scattering on charged traps and/or temperature dependent screening. Here we try to address, very briefly, some of the most important theoretical suggestions related to the physics of MIT in 2D electron/hole systems. Some of these theoretical models which are more related to kind of the physics addressed in this thesis will be discussed in depth later.

The possibility that a metallic state can exist at zero magnetic field in two dimensions was first suggested by Finkelstein [22, 23] and Castellani *et al.* [24] (and reconsidered again later by Castellani *et al.* [25]). In this theory, the combined effects of interactions and disorder were studied by perturbative renormalization group (RG) methods. It was found that for a weakly-disordered 2D system, an interaction parameter scales to infinitely large values — thus out of the perturbative regime — before zero temperature is reached. Unfortunately the RG procedure fails as soon as this dimensionless coupling exceeds unity. However, as the temperature is lowered, the resistivity first increases slightly and then begins to decrease, just as the coupling becomes too large. This suggests that a low-temperature metallic state might be established. The theory does not contain a metal-insulator transition (in the absence of a magnetic field), nor is the nature of the possible metallic state revealed. However, an external magnetic field, via Zeeman splitting, will drive the system back to the insulating state, in agreement with the experiments. This scenario has not received general acceptance because the divergences which occur at non-zero temperature cause the theory to become uncontrolled. One should also realize that the approach is perturbative and based on a Fermi liquid starting point. In the present context, as discussed earlier, r_s is so large that the theory's detailed applicability is in question.

Nevertheless, we may consider what happens within the RG scenario as the temperature is reduced further. The RG flow not only leads to a divergent interaction coupling but also to a divergent spin susceptibility. This has been interpreted as signaling either the development of local moments [23] or of ferromagnetism [26]. The latter is expected to occur at sufficiently large r_s in the 2D interacting electron system [28]. The onset of such time reversal breaking effects entails a crossover to a situation (a different universality class) in which the previously diverging coupling remains finite, even decreases, and eventually, at low temperature, an insulating state is once again obtained. Thus, carried to its conclusion, the renormalization group description appears to indicate an intermediate region of metallic-like behavior but so far fails to produce a metallic state at zero temperature [27]. However, the details of the various possible behaviors have not yet been worked out.

Pursuing the theme that the metallic state in these systems is very unlikely to be Fermi Liquid (since if the interaction would turned off the metallic behavior would disappear and the system become an insulator) , the effect of disorder on a model of a 2D non-Fermi liquid was discussed by Chakravarty *et al* [29]. While the origin of the non-Fermi liquid was not explained, they showed that for sufficiently

strong interactions (which would occur at low density), a non-Fermi liquid state of interacting electrons is stable in the presence of disorder and is a perfect conductor (as conjectured earlier by Dobrosavljević *et al* [30]. Otherwise, the disorder leads to localization as in the case of non-interacting electrons.

A number of authors have suggested a percolation-type description of the metal-insulator transition. He and Xie [31] proposed such a transition in the two dimensional electron system in Si MOSFETs at low electron densities, with percolation occurring between a conducting liquid phase separated by regions of an insulating, low density, *vapor* phase. A percolation transition involving non-interacting electrons was developed by Meir [32]. He considered the system to be inhomogeneous, consisting of electron (or hole) puddles connected by quantum point contacts. The model is capable of explaining the metallic $\rho(T)$ for n_s close to n_c, but it predicts the drop of resistance to be no more than a factor two in a degenerate system, while experimentally it is more than ten in Si MOSFETs. Meir remarked that in a non-degenerate system, the drop might be arbitrarily large. However, in the experiment, the dramatic drop of the resistance occurs only when the electron system is degenerate [33]. The drop is also substantial at larger n_s where the inhomogeneities, if any, are weak. In fact, all the percolation scenarios require the metallic phase to be inhomogeneous on some scale. It would be of interest to examine this issue experimentally. Recent results of Ilani *et al.* [14] seem to indicate an inhomogeneous insulating phase, but a homogeneous metallic one.

A very interesting suggestion, which is central to the study of the MIT presented in this thesis, came recently from Si and Varma [34]. Building on some of the theoretical work that has just been discussed, they developed a theory which leads to a metal-insulator transition without, however, giving a clear description of the metallic phase. A feature of the earlier RG approach is that the compressibility (proportional to the inverse screening length $1/s$) is unrenormalized. This is valid at high enough density so that the screening is good and s is less than the mean free path ℓ. Si and Varma pointed out that at the low electron densities (large r_s) of the experiments, one might expect $s > \ell$ and thus a renormalization of the compressibility. As already mentioned, no perturbative singularity is found in the compressibility due to interactions. However, the correlation energy contribution of the zero-point fluctuations of plasmons is altered due to disorder with a magnitude which depends also on r_s. The leading order contribution in powers of $(k_F\ell)^{-1}$ can be calculated for arbitrary r_s. Including this contribution the compressibility κ may be written in the form[35]

$$\frac{\kappa_0}{\kappa} = \frac{\kappa_0}{\kappa_{pure}} + 0.11 r_s^3/(\omega_0\tau) + O\left((r_s^4)/(\omega_0\tau)^2\right). \tag{1.2}$$

Here κ_{pure} is the compressibility for zero disorder, $\kappa_0 = N(0)$, and ω_0 is the Rydberg. This has an important bearing on the Metal-Insulator transition because the screening length s is given by

$$s/s_0 = \kappa_0/\kappa, \tag{1.3}$$

where $s_0 = a_0/2$.

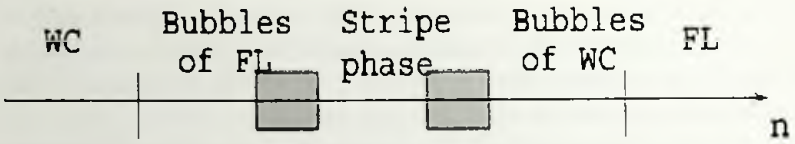


Figure 1.7: Phase diagram of the 2D electron system at $T = 0$ suggested by Spivak [38]. The symbols WC and FL correspond to the Wigner crystal and the Fermi liquid phases, respectively. The shaded regions correspond to phases which are more complicated than the bubble and the stripe phases. (from Spivak [38])

The above line of reasoning is of particular interest because as discussed earlier a sharp variation in the compressibility is indeed observed to accompany the transition from the metallic-like to insulating-like state as density is decreased (as shown in Fig. 1.4). In such systems for large r_s, the interaction becomes unscreened at low temperature and the compressibility approaches zero as one nears a metal-insulator transition. This is what one expects if the insulating state is one in which the long-range Coulomb interaction becomes unscreened, as would also be the case in a disordered Wigner solid [36].

Spivak [38] predicted the existence of an intermediate phase between the Fermi liquid and the Wigner crystal with a first order transition in a clean electron system. The suggested phase diagram is shown in Fig.1.7. In analogy with He^3, where m^* increases approaching the crystallization point, Spivak [37] suggested in an earlier paper that the renormalization of m^* is dominant compared to that of the g-factor as the transition is approached, and that m^* should increase with magnetic field. Although the increase of the mass is in agreement with the experimental results of Shashkin *et al.* [39, 40, 41], the suggested increase of m^* with the degree of spin polarization is not confirmed by their data.

1.3 The Plan of The Thesis

Recent progress in semiconductor technology has enabled the fabrication of high quality 2D samples with very low randomness in which measurements can be made at very low carrier densities[4]. The strongly interacting regime ($r_s \gg 1$) has thus become experimentally accessible. Experiments on the low-disordered 2D electron(hole) system in Si-MOSFET and p-GaAs/AlGaAs samples demonstrated that there are surprising and dramatic differences between the behavior of strongly interacting systems at $r_s > 10$ as compared with weakly-interacting systems where $r_s \ll 1$: with increasing electron density, one can cross from the regime with insulating behavior to a regime with metallic properties. The *metal-insulator Transition* (MIT) in 2D electron (hole) system is observed in both transport and ther-

[4]It is very interesting to note that even in the earlier time physisict were able to produce hight quality samples.

modynamic experiments. In chapter 1 we try to give a very brief introduction and address the current state of the most important experimental results as well as the related school of thoughts in this very interesting subject.

Clearly the experimental finding on the MIT is in contradiction with the scaling theory of localization which predicts insulating phase in 2D disordered system for all densities. Although the scaling theory does not explicitly consider the effect of the Coulomb interaction between electrons, early theoretical works (in 1980's) predicted that weak electron-electron interactions ($r_s \ll 1$) increase the localization even further. To date, No analytical theory has been developed in the limit of strong interactions ($r_s \gg 1$). The old problem of the interplay of disorder (Anderson localization) and electron-electron interaction (Mott localization) is presented here in an extreme limit. Various explanations have been suggested, ranging from non-Fermi liquid states and different kinds of superconductivity to single-particle physics based on temperature-dependent scattering on charged traps and/or temperature-dependent screening. Chapter 2 briefly reviews some of the most importants of these theoretical models. We will put more emphasize on the models which are more in the same line of the research presented in this thesis.

There also have been suggestions that the observed (continuous) metal-insulator transition in 2D electron system at low temperatures, might be a reminiscence of the more robust zero temperature Quantum Phase Transition (QPT), driven by interaction and disorder in the system. These type of transitions, which are accessed at zero temperature by variation of a non-thermal control parameter, can influence the behavior of electronic systems over a wide range temperatures. Generally speaking, quantum phase transitions occur as a result of competing ground state phases. The cuprate superconductors which can be tuned from a Mott insulating to a d-wave superconducting phase by carrier doping are a paradigmatic example. At the generic quantum critical point, the ratio of the thermal expansion and the specific heat divergs. Due to the importance of such a phenomena (QPT) in relation to the MIT, in chapter 3, we change a gear and will address the universality of QPT in cuprates. Using elementary power counting arguments we have discovered an empirical strategy which should make it possible to decide if the 'quantum criticality' of the cuprates has to do with universality. We will use these arguments later in the last chapter in connection with the MIT transition in 2D systems.

In chapter 4 we will focus on the recent and at the same time very interestinf experimental finding which probe the change in the thermodynamic behavior of 2D electron(hole) system across MIT: Compressibility measurement. These experiments are important, at least, for two reasons: first, they are quite free of any complications related to the quantum coherency which one may encounter in transport measurements and second, if a zero temperature phase transition exist, it may reveal itself also in the thermodynamics like the electronic compressibility, κ . These experiments, clearly show that electronic compressibility, exhibit an unusual change across the MIT point in 2D electron/hole systems. In fact, the very fundamental reason for such behavior and its relationship with the local properties of electron(hole) system is the central focus of this thesis.

Over the years, it has been suggested that the interacting 2D electrons in the presence of disorder can be understood by the droplet, or puddle, model. The electrons are separated into "liquid" droplets with local density higher than the average density and "vapor" islands with lower density. For a given disorder, due to the potential fluctuation of the donors, as the Fermi energy is decreased, the "vapor" region increases. In this model, a crossover from metallic behavior to insulating behavior should be seen around the percolation threshold. Using an artificial way of making such droplets, there has been an attempt (within DFT-LDA) to explain the observed anomaly in compressibility of 2D electrons. In chapter 5 we will make the case that for the highly mobile (weakly disordered) and strong interacting electrons(hole) this kind of physics fails to survive. This will make the stage ready for the chapter 6.

At low enough density, the screening length of a disordered interacting electron gas may become larger than its mean free path. The compressibility, being proportional to the inverse screening length, is expected to vanish as the transition to the insulating state is approached. This, indeed, is the most important physical concept behind the line of a research carried out in this thesis, suggested by Si and Varma. In chapter 6, we will develop a DFT formulation of this theory as a first order correction to LDA: A density functional theory formalism, applicable to the metallic side of the MIT transition, incorporating the physics of highly unscreened (interacting) and highly mobile (weakly disordered) electrons.

Chapter 7 and 8 are basically the out comes of the Si-Varma theory focusing in both bulk (chemical potential and compressibility) and local (density inhomogeneity) properties. By employing the numerical simulation, we will show that, although the "critical" density where the compressibility changes its behavior implied by these calculations is smaller than the experimentally observed critical density, but such a DFT formulation for 2D electrons indicate a substantial improvement of the description of the behavior of compressibility near MIT transition point. At the same time, we will see, due to the fact that the screening length becomes large, it forces the electron liquid to become increasingly incompressible with the obvious effect that density fluctuations are suppressed.

There have been recent experimental findings that show there is a sharp enhancement of the spin susceptibility as the metal-insulator transition is approached; indications exist that in silicon MOSFETs, the spin susceptibility may actually diverge at some sample-independent electron density $n_\chi \approx 8 \cdot 10^{10}$. Based on the Fermi-liquid theory, this observation might also be related to the enhancement of the effective mass of the charge carriers at very low densities. Finally, in chapter 9, we will discuss the possibility of the existence of a QPT in disordered 2D electron(hole) system. By using the zero temperature scaling laws, already derived in chapter 3, we will address the possible relationship between the effective mass divergence and the screening properties of the 2D electron system at low densities and low disorders.

Chapter 2

Theoretical Background

In two-dimensional electron systems, the electrons are confined to move in a plane in the presence of a random potential. According to the scaling theory of localization [2], these systems lie on the boundary between high and low dimensions insofar as the metal-insulator transition is concerned. The carriers are always strongly localized in one dimension, while in three dimensions, the electronic states can be either localized or extended. In the case of two dimensions the electrons may conduct well at room temperature, but a weak logarithmic increase of the resistance is expected as the temperature is reduced. This is due to the fact that, when scattered from impurities back to their starting point, electron waves interfere constructively with their time reversed paths. While this effect is weak at high temperatures due to inelastic scattering events, quantum interference becomes increasingly important as the temperature is reduced and leads to localization of the electrons, albeit on a large length scale; this is generally referred to as "weak localization". Indeed, thin metallic films and two-dimensional electron systems fabricated on semiconductor surfaces were found to display the predicted logarithmic increase of resistivity [5], providing support for the weak localization theory.

The scaling theory does not explicitly consider the effect of the Coulomb interaction between electrons. The strength of the interactions is usually characterized by the dimensionless Wigner-Seitz radius, $r_s = 1/(\pi n)^{1/2} a_B$ (here n is the electron density and a_B is the Bohr radius in a semiconductor). In the experiments mentioned, the Coulomb interactions are relatively weak. Indeed, these experiments are in agreement with theoretical predictions [3] that weak electron-electron interactions ($r_s \ll 1$) increase the localization even further. As the density of electrons is reduced, however, the Wigner-Seitz radius grows and the interactions provide the dominant energy of the system. No analytical theory has been developed to date in the strongly-interacting limit ($r_s \gg 1$). Finkelstein [22, 23] and Castellani *et al.* [24] predicted that for weak disorder and sufficiently strong interactions, a 2D system should scale toward a *conducting* state as the temperature is lowered. However, the scaling procedure leads to an increase in the effective strength of the interactions and to a divergent spin susceptibility, so that the perturbative approach

breaks down as the temperature is reduced toward zero. Therefore, the possibility of a 2D metallic ground state stabilized by strong electron-electron interactions was not seriously considered.

The experimental findings described in previous chapter were quite unexpected. Once accepted, they elicited strong and widespread interest among theorists, with proposed explanations that included unusual superconductivity [42], charging/discharging of contaminations in the oxide [43], the formation of a disordered Wigner solid [36], inter-subband scattering [45] and many more (for a review, see [8]). It is now well-documented that the metallic behavior in zero magnetic field is caused by the delocalizing effect associated with strong electron-electron interactions over ruling quantum localization. In the "ballistic regime" deep in the metallic state, the conductivity is linear with temperature [46]. Closer to the transition, in the "diffusive" regime, the temperature dependence of the resistance is well described by a renormalization group analysis of the interplay of strong interactions and disorder [47]. Within both theories (which consider essentially two limits of the same physical process) an external magnetic field quenches the delocalizing effect of interactions by aligning the spins, and causes a giant positive magnetoresistance. However, the metal-insulator transition itself, as well as the dramatic increase of the spin susceptibility, inverse compressibility, and effective mass in its close vicinity, still lack adequate theoretical description; in this region the system appears to behave completely different from weakly interacting Fermi liquid.

2.1 2D Electron System

The system we are interested in is an electron gas in two dimension with a uniform positive background with no complications arising from the lattice structures, the so called Jellium model. This situation is indeed realized experimentally in MOSFETs (and heterostructures) in which an insulator is typically sandwiched between a metallic plate and a semiconductor. By applying an electric field a two-dimensional charge layer accumulates on the surface of the semiconductor adjacent to the insulator, whose density can simply be changed by varying the field strength. Since the thickness of the insulating layer is typically more than 100 nm, for the phenomena in which the large average inter-electron distance is of order of 10-100 nm the positive (capacitive) charge on the insulator provides a uniform background to a first approximation. In Si samples, surface roughness is the principle source of disorder at high densities, while at low densities (in the regime where the transition take place) ionized impurity scattering takes over due to the fact that there is much less screening. In GaAs, remote impurity scattering dominates which is mostly small-angle. This is the main reason for the mobility in these samples to be large. Peak electron mobilities in these samples exceed those in the samples used in previous studies by an order of magnitude, reaching more than $4 \times 10^4\ cm^2/Vs$ at $T = 4.2\ K$. The very high quality of the samples allow studies of the 2D system

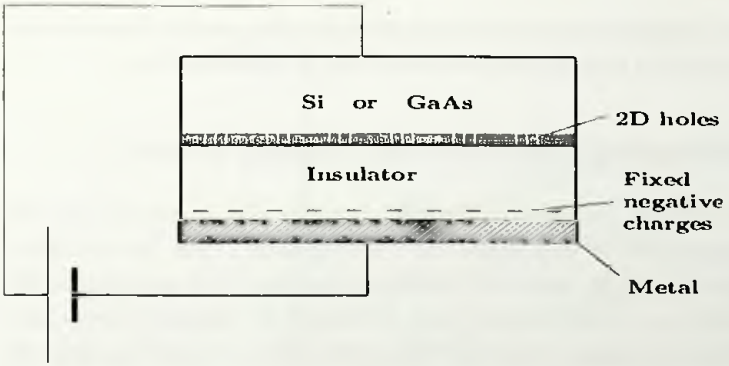


Figure 2.1: Sketch of a MOSFET. Holes (or electrons) are trapped at the interface of the semi-conductor and the insulator due to the band gap difference between them, the dipole layer and the applied electric field. Two-dimensional electrons (holes) may also be found by layered structures (heterostructures) of semi-conductors with different band gaps such as *GaAs* and *AlGaAs*.

in a very dilute regime, i.e., at electron densities below $10^{11}\ cm^{-2}$.

Instead of being small compared to the Fermi energy, the electron-electron interaction energy, E_{e-e}, is the dominant parameter at these low densities. Estimates for Si MOSFETs at $n_s = 10^{11}\ cm^{-2}$ yield

$$E_{e-e} \sim \frac{e^2}{\varepsilon}(\pi n)^{1/2} \approx 10meV \tag{2.1}$$

while

$$E_F = \frac{\pi\hbar n}{2m^*} \approx 0.58meV \tag{2.2}$$

where e is the electron charge, ε is the dielectric constant, E_F is the Fermi energy, and m^* is the effective electron mass.

Neglecting disorder, the problem is characterized by r_s, defined as the ratio of the potential energy to the kinetic energy:

$$r_s \equiv \frac{E_{e-e}}{E_F} = \frac{m^* e^2}{4\pi\varepsilon\hbar^2\sqrt{\pi n}}. \tag{2.3}$$

The dimensionless quantity r_s can also be expressed as the radius of the circle whose area is equal to the area per conduction electron, measured in units of Bohr radius a_B. Thus according to Eq.(2.1) and Eq.(2.2), r_s assumes values above 10 in these samples. In the dense electron limit ($r_s \ll 1$), the kinetic energy dominates and metallic behavior is expected. In the very dilute regime ($r_s \gg 1$), the Coulomb energy dominates and 2D electrons are expected to form a Wigner crystal if the disorder is weak; a numerical simulation [4] predicted that this should occur at $r_s \approx 37 \pm 5$. Subsequent work by Chui and Tanatar [48] showed that solidification

is occurring at even higher density when disorder is present, since the introduction of disorder is expected to turn the first order MIT into a continuous one.

2.2 Non-interacting Electrons and Scaling Theory

Traditionally the study of electronic conduction in solids was concerned with the case that the concentration of impurities was small. In such a case the correlation between the impurities can be neglected and the conductivity is proportional to the concentration. Moreover, most theories were developed for infinite systems even though it was known for many years that the conductance of actual systems depends on their size. Attempts to go beyond these idealizing conditions have been limited and did not bring about any conceptual change in general views of conductance. In 1978, Mott [49] discussed the existence of a *mobility edge* in the metal-insulator transition in disordered systems. He argued that when the concentration of impurities increases, the localized electron waves around the impurities start overlapping with each other until a continuous impurity band is formed. However, the bottom of this band is still localized and there must be a sharp boundary between the localized and extended states. The mobility is finite in this case if the electron energy is above the mobility edge ε_c and is zero below. Hence, there can be a minimum conductivity above which the system is conductive.

Subsequently, in 1979, Abrahams *et al.* [2] reported a scaling theory which opened the door to the basic understanding of conduction in disordered systems. According to this theory the conductance of disordered systems varies continuously to zero without a minimum. The scaling theory triggered substantial theoretical and experimental avdances in this hitherto neglected area and Anderson localization became then central subject in the study of conductance in disordered systems.

The basic idea behind the scaling theory is simple and beautiful. Consider a system with size L in each spatial direction with dimension d. In scaling theory the conductance is considered as a function of L and for large enough system the conductance, $g_L = g(L)$, should be proportional to the cross section L^{d-1} and inversely proportional to length L. Therefore, the conductance can be expressed as

$$g(L) = \sigma L^{d-2} \tag{2.4}$$

where σ is the conductivity which is independent of L. The immediate result of this formulation is, for $d = 2$, $g(L) = \sigma$, and it appears as though there is no L dependence. Now consider another system with size L'; the conductance $g(L')$ is assumed to depend on L' in accordance with

$$g(L') = f(g(L), L'/L) \tag{2.5}$$

where f is a universal function which is independent of microscopic structure of the system. This is a *scaling hypothesis.* Its consequence can be seen by taking a logarithmic differential of this equation with respect to variable $x = L'/L$. The derivative is

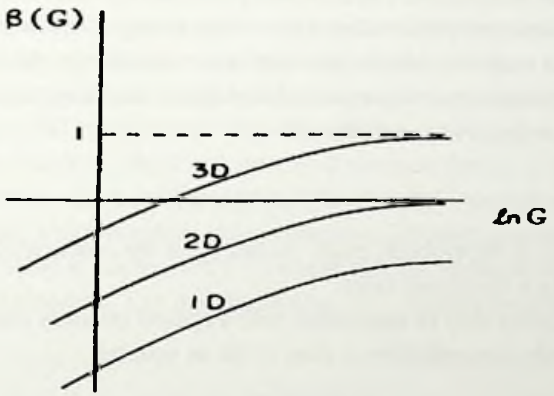


Figure 2.2: The scaling function for non-interacting electrons with disorder deduced by Abrahams *et al.* [2].

$$\frac{d\ln g}{d\ln L} = \beta(g) \tag{2.6}$$

where $\beta(g)$ is a universal function. In the metallic regime Eq.2.4 suggests

$$\beta(g) = d - 2 \tag{2.7}$$

In the opposite limit, the system is expected not to conduct. In this case the conductance may decrease exponentially with L. If

$$g_L \sim \exp(-\alpha L) \tag{2.8}$$

as $L \to \infty$, we expect

$$\beta(g) = \ln g \tag{2.9}$$

Let us assume that $\beta(g)$ is a continuous function which connects the two limits given in Eq.2.7 and Eq.2.8. Such a function is schematically shown in Fig. 2.2, where $\beta(g)$ is plotted against $\ln g$. This type of variation of $\beta(g)$ agrees with numerical results [50]. In three dimensions $\beta(g)$ is negative up to a certain value of g and beyond this value it increases toward the asymptotic value 1. The point at which $\beta(g) = 0$ is called fixed point, in the sense that $d\ln g/dL$ vanishes, i.e. there is no L dependence. However, the location of this point may depend on impurity or electron energy.

By assuming a linear variation of $\beta(g)$ in vicinity of fixed point one easily finds that for sufficiently long L, the conductivity is

$$\sigma = \sigma_c(g - g_c)^{\nu} \tag{2.10}$$

where g_c is the conductance at the fixed point and σ_c is a certain constant. ν is a positive parameter proportional to $1/g_c$. This is important as it shows σ vanishes as $g \to g_c$. We may translate the above relation into energy space, because near the fixed point the conductance is expected to depend on electron energy. By assuming that $g(L) \sim \varepsilon$ and $g_c \sim \varepsilon_c$, we arrive at

$$\sigma = \sigma_c(\varepsilon - \varepsilon_c)^{\nu} \tag{2.11}$$

The constant ε_c is the mobility edge. At this point the conductivity vanishes instead of approaching a minimum value.

The fixed point may be associated with a critical impurity concentration n_i^c. For a given impurity concentration n_i near n_i^c let us assume

$$g(L) - g_c \propto (n_i - n_i^c) \tag{2.12}$$

even though such a proportionality is not guaranteed. We obtain

$$\sigma = \sigma_c(n_i - n_i^c)^{\nu}. \tag{2.13}$$

Thus, the conductivity vanishes when the impurity concentration approaches a certain value n_i^c.

It is of course important conceptually to recognize the appearance of a point in three dimensions where the conductivity vanishes. In contrast no such a point appears in one and two dimension since $\beta(g)$ stays below the $\ln g$ axis as shown in Fig.2.2. Hence, according to the scaling theory the electrons are localized at absolute zero in one and two dimensions regardless of the amount of disorder. One dimension is somewhat special in the sense that a single impurity site is suffices to block the single conducting channel. Therefore, let us find how the conductance of a 2D system varies with its length L. For large g, Fig.2.2 suggests that $\beta(g)$ approaches zero. Let us assume that this (asymptotic) behavior is simply proportional to $1/g$

$$\beta(g) = -\frac{g_a}{g}. \tag{2.14}$$

Such an asymptotic form has actually been justified to be correct to order $(1/g^3)$ based on a nonlinear-σ model [51, 52].

The dimension of conductivity is given in 2D by e^2/h. Hence, the constant g_a may be expressed by

$$g_a = \frac{e^2\lambda}{2\pi^2\hbar}, \tag{2.15}$$

where λ is a dimensionless constant. In combination with Eq.2.6, Eq.2.14 yields

$$g(L) = g(L_0) - g_a \log(L/L_0), \tag{2.16}$$

where $g(L_0)$ and L_0 are (unknown) constants. Therefore within this approximation and for large g a logarithmic decrease of conductance with L takes place.

The above logarithmic variation has been derived for absolute zero. One can also show that the temperature variation of conductance is logarithmic as well. According to the general conductivity formula, which is expressed by the current-current correlation function, phase coherence of electron waves is crucial to the conductivity. Electrons are scattered elastically and diffuse around until inelastic scattering takes place which destroys phase coherence. Hence, an elastic mean free path ℓ_e is expected to be given by $(D\tau_i)^{1/2}$, where τ_i is the inelastic scattering time and D is a diffusion constant. The inelastic scattering time may be assumed to vary as

$$\tau_i \sim T^{-p}, \tag{2.17}$$

because $\tau_i \to \infty$ at absolute zero, where p is of order 1. Hence, due to localization, a finite temperature correction, $\Delta\sigma(T)$, to the Drude-type conductivity, is expected in the form

$$\Delta\sigma(T) = \frac{e^2}{2\pi^2\hbar} p \log T/T_0. \tag{2.18}$$

This means that below a certain temperature T_0, the conductance in two dimensions decreases with temperature. The prefactor

$$\frac{e^2}{2\pi^2\hbar} = 1.2 \times 10^{-5}\Omega^{-1} \tag{2.19}$$

is the natural conductivity unit. Note that Eq.2.18 is a correction. The logarithmic decrease can be considered as a precursor to a complete localization at absolute zero, and represents what is called *weak localization*.

Clearly the above hypothesis is associated with systems of non-interacting electrons. One might expect turning on the interaction would gives you quite different scenario from the above. However, early work by Altshuler *et al.* [3] showed that for weak interactions these behavior essentially remains unchanged, that is, the electrons are expected to be localized even further. Clearly this is not the case for the experiments already summarized in chapter 1, as these systems are the true examples of the strongly interacting electron physics.

2.3 Interactions in Disordered Electron Systems

In the study of electronic properties of metals the success of Fermi liquid theory manifested e.g. by the empirical success of the Drude relation for conductivity, has long given the impression that interaction effects are not important. How can we in general understand that Coulomb effects are strong in the regimes of weak localization? This question may be answered by recalling that weak localization is related to diffusive motion of electrons. Thus the average time they spend at an impurity site can be longer than the case of plane-wave electrons. This causes enhancement

of their (time averaged) interactions. Fermi-liquid theory for interacting electrons survives in three dimensions in the presence of a dilute concentration of impurities. Some noteworthy differences from the clean case are [35]:

1. Because of the lack of momentum conservation, the concept of a Fermi-surface in momentum space is lost but it is preserved in energy space, i.e., a discontinuity in particle occupation as a function of energy occurs at the chemical potential. Momentum of particles may be defined after impurity averaging. General techniques for calculating impurity-averaged quantities are well developed; see for example [53, 54].

2. In the presence of impurities, the density-density correlation (and spin-density correlation, if spin is conserved) at low frequencies and small momentum must have a diffusive form (as required by particle-number conservation and the continuity equation)

$$\pi(q,\omega) = \kappa \frac{Dq^2}{i\omega + Dq^2}, q \ll \ell^{-1} \text{and} \omega \ll \tau^{-1}. \tag{2.20}$$

   Here $\kappa = dn/d\mu$ is the compressibility and D is the diffusion constant. For non-interacting electrons $D = \frac{1}{3} v_F^2 \tau$, where τ and v_F are the scattering time and Fermi velocity, respectively. Interactions renormalize D and κ [54].

3. Because of statement (1), the *impurity-averaged* single-particle spectral function at a fixed k is spread out over an energy Γ, so that for frequencies within a range Γ of the chemical potential, it has both a hole-part (for $\omega < \mu$) and a particle part (for $\omega > \mu$). This is an important technical point in microscopic calculations.

4. Actually even the first order interaction dressed by diffusion fluctuations is singular. For the small q (for singular properties), one need to consider interactions only in the s-wave channel. One then has two interaction parameters, one in the singlet channel and the other in the triplet channel (See Fig.2.3). Considering Coulomb interactions, the effective interaction in the singlet channel sums the polarization bubbles connected by Coulomb interactions. Using Eq. 2.20 for the polarization bubble, it can be shown [55] that for small momentum transfer the interaction in the spin singlet ($S = 0$) channel, Fig.2.3, becomes

$$V_{\text{singlet}} = 2\kappa. \tag{2.21}$$

   In the non-interacting limit $\kappa = N(0)$, independent of density. Consider now the ladder-type interactions illustrated in Fig.2.4. These involve both the singlet and the triplet interactions. The momentum carried by the interaction lines is however to be integrated over. Therefore the triplet-interactions do not have a universal behavior, unlike the singlet interactions.

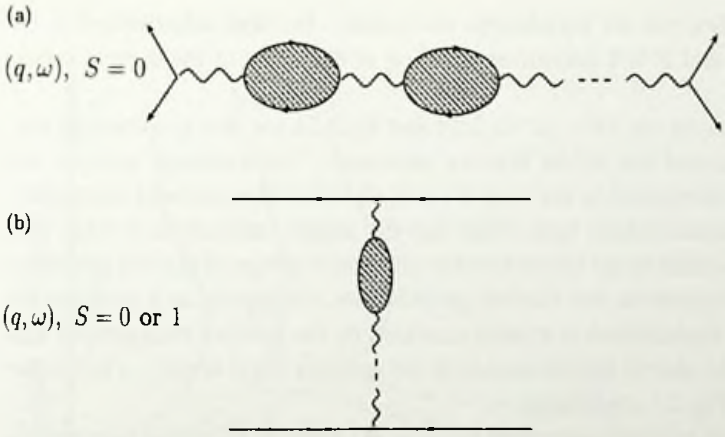


Figure 2.3: Effective interactions can be split up into singlet and triplet channels. In the singlet-only channel (a), the density-density interaction is screened by the Coulomb interaction and is universal at long wavelengths. In the triplet channel and in the singlet channel for large momentum, the screened density-density interaction appears only in the cross channel and is therefore non-universal.

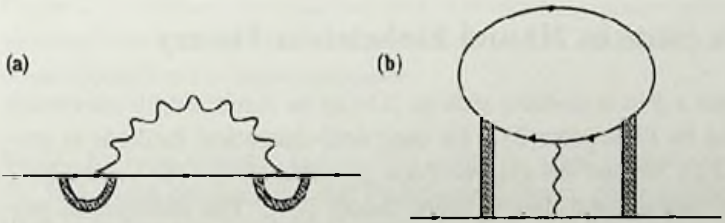


Figure 2.4: Simplest processes contributing to the singular self-energy. (a) Exchange process. (b) Hartree process.

In fact more than two decades ago, Altshuler, Aronov and collaborators [3, 135, 136] calculated the logarithmic corrections to first-order in the interactions for various physical quantities, within the Hartree-Fock approximation. To these one can add the contribution already discussed due to weak-localization. The corrections to the single-particle density of states, the specific heat and the conductivity over the non-interacting values are respectively:

$$\frac{\delta N}{N} = \frac{1}{4\pi\varepsilon_F\tau}\ln|\omega\tau|\left|\ln\left|\frac{\omega}{\tau(Ds^{-2})^2}\right|\right|, \tag{2.22}$$

$$\frac{\delta C}{C} = \frac{1}{\pi\varepsilon_F\tau}\left(1-\frac{3}{2}F\right)\ln|T\tau|, \tag{2.23}$$

$$\frac{\delta\sigma}{\sigma} = \frac{1}{4\pi^2}\left(2-\frac{3}{2}F\right)\ln|T\tau|. \tag{2.24}$$

The compressibility has no logarithmic correction. In these equations, s is the screening length and F is a parameter which is of the order of the dimensionless interaction r_s.

The first terms on the r.h.s. in Eq.2.23 and Eq.2.24 are due to exchange processes and the second due to the Hartree processes. The exchange process originates from the interaction in the singlet channel; hence the universal coefficient. The second contribution uses both triplet and the singlet interactions at large momentums. In first-order of the interaction the difference in sign of the two processes is natural. In pure systems, the Hartree process does not appear, as it involves the $q = 0$ interaction alone which is exactly canceled by the positive background. For disordered systems, due to the fluctuation in the (ground state) density, a first order Hartree process, Fig.2.4 contributes.

In the presence of a magnetic field the $S_z = \pm 1$ parts of the triplet interactions acquire a low-energy cut-off. Therefore the logarithmic correction to the resistivity is suppressed leading to negative magneto-resistance proportion to $F(H/kT)^2$ for small H/kT but $g\mu_B H >> \tau_{so}^{-1}, \tau_s^{-1}$ where τ_{so}^{-1} and τ_s^{-1} are spin-orbit and spin-scattering rates respectively, for appropriate impurities.

2.4 Metallic State in 2D and Finkelstein Theory

The possibility that a $B = 0$ metallic state in 2D can be stabilized by interaction was first suggested by Finkelstein[23]. He used field-theoretical methods to generalize Eqs.2.22-2.24 beyond the Hartree-Fock approximation. His results have been rederived in conventional diagrammatic theory [24]. The interference processes leading to weak localization are neglected. The theory may be regarded as first order in $1/k_F\ell$. In effect, the method consists in replacing the parameter F (see Eqs.2.22-2.24) by a scattering amplitude γ_t for which scaling equations are derived. This analogue of the F_o^s parameter of Fermi-liquid theory is fixed by imposing that the compressibility remains unrenormalized, i.e., does not acquire logarithmic corrections. A second important quantity is a scaling variable z, which is analogous to the dynamical scaling exponent z which gives the relative scaling of temperature (or frequency) with respect to the length scale. A very unusual feature of the theory is that z itself scales! Scaling equations are derived for γ_t and$z$ to leading order $1/k_F\ell$. As $T \to 0$, both γ_t and z diverge. The divergence in z usually means that the the momentum dependence of the fluctuations is unimportant compared to their frequency dependence [35]. The divergence in γ_t as $T \to 0$ in such a case has been interpreted to imply divergent spatially localized magnetic-fluctuations; In other words, it implies the formation of local moments [56, 23]. At the same time, the scaling equations show conductivity flows to a finite value.

The scaling trajectories of Finkelstein's theory are shown schematically in Fig. 2.5. While in the non-interacting theory with disorder, one always has an insulator, the system flows always towards a metal. However the theory cannot be trusted beyond $\gamma_t \sim 1$ beyond which it becomes uncontrolled. The theory also cannot be

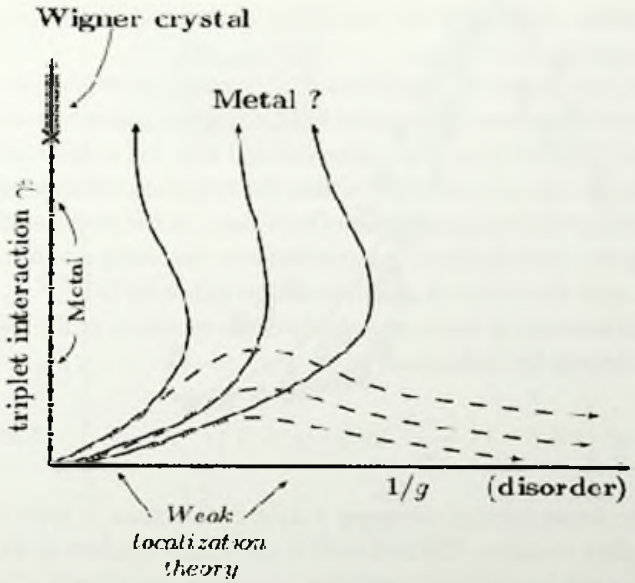


Figure 2.5: Schematic renormalization group flow for the disordered interacting electron problem according to the Finkelstein theory. dashed lines represent the effect on the solid lines on applying a magnetic- field which couples to spins alone.

trusted for large disorder, $k_F\ell \sim O(1)$, even for small interactions.

It is worth emphasizing that Finkelstein's theory gives an effect of the interactions in a direction opposite to the leading perturbative results. The perturbative results themselves of course are valid only for small r_s while Finkelstein theory is strictly valid only for $r_s < 1$. One possibility is that the Finkelstein result itself is a transient and the correct theory scales back towards an insulator. Another possibility is that it correctly indicates (at least for some range of r_s and disorder) a strong-coupling singular Fermi-liquid metallic fixed line.

Given the fact that in Fikelstein theory an interaction parameter scales to infinitely large values before zero temperature is reached (and the theory thus becomes uncontrolled), this scenario has not received general acceptance. It should also be noted that the theory may not be applicable to the current experiments (mentioned in Chapter 1) since it was developed for the diffusive regime: $T \ll \hbar/\tau$, where τ is the elastic mean-free time extracted from the Drude conductivity (Boltzmann constant is assumed to be equal to 1 throughout this section). This condition corresponds to the low-temperature limit $T \ll T_F\rho/(h/e^2)$. Since the Fermi temperature, T_F, is rather low at the small carrier densities considered here, the above condition is satisfied only close to the transition, where ρ becomes of the order of h/e^2. In the experiments, however, the characteristic decrease of the resistance with decreasing temperature often persists into the relatively high-temperature *ballistic* regime $T > \hbar/\tau$ (or $T > T_F\rho/(h/e^2)$). This observation has been interpreted [57, 58] as

evidence that the mechanism responsible for the strong temperature dependence cannot originate from quantum interference.

Based on the earlier calculations by Finkelstein [23], recently, Punnoose and Finkelstein [47] have tried to improve the previous RG calculations using two distinct (electron) valleys in Si-MOSFETs. They demonstrated that, the temperature dependence of the resistivity can be understood within the renormalization group theory describing the effect of the electron-electron interactions on the propagation of diffusive collective modes, with the *delocalizing* component becoming dominant in dilute 2D systems. A brief discussion of the method is as following [9]:

In 2D, the renormalization group equation describing the evolution of the resistivity, which has been derived by Finkelstein, is:

$$\frac{dg}{d\xi} = g^2\left[1+1-3\left(\frac{1+\gamma_2}{\gamma_2}\ln(1+\gamma_2)-1\right)\right]. \tag{2.25}$$

Here $\xi = -\ln(T\tau/\hbar)$, the dimensionless parameter g is the conductance in units of $e^2/\pi h$ and γ_2 is the coupling constant. The first term in the square brackets of Eq. 2.25 corresponds to the weak localization correction (quantum interference [2]), while the second term is the contribution of the singlet density mode which is due to the long range nature of the Coulomb interaction [3]. The last term describes the contribution of the three triplet modes. Note that the last two terms have opposite signs, favoring localization and delocalization, respectively. The resulting dependence $g(\xi)$ becomes delocalizing when $\gamma_2 > \gamma_2^* = 2.04$. This requires the presence of rather strong electron correlations.

In the case of two distinct valleys (as in (100) silicon MOSFETs), Eq. 2.25 can be easily generalized to:

$$\frac{dg}{d\xi} = g^2\left[2+1-15\left(\frac{1+\gamma_2}{\gamma_2}\ln(1+\gamma_2)-1\right)\right]. \tag{2.26}$$

The difference between the numerical factors in Eq. 2.25 and Eq. 2.26 results from the number of degrees of freedom in each case. The weak localization term becomes twice as large. The difference in the number of the multiplet modes increases the coefficient of the γ_2 term from 3 to 15. As a result of these modifications, the value of γ_2 required for the dependence $g(\xi)$ to become delocalizing is reduced to $\gamma_2^* = 0.45$. which makes it easier in the case of two valleys to reach the condition where the resistivity decreases with decreasing temperature.

In conventional conductors the initial values of γ_2 are small, and the net effect is in favor of localization. In dilute systems, however, this value is enhanced due to electron-electron correlations. In addition, γ_2 also experiences logarithmic corrections due to the disorder [22, 23, 17, 25]. The equation describing the renormalization group evolution of γ_2 is the same for the case of one and two valleys:

$$\frac{d\gamma_2}{d\xi} = g\frac{(1+\gamma_2)^2}{2}. \tag{2.27}$$

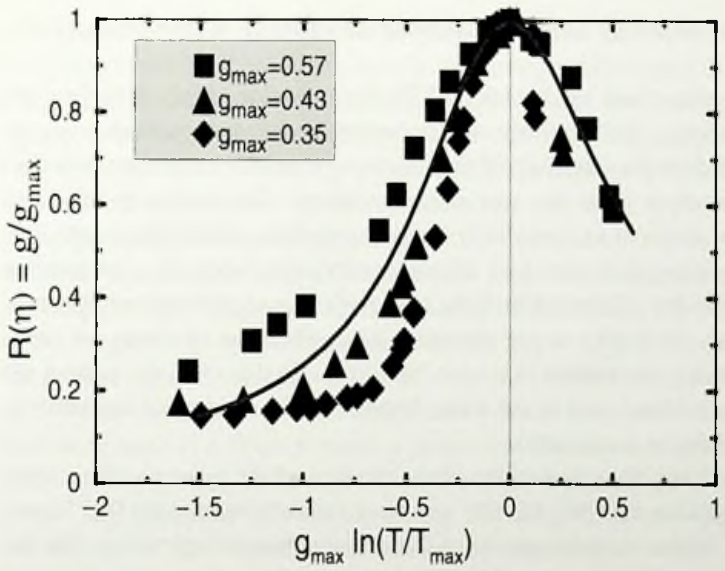


Figure 2.6: Data for silicon MOSFETs at $n_s = 0.83$, 0.88 and $0.94 \cdot 10^{11}$ from Pudalov *et al.* [59] are scaled according to Eq. 2.28. The solid line corresponds to the solution of the renormalization group, Eqs. 2.26 and 2.27; no adjustable parameters were used in this fit. (From Punnoose and Finkelstein [47]).

It follows from this equation that γ_2 increases monotonically as the temperature is decreased. Once it exceeds γ_2^*, the resistivity passes through a maximum. Even though the initial values of g and γ_2 are not universal, the flow of g can be described by a universal function $R(\eta)$ [22]:

$$g = g_{\max} R(\eta) \text{ and } \eta = g_{\max} \ln(T_{\max}/T), \tag{2.28}$$

where $T_{\max}$ is the temperature at which g reaches its maximum value $g_{\max}$.

In Fig. 2.6, the theoretical calculations are compared with the experimental data obtained by Pudalov *et al.* [59]. Since the renormalization group equations were derived in the lowest order in g and cannot be applied in the critical region where $g > 1$, only curves with maximum g ranging from $g_{\max} \approx 0.3$ to $g_{\max} \approx 0.6$ are shown in the figure. The decrease in the resistivity by up to a factor of five, as well as its saturation, are both captured in the correct temperature interval by this analysis without any adjustable parameters.

In agreement with experiment, this universal behavior can be observed only in ultra clean samples (with negligible inter-valley scattering) and will not be found in samples that are only moderately clean. In disordered silicon MOSFETs, a description in terms of an effective *single* valley is relevant, and the large value of $\gamma_2^* = 2.04$ makes it difficult for the non-monotonic $\rho(T)$ to be observed as the initial values of γ_2 are usually much less than 2.04. As a result, for g near the

critical region, the resistivity becomes insulating-like instead of going through the maximum.

Therefore, in ultra clean silicon MOSFETs, the behavior of the resistivity not far from the transition is quantitatively well described by the renormalization group analysis that considers the interplay of the electron-electron interactions and disorder when the electron band has two distinct valleys. The theory in this case remains in control down to exponentially low temperatures, unlike the single valley case, where γ_2 diverges at $\eta \approx 1$, or at temperatures just below T_{max}. A parallel magnetic field provides a Zeeman splitting and gives rise to positive magnetoresistance [60, 23, 24]. In a very strong magnetic field, when the electrons are completely spin-polarized, the system becomes "spinless". In this case, the system always scales to an insulator, and in the weak disorder limit, a universal logarithmic temperature dependence is expected.

For the ballistic region, calculations in the random phase approximation were carried out two decades ago [61, 62, 63] and were recently refined by Das Sarma and Hwang [64]. These theories predict a linear temperature dependence for the conductivity with metallic-like slope ($d\sigma/dT < 0$) regardless of the strength of the interactions. The spin degree of freedom is not important in this theory and enters only through the Fermi energy, which is a factor of two larger for the spin-polarized than for the unpolarized system. Therefore, the application of a (parallel) magnetic field does not eliminate the metallic temperature dependence.

The two limits of diffusive and ballistic regimes had until recently been assumed to be governed by different physical processes. However, Zala *et al.* [46] have shown that the temperature dependence of the conductivity in the ballistic regime originates from the same physical process as the Altshuler-Aronov-Lee corrections: coherent scattering of electrons by Friedel oscillations. In this regime the correction is linear in temperature, as is the case for the results mentioned in the previous paragraph. However, the value and even the *sign* of the slope depends on the strength of electron-electron interaction, the slope being directly related to the renormalization of the spin susceptibility [65, 69]. By aligning the spins, a magnetic field causes a positive magnetoresistance and changes the temperature dependence of the conductivity from metallic-like to insulating-like [69, 67, 68], in agreement with experiments.

2.5 Non-Fermi Liquid Approaches

The calculations discussed in the previous section [24, 22, 23, 46] use the Fermi liquid as a starting point. As discussed earlier, r_s becomes so large that the theory's applicability is in question near the transition [70] (for a review, see [35]. Moreover, the behavior of the effective mass reported by Shashkin *et al.* [40, 41] in the vicinity of the transition is unlikely to be consistent with a Fermi liquid model.

Very little theory has been developed for strongly interacting systems for which

r_s is large but below the expected Wigner crystallization. Several candidates have been suggested for the ground state of the strongly interacting 2D system, among them (i) a Wigner crystal characterized by spatial and spin ordering [71], (ii) an itinerant ferromagnet with spontaneous spin ordering [72], and (iii) a paramagnetic Fermi liquid [73]. According to numerical simulations [4], Wigner crystallization is expected in a very dilute regime, when r_s reaches approximately 35. This value has already been exceeded in the best p-type GaAs/AlGaAs heterostructures; however, no dramatic change in transport properties has occurred at the corresponding density. Recent detailed numerical simulations [74] have predicted that in the range of the interaction parameter $25 < r_s < 35$ prior to the crystallization, the ground state of the system becomes an itinerant ferromagnet. Chakravarty *et al.* [36] have proposed a more complicated phase diagram where the low-density insulating state is a Wigner glass, a phase that has quasi-long-range translational order and competing ferromagnetic and antiferromagnetic spin-exchange interactions. The transition between insulating and metallic states within this theory is the melting of the Wigner glass, where the transition is second order due to the disorder.

An interesting suggestion comes from Spivak [38]. He predicted that the existence of an intermediate phase between the Fermi liquid and the Wigner crystal phases is a generic property of the two-dimensional *pure* electron liquid in MOSFETs at zero temperature. The physical reason for the existence of these phases is a partial separation of the uniform phases. This phenomenon is due to a tendency for phase separation which originates from the existence of a first-order phase transition between the Fermi liquid and the Wigner crystal phases as a function of n. The difference between the crystal-liquid phase transition in MOSFETs and the usual first order phase transitions in neutral systems is the following. In neutral systems with first order phase transitions the energy of the surface between the phases is positive and the minimum of the free energy corresponds to a minimal surface area and to a global phase separation. In charged systems, like electrons on a positive frozen background, global phase separation does not occur because of a large Coulomb energy associated with a non-uniform distribution of electron density. The electron liquid in MOSFETs, in a sense, is a system intermediate between these two limiting cases. Similarly to the neutral systems with first order phase transitions, the electron liquid in MOSFETs exhibits phase separation. On the other hand the surface energy of a minority phase droplet of a large enough radius turns out to be negative. As a result at different n there is a variety of intermediate phases in this system which are different both from the Fermi liquid and from the Wigner crystal. In the next section we introduce the idea of Spivak and later in Chapter 8 we will discuss the properties of these phases in disordered case in combination with Si-Varma physics and a possible explanation of experimental results on local properties of low density electron gas in Si MOSFETs. Due to the importance of this point of view and its relationship to the physics being addressed in this thesis, we are going to discuss, in more detail, this type of phase separation in pure 2D systems and the properties of these phases.

2.5.1 Spivak's Theory of Phase Separation

Consider a two-dimensional electron liquid of density n in a MOSFET separated by a distance d from a metallic gate [38]. Electrons interact via Coulomb interaction while a global electric neutrality of the system is enforced by the metallic gate with a positive charge density en. The energy density of the system per unit area $\varepsilon(n) = \varepsilon^{(C)} + \varepsilon^{(el)}$ is a sum of the energy density of the capacitor $\varepsilon^{(C)} = (en)^2/2C$ and the internal energy density of the electron liquid $\varepsilon^{(el)}$. In the case of a uniform electron distribution the capacity per unit area is $C = C_0 = 1/d$.

At high electron densities $na_B^2 \gg 1$ the kinetic energy of electrons is larger than the potential energy and the interaction can be taken into account by perturbation theory. (Here a_B is the electron Bohr radius.) In this case the system can be described by Fermi liquid theory, the difference between the effective m^* and the bare m electron masses is small, and $\varepsilon^{(el)} = \varepsilon_L^{(el)} \sim n^2/m$. On the other hand, in the opposite limit $na_B^2 \ll 1$ (but still $nd^2 \gg 1$) the potential Coulomb energy of electrons is much larger than the kinetic energy and the ground state of the system is a Wigner crystal with $\varepsilon^{(el)} = \varepsilon_W^{(el)} = -e^2 n^{3/2}$ (see, for example, [49]). Thus at zero temperature there is a critical electron concentration n_c where the phase transition between the Fermi liquid and the Wigner crystal phases takes place. According to Landau mean field theory this transition is of the first order. The n-dependence of the energy densities of the Fermi liquid $\varepsilon_L(n) = \varepsilon^C + \varepsilon_L^{(el)}$ and the Wigner crystal $\varepsilon_W(n) = \varepsilon^{(C)} + \varepsilon_W^{(el)}$ phases near the critical density n_c is shown schematically in Fig.2.7.

In the limit of small densities $nd^2 \ll 1$, due to the existence of the image charges in the gate, the interaction between adjacent electrons has a dipole character. In this case the ratio between the potential and the kinetic energy decreases as n decreases. Therefore, for the small electron density, n, system is a weakly interacting Fermi liquid. Thus we arrive at the conclusion that there exists another critical point $n_c^{(1)} \sim 1/d^2$ which corresponds to a second Wigner crystal-Fermi liquid transition. The phase diagram of the electron system at $T = 0$ is shown in Fig.2.7b. If $d < d^* \sim 38a_B$, than the system is in the liquid state at any value of n. Here the factor 38 is the result of numerical simulations [4].

Now let us discuss the mean field description of such a phase separation. Assuming $C = C_0$, the qualitative picture of the phase transition is the same as the picture of any first order phase transition in neutral systems. In particular, there is an interval of electron densities $n_W < n < n_L$ shown in the Fig.2.7 where there is a phase separation, which means that there is a spatially nonuniform distribution of the Wigner crystal and Fermi liquid phases coexisting in equilibrium. In the case of large d one can linearize $\varepsilon_{L,W}^{(el)}(n)$ near the point $n = n_c$. As a result, we have

$$n_{L,W} = n_c \pm \frac{(\mu_W - \mu_L)}{2e^2 d}, \tag{2.29}$$

where $\mu_{W,L} = (d\varepsilon_{W,L}^{(el)}/dn)|_{n=n_c}$. One can get from Eq.2.29 an estimate $n_c a_B/d$ for

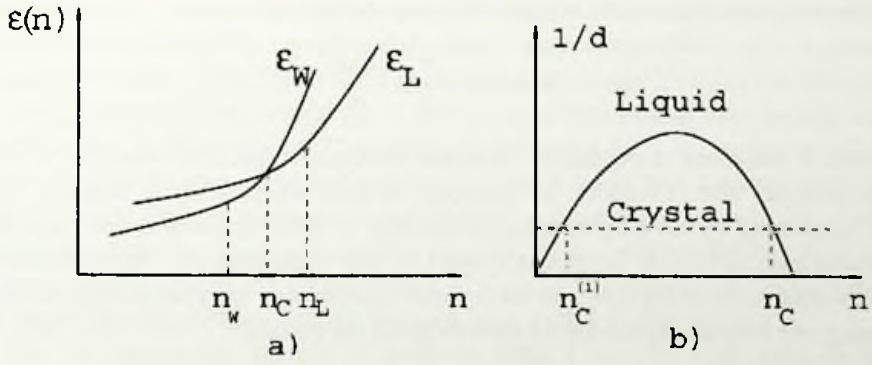


Figure 2.7: a) The dependence of the energy densities of the Wigner crystal and the Fermi liquid phases $\varepsilon_{W,L}(n)$ on the electron density n. Symbols W and L correspond to the Wigner crystal and the Fermi liquid phases respectively. b) The effective phase diagram of the 2D electron system at zero temperature.(from Sipvak [38]

the size of the interval of electron densities where the phase separation occurs. Values of d/a_B in various MOSFETs range from of $O(1)$ to $O(50)$.

The relative fractions of these phases x_W are determined by the tangent construction near n_c. At $(n_L - n) \ll (n_L - n_W)$ the fraction of the area occupied by the Wigner crystal $x_W \ll 1$ is small while in the case $(n - n_W) \ll (n_L - n_W)$ the fraction of the area occupied by the Fermi liquid $x_L \ll 1$ is small.

$$x_{W,L} = \pm \frac{n - n_{W,L}}{n_c}. \tag{2.30}$$

The compressibility of the system $\nu = d^2\varepsilon/dn^2$ should exhibit jumps of order $e^2 d$ at points $n = n_L, n_W$.

The crucial difference between first order phase transitions in neutral systems and in the system of electrons in MOSFETs arises when one considers the shapes of the minority phases. In the case of neutral systems the surface energy density σ is positive. Therefore in equilibrium the system should have a minimal area of the surface separating the phases, leading to global phase separation. On the other hand, in the three dimensional charged systems the global phase separation is impossible because of the large Coulomb energy associated with the charge separation. It is possible, however, that in this case the electron system consists of bubbles and stripes of different electron density, provided the tendency for phase separation is strong enough.

The situation in MOSFETs is very different. On the one hand, in the approximation when $C = C_0$ global phase separation is possible at an arbitrary value of $(\mu_W - \mu_L)$. On the other hand, it turns out that for large droplets of the minority phase the surface energy is *negative*. To prove this one has to take into account the

finite size corrections to the standard formula for the capacitance

$$C = C_0 + \frac{R}{A} \ln \frac{16\pi R}{d} \tag{2.31}$$

where A and d are the capacitor area and thickness respectively and $R = \sqrt{A}$ is the capacitor size. Consider, for example, the case when $x_W \ll 1$. Then x_W can be determined by the tangent construction near n_c in the approximation when the second term in Eq.2.31 is neglected and $C = C_0$. Expanding $\varepsilon(n)$ with respect to the second term in Eq.2.31 and taking into account also the microscopic surface energy we have an expression for the energy of the surface

$$E_{(surf)} = -\frac{1}{2} N_W e^2 (n_W - n_L)^2 d^2 R_W \ln \frac{16\pi R_W}{d} + N_W \sigma 2\pi R_W \tag{2.32}$$

It is assumed that the Wigner crystal phase embedded into the liquid consists of droplets of radius R_W and concentration N_W and take into account that inside the droplet $n \sim n_W$. Thus, at large R_W the surface energy Eq.2.32 turns out to be negative. We have to find a minimum of Eq.2.32 at a given total area occupied by the minority phase, which gives us the characteristic size of the droplet

$$R_W \sim \frac{d}{16\pi} \exp(\gamma) \tag{2.33}$$

with $\gamma = (e^2\sigma)/2\pi(\mu_W - \mu_L)^2$. The analogous calculation for the case $x_L \ll 1$ gives an expression for the radius of liquid droplets embedded into the crystal which is identical to Eq.2.33.

At the point of the transition the values of σ and $(\mu_W - \mu_L)^2/e^2$ are of the same order and at present nothing is known about the value of the dimensionless parameter γ. Even the fact that $\sigma > 0$ is not proved. It is important to note that in the case of the weakly first order phase transitions one we always finds $\gamma \ll 1$. Here it is assumed that $\gamma \geq 1$, which means that the (*clean*) 2D electron liquid on a *frozen* positive background does not exhibit a phase transition.

One can estimate the dependence of $x_{W,L}(T, H_\parallel)$ on the temperature T and the magnetic field $H_\parallel$ parallel to the film. It is determined by the corresponding dependence of the free energies for the Fermi liquid and Wigner crystal phases. At small T and $H_\parallel$ one can neglect the T and $H_\parallel$ dependences of $\varepsilon_{W,L}$ and we have the following expression for the free energy densities of the liquid and the Wigner crystal phases

$$F_{W,L}(H_\parallel) = \varepsilon_{W,L} - M_{W,L} H_\parallel n - T n S_{W,L} \tag{2.34}$$

where S_W and S_L are the entropies of the crystal and the liquid phases respectively, while M_W and M_L are the corresponding spin magnetizations per electron.

As a result, one can obtain how $x_{W,L}(T, H_\parallel)$, and $n_{cW,L}(T, H_\parallel)$ depend on T and $H_\parallel$ by making the following substitution in Eqs.2.29 and 2.30

$$(\mu_W - \mu_L) \to (\mu_W - \mu_L) - (M_W - M_L) H_\parallel - T(S_W - S_L) \tag{2.35}$$

At small $\mu_B H_\| \ll T \ll E_F$ we have $M_{W,L} = \chi_{W,L} H_\|$, where χ_W and χ_L are linear susceptibilities of the crystal and the liquid respectively. (Here μ_B is the Bohr magneton and E_F is the Fermi energy.) At low temperature $T \ll E_F$ the spin susceptibility of the Wigner crystal $\chi_W \sim \mu_B^2/T \gg \chi_L$ is much larger than the spin susceptibility of the Fermi liquid. The entropy of the crystal $S_W \sim \ln 2 \gg S_L \sim T/E_F$ is mainly due to the spin degrees of freedom and much larger than the entropy of the Fermi liquid. Thus x_W increases linearly with T and quadratically with $H_\|$, which means that both the temperature and the magnetic field parallel to the film drive the electron system toward the crystallization [75]. (It is assumed that the temperature is larger than the exchange energy between spins in the Wigner crystal).

In the intermediate interval of magnetic fields $T < \mu_B H_\| < E_F$ spins in the Wigner crystal are completely polarized while the Fermi liquid is still in the linear regime. In this case x_W increases linearly with $H_\|$.

At high magnetic field $H_\| > H_\|^c \sim E_F/\mu_B$ both Fermi liquid and Wigner crystal are spin polarized and $x_W(T, H_\|)$ saturates as a function of $H_\|$. We assume that $\varepsilon_L(H_\| = 0) < \varepsilon_L(H > H_\|^c)$ and, therefore $x_W(H_\| = 0) < x_W(H > H_\|^c)$. On the other hand, the spin entropy of the Wigner crystal is frozen in this case. As a result, at $H_\| > H_\|^c$ the temperature dependence of $x_{W,L}(T, H_\|)$ is suppressed significantly.

Spivak's idea of phase separation in pure 2D electron liquid could qualitatively explain some of the experimental results on 2D electrons in MOSFETs. The most significant outcomes are:

1. The existence of the metal-insulator phase transition.
2. The temperature dependence of the resistance in the metallic phase.
3. The temperature dependence of the resistance in the metallic phase at large $H_\|$.
4. The positive magnetoresistance of the metallic phase in the magnetic field parallel to the film.
5. The n-dependence of $H_\|^c$.

Of course the above consideration is based only on the assumption about the existence of the first order phase transition between the uniform Fermi liquid and the Wigner crystal phases and on electrostatic properties of two dimensional electron system. Clearly the effect of disorder has been completely ignored. One can speculate that the effect of a disorder potential of finite amplitude, in some regimes, would cause the system to acquire a glassy behavior, characteristic for crystals in the presence of disorder. Experimental indications of glassy behavior of the electronic system in Si MOSFETs have already been reported [78, 76, 77].

It is clear that here only bubble phases, which exist near the critical concentrations n_L and n_W, have been considered on a mean field level. In the interval $n_L < n < n_W$ the system, will probably exhibit a sequence of quantum phase transitions. In particular, it is likely that at electron densities close to n_c there are stripe

phases [79, 80, 81, 82]. This consideration is, in many respects, similar to the quantum critical point of strongly correlated electron systems considered by different people [83, 86, 84][laughlin,sachdev,millis]. In particular, the Fermi liquid state with densities close to n_W will demonstrate very large sensitivity to imperfections, which is characteristic for the "almost critical" quantum state [83].

2.6 Conclusions

According to the scaling theory there is no metallic state in 2D as it does not explicitly consider the effect of the Coulomb interaction between electrons. Early theoretical predictions based on HF approximation gives even more localized system for weakly interacting system. Triggered by the recent experimental results, some theoretical explanations have been put forward to tackle the transport properties of high mobility 2D electron/hole systems in MOSFETs. Except for the theory proposed by Si and Varma, which is going to be discussed in Chapter 6, the current theories assume the mean free path of the electrons is always smaller than the screening length. This results in constant interaction parameter in singlet channel, which is responsible for the renormalization of the density of sates of interacting electrons near the MIT. Moreover changes in the thermodynamic behavior of such system has not been addressed explicitly. Non-Fermi liquid behavior and its connection to the recent zero temperature MIT also has been studied by several authors. Within this perspective, a combination of special electrostatic configuration of the MIT experiments in Si-MOSFETs and long-range Coulomb interactions between electrons in such a system, might cause the system to phase separate into high and low density regions near MIT. We are going to address this more in Chapter 8. The possibility of the existence of quantum criticality in these systems also has been suggested. Due to the importance of such point of view, in the next chapter, we are going to introduce the phenomena of quantum critical phase transition very briefly and address the quantum criticality in cuprates superconductors. Later in the last chapter (Chapter 9) we will come back to discuss the MIT phenomena in MOSFETs in this framework.

Chapter 3

Quantum Phase Transition in 2D Correlated Electron Systems

Phase transitions play an essential role in nature. Everyday examples include the boiling of water or the melting of ice, more complicated is the transition of a metal into the superconducting state upon lowering the temperature. The universe itself is thought to have passed through several phase transitions as the high-temperature plasma formed by the Big Bang cooled to form the world as we know it today.

Phase transitions occur upon variation of an external control parameter; their common characteristics is a qualitative change in the system properties. The phase transitions mentioned so far occur at finite temperature; here macroscopic order (e.g. the crystal structure in the case of melting) is destroyed by thermal fluctuations. During recent years, a different class of phase transitions has attracted the attention of physicists, namely transitions taking place at zero temperature. A non-thermal control parameter such as pressure, magnetic field, or chemical composition, is varied to access the transition point. There, order is destroyed solely by quantum fluctuations which are rooted in the Heisenberg uncertainty principle.

Quantum phase transitions [87, 88, 89, 112] have become a topic of vivid interest in current condensed matter physics. At first glance it might appear that the study of such special points in the phase diagram is a marginal problem of interest only to specialists, as such transitions occur at only one special value of a control parameter at the experimentally impossible temperature of absolute zero. However, experimental and theoretical developments in the last decades have clearly established the contrary. They have made clear that the presence of such zero-temperature quantum critical points holds the key to so-far unsolved puzzles in many condensed matter systems. Examples include rare-earth magnetic insulators [91], heavy-fermion compounds [92, 93] high-temperature superconductors [94, 86, 84], and two-dimensional electron gases [87, 7]. In this chapter we are going to address the high T_c system in this context and will be back to the MOSFETs in the last chapter. We start with summarizing the basics of finite-temperature phase transitions, extend those general concepts to $T = 0$ highlighting the interplay

between classical and quantum fluctuations near a quantum critical point, and illustrate the correspondence between quantum and calssical transitions. This part is mainly a selection of the most important concepts which one can find in great detail in the litratures [87, 88, 89, 112, 94, 86, 84, 85] Then, we introduce a very interesting thermodynamical tool to probe QPTs in critical systems, namely the "Grüneisen ratio", $\Gamma = \alpha / c_p$, where α is the thermal expansion and c_p is the specific heat. After that we discuss the corresponding Rosch-Si scaling analysis and by using this concept we will discuss the possibility of QPT in cuprate superconductors.

3.1 Phase transitions: Basic Concepts

We start out with briefly mentioning the basic concepts of phase transitions and criticality [95, 96] which are necessary for the later discussions. Phase transitions are traditionally classified into first-order and continuous transitions. At first-order transitions the two phases co-exist at the transition temperature. In contrast, at continuous transitions the two phases become the same. The transition point of a continuous phase transition is also called critical point. The study of phase transitions, continuous phase transitions in particular, has been one of the most fertile branches of theoretical physics in the last decades.

In the following we concentrate on systems near a continuous phase transition. Such a transition can usually be characterized by an order parameter – this is a thermodynamic quantity that is zero in one phase (the disordered) and non-zero and non-unique in the other (the ordered) phase. Very often the choice of an order parameter for a particular transition is obvious as, e.g., for the ferromagnetic transition where the total magnetization is an order parameter. However, in some cases finding an appropriate order parameter is complicated and still a matter of debate, e.g., for the interaction-driven metal–insulator transition in electronic systems (the Mott transition [97]).

While the thermodynamic average of the order parameter is zero in the disordered phase, its fluctuations are non-zero. If the critical point is approached, the spatial correlations of the order parameter fluctuations become long-ranged. Close to the critical point their typical length scale, the correlation length ξ, diverges as

$$\xi \propto |t|^{-\nu}, \tag{3.1}$$

where ν is the correlation length critical exponent and t is some dimensionless measure of the distance from the critical point. If the transition occurs at a non-zero temperature T_c, it can be defined as $t = |T - T_c|/T_c$.

In addition to the long-range correlations in space there are analogous long-range correlations of the order parameter fluctuations in time. The typical time scale for a decay of the fluctuations is the correlation (or equilibration) time τ_c. As the critical point is approached the correlation time diverges as

$$\tau_c \propto \xi^z \propto |t|^{-\nu z}, \tag{3.2}$$

where z is the dynamic critical exponent. Close to the critical point there is no characteristic length scale other than ξ and no characteristic time scale other than τ_c. (Note that a microscopic cutoff scale must be present to explain non-trivial critical behavior, for details see, e.g., Goldenfeld [96]. In a solid such a scale is, e.g., the lattice spacing.)

The divergencies (3.1) and (3.2) are responsible for the so-called critical phenomena. At the phase transition point, correlation length and time are infinite, fluctuations occur on *all* length and time scales, and the system is said to be scale-invariant. As a consequence, all observables depend via power laws on the external parameters. The set of corresponding exponents – called critical exponents – completely characterizes the critical behavior near a particular phase transition.

Let us illustrate the important concept of scaling in more detail. To be specific, consider a classical ferromagnet with the order parameter being the magnetization $M(\mathbf{r})$. External parameters are the reduced temperature $t = |T - T_c|/T_c$ and the external magnetic field B conjugate to the order parameter. Close to the critical point the correlation length is the only relevant length scale, therefore the physical properties must be unchanged if we rescale all lengths in the system by a common factor, and at the same time adjust the external parameters in such a way that the correlation length retains its old value. This gives rise to the homogeneity relation for the singular part of the free energy density,

$$f(t,B) = b^{-d} f(t\, b^{1/\nu}, B b^{y_B}). \tag{3.3}$$

Here y_B is another critical exponent. The scale factor b is an arbitrary positive number. Analogous homogeneity relations for other thermodynamic quantities can be obtained by differentiating f. The homogeneity law (3.3) was first obtained phenomenologically by Widom and Kadanoff [98]; within the framework of the renormalization group theory [99] it can be derived from first principles.

In addition to the critical exponents ν, y_B and z defined above, a number of other exponents is in common use. They describe the dependence of the order parameter and its correlations on the distance from the critical point and on the field conjugate to the order parameter. The definitions of the most commonly used critical exponents are summarized in Table 3.1.

Note that not all the exponents defined in Table 3.1 are independent from each other. The four thermodynamic exponents $\alpha, \beta, \gamma, \delta$ can all be obtained from the free energy (3.3) which contains only two independent exponents. They are therefore connected by the so-called scaling relations

$$2 - \alpha = 2\beta + \gamma, \quad 2 - \alpha = \beta(\delta + 1)\,. \tag{3.4}$$

Analogously, the exponents of the correlation length and correlation function are connected by two so-called hyperscaling relations

	exponent	definition	conditions
specific heat	α	$C \propto \lvert t\rvert^{-\alpha}$	$t \to 0, B = 0$
order parameter	β	$m \propto (-t)^{\beta}$	$t \to 0$ from below, $B = 0$
susceptibility	γ	$\chi \propto \lvert t\rvert^{-\gamma}$	$t \to 0, B = 0$
critical isotherm	δ	$B \propto \lvert m\rvert^{\delta}\mathrm{sign}(m)$	$B \to 0, t = 0$
correlation length	ν	$\xi \propto \lvert t\rvert^{-\nu}$	$t \to 0, B = 0$
correlation function	η	$G(r) \propto \lvert r\rvert^{-d+2-\eta}$	$t = 0, B = 0$
dynamic	z	$\tau_c \propto \xi^z$	$t \to 0, B = 0$

Table 3.1: Commonly used critical exponents for magnets, where the order parameter is the magnetization m, and the conjugate field is a magnetic field B. t denotes the distance from the critical point and d is the space dimensionality. (The exponent y_B defined in (3.3) is related to δ by $y_B = d\,\delta/(1+\delta)$.)

$$2 - \alpha = d\nu\,, \quad \gamma = (2-\eta)\nu\,. \tag{3.5}$$

(Hyperscaling relations are violated in theories with mean-field critical behavior due to the presence of a dangerously irrelevant variable.) Since statics and dynamics decouple in classical statistics, the dynamic exponent z is completely independent from all the others.

One of the most remarkable features of continuous phase transitions is universality, i.e., the fact that the critical exponents are the same for entire classes of phase transitions which may occur in very different physical systems. These universality classes are determined only by the symmetries of the order parameter and by the space dimensionality of the system.

This implies that the critical exponents of a phase transition occurring in nature can be determined exactly (at least in principle) by investigating *any* simple model system belonging to the same universality class. The mechanism behind universality is again the divergence of the correlation length. Close to the critical point the system effectively averages over large volumes rendering the microscopic details of the Hamiltonian unimportant.

3.2 Criticality and Quantum Mechanics

The question to what extent quantum mechanics is important for understanding a continuous phase transition has at least two aspects. On the one hand, quantum mechanics can be essential to understand the existence of the ordered phase, (e.g., superconductivity) – this depends on the particular transition considered. On the other hand, one may ask whether quantum mechanics influences the asymptotic

critical behavior. For this discussion we have to compare two energy scales, namely $\hbar\omega_c$, which is the typical energy of long-distance order parameter fluctuations, and the thermal energy $k_B T$. We have seen in the preceeding section that the typical time scale τ_c of the fluctuations diverges as a continuous transition is approached. Correspondingly, the typical frequency scale ω_c goes to zero and with it the typical energy scale

$$\hbar\omega_c \propto |t|^{\nu z} . \tag{3.6}$$

Quantum mechanics will be important as long as this typical energy scale is larger than the thermal energy $k_B T$; on the other hand, for $\hbar\omega_c \ll k_B T$ a purely classical description can be applied to the order parameter fluctuations. In other words, the character of the order parameter fluctuations crosses over from quantum to classical when $\hbar\omega_c$ falls below $k_B T$.

Now, for any transition occurring at some finite temperature T_c quantum mechanics will become unimportant for $|t| < T_c^{1/\nu z}$, in other words, the critical behavior asymptotically close to the transition is entirely classical. This justifies to call all finite-temperature phase transitions "classical". Quantum mechanics can still be important on microscopic scales, but classical thermal fluctuations dominate on the macroscopic scales that control the critical behavior. If, however, the transition occurs at zero temperature as a function of a non-thermal parameter r like pressure or magnetic field, the behavior is always dominated by quantum fluctuations. Consequently, transitions at zero temperature are called "quantum" phase transitions.

The interplay of classical and quantum fluctuations leads to an interesting phase diagram in the vicinity of the quantum critical point. Two cases need to be distinguished, depending on whether long-range order can exist at finite temperatures.

Fig. 3.1a describes the situation where order only exists at $T = 0$, this is the case, e.g., in two-dimensional magnets with SU(2) symmetry where order at finite T is forbidden by the Mermin-Wagner theorem. In this case there will be no true phase transition in any real experiment carried out at finite temperature. However, the finite-T behavior is characterized by three very different regimes, separated by crossovers, depending on whether the behavior is dominated by thermal or quantum fluctuations of the order parameter. In the thermally disordered region the long-range order is destroyed mainly by thermal order parameter fluctuations. In contrast, in the quantum disordered region the physics is dominated by quantum fluctuations, the system essentially looks as in its quantum disordered ground state at $r > r_c$. In between is the so-called quantum critical region [100], where both types of fluctuations are important. It is located near the critical parameter value $r = r_c$ at comparatively high (!) temperatures. Its boundaries are determined by the condition $k_B T > \hbar\omega_c \propto |r - r_c|^{\nu z}$: the system "looks critical" with respect to the tuning parameter r, but is driven away from criticality by thermal fluctuations. Thus, the physics in the quantum critical region is controlled by the thermal excitations of the quantum critical ground state, whose main characteristics is the *absence* of conventional quasiparticle-like excitations. This causes unusual finite-temperature

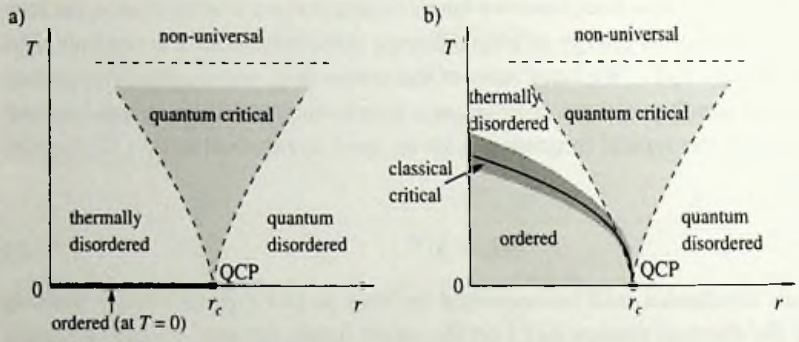


Figure 3.1: Schematic phase diagrams in the vicinity of a quantum critical point (QCP). The horizontal axis represents the control parameter r used to tune the system through the quantum phase transition, the vertical axis is the temperature T. a) Order is only present at zero temperature. The dashed lines indicate the boundaries of the quantum critical region where the leading critical singularities can be observed; these crossover lines are given by $k_B T \propto |r - r_c|^{\nu z}$. b) Order can also exist at finite temperature. The solid line marks the finite-temperature boundary between the ordered and disordered phases. Close to this line, the critical behavior is classical. (from Vojta [85])

properties in the quantum critical region, such as unconventional power laws, non-Fermi liquid behavior etc.

Universal behavior is only observable in the vicinity of the quantum critical point, i.e., when the correlation length is much larger than microscopic length scales. Quantum critical behavior is thus cut off at high temperatures when $k_B T$ exceeds characteristic microscopic energy scales of the problem – in magnets this cutoff is, e.g., set by the typical exchange energy.

If order also exists at finite temperatures, Fig. 3.1b, the phase diagram is even richer. Here, a real phase transition is encountered upon variation of r at low T; the quantum critical point can be viewed as the endpoint of a line of finite-temperature transitions. As discussed above, classical fluctuations will dominate in the vicinity of the finite-T phase boundary, but this region becomes narrower with decreasing temperature, such that it might even be unobservable in a low-T experiment. The fascinating quantum critical region is again at finite temperatures above the quantum critical point.

A quantum critical point can be generically approached in two different ways: $r \to r_c$ at $T = 0$ or $T \to 0$ at $r = r_c$. The power-law behavior of physical observables in both cases can often be related. Let us discuss this idea by looking at the entropy S. It goes to zero at the quantum critical point (exceptions are impurity transitions), but its derivatives are singular. The specific heat $c = T\partial S/\partial T$ is expected to show power-law behavior, as does the quantity $\alpha = \partial S/\partial r$. Using scaling arguments (see below) one can now analyze the ratio of the *singular parts* of α and c: the scaling dimensions of T and S cancel, and therefore $alpha/c$ scales as the inverse of the tuning parameter r. Thus, one obtains a *universal* divergence in the low-

temperature limit, $\alpha/c \propto |r-r_c|^{-1}$; similarly, $\alpha/c \propto T^{-1/\nu z}$ at $r = r_c$. Note that α/c does not diverge at a finite-temperature phase transition. At a pressure-tuned phase transition $r \equiv p$, then α measures the thermal expansion, and α/c_p is the so-called Grüneisen parameter [102]. The scaling argument presented here can be invalid above the upper-critical dimension.

3.3 Scaling Analysis and Qunatum Correspondence

For deeper understanding of the relation between classical and quantum behavior, and the possible quantum–classical crossover, we have to recall general features from quantum statistical mechanics.

The starting point for the derivation of thermodynamic properties is the partition function

$$Z = \mathrm{Tr} e^{-H/k_B T} \tag{3.7}$$

where $H = H_{\mathrm{kin}} + H_{\mathrm{pot}}$ is the Hamiltonian characterizing the system. In a classical system, the kinetic and potential part of H commute, thus Z factorizes, $Z = Z_{\mathrm{kin}} Z_{\mathrm{pot}}$, indicating that in a classical system statics and dynamics decouple. The kinetic contribution to the free energy will usually not display any singularities, since it derives from the product of simple Gaussian integrals. Therefore one can study classical phase transitions using effective time-independent theories, which naturally live in d dimensions, where d is the spatial dimensionality of the system under consideration.

In contrast, in a quantum problem the kinetic and potential parts of H in general do not commute, the quantum mechanical partition function does *not* factorize, which implies that statics and dynamics are always coupled. An order parameter field theory needs to be formulated in terms of space and time dependent fields. The canonical density operator $e^{-H/k_B T}$ looks exactly like a time evolution operator in imaginary time τ if one identifies $1/k_B T = \tau = -i\Theta/\hbar$ where Θ denotes the real time. Therefore it proves convenient to introduce an imaginary time direction into the system – formally this is done in the path integral representation of the partition function. At zero temperature the imaginary time acts similarly to an additional space dimension since the extension of the system in this direction is infinite. According to (3.2), time scales like the z-th power of a length. (Note that $z = 1$ for many transitions in clean insulators, however, in general other values of z including fractional ones can occur.) The classical homogeneity law (3.3) for the free energy density can now easily be adopted to the case of a quantum phase transition. At zero temperature it reads

$$f(t,B) = b^{-(d+z)} f(t b^{1/\nu}, B b^{y_B}) \tag{3.8}$$

where now $t = |r-r_c|/r_c$. This shows that a quantum phase transition in d space dimensions is related to a classical transition in $(d+z)$ space dimensions.

For a quantum system, it is now interesting to discuss the quantum to classical crossover upon approaching a *finite-temperature* phase transition – this turns out

to be equivalent to a dimensional crossover. Under the quantum–classical mapping the temperature of the quantum problem maps onto the inverse length of the imaginary time axis. (Note that the temperature of the classical problem corresponds to a coupling constant in the quantum model within this mapping.) Close to the transition (Fig. 3.1b) the behavior is determined by the relation between the characteristic correlation time, τ_c, and the extension in imaginary time direction, $\beta = 1/k_B T$. The crossover from quantum to classical behavior will occur when the correlation time τ_c becomes larger than β which is equivalent to the condition $|t|^{\nu z} < k_B T$; in other words, once $\tau_c > \beta$ the system realizes that it is effectively only d-dimensional and not $(d+z)$-dimensional. The corresponding crossover scaling is equivalent to finite size scaling in imaginary time direction.

In contrast, when approaching a *zero-temperature* transition by lowering the temperature at $r = r_c$, both $1/k_B T$ and τ_c diverge, such that quantum effects are always important, and we get a truly $(d+z)$-dimensional system. Recognizing that quantum critical singularities can be cut off both by tuning r away from r_c at $T = 0$ and by raising T at $r = r_c$, it is useful to generalize the homogeneity law (3.8) to finite temperatures,

$$f(t, B, T) = b^{-(d+z)} f(t\, b^{1/\nu}, B\, b^{y_B}, T\, b^{z}) \,. \tag{3.9}$$

The homogeneity law for the free energy immediately leads to scaling behavior of both static and dynamic observables. Consider an observable $O(k, \omega)$, e.g., a magnetic susceptibility, measured at wavevector k and frequency ω, then the existence of a single length scale ξ and a single time scale $\omega_c^{-1} = \xi^z$ implies

$$\begin{aligned} O(t, k, \omega, T) &= \xi^{d_O}\, O_1(k\xi, \omega\xi^z, T\xi^z) \\ &= T^{-d_O/z}\, O_2(kT^{-1/z}, \omega/T, T\xi^z) \end{aligned} \tag{3.10}$$

where O_1 and O_2 are different forms for the scaling function associated with the observable O, and d_O is the so-called scaling dimension of O. For simplicity we have set the external field $B = 0$ and assumed that k is measured relative to the ordering wavevector.

Precisely at the critical point, where the correlation length is infinite, the only length is set by the measurement wavevector k, analogously, the only energy is ω, leading to

$$O(t = 0, k, \omega, T = 0) = k^{-d_O}\, O_3(k^z/\omega) \,. \tag{3.11}$$

Similarly, for $k = 0$, but a finite temperatures, we obtain

$$O(t = 0, k = 0, \omega, T) = T^{-d_O/z}\, O_4(\omega/T) \,, \tag{3.12}$$

this is known as "ω/T" scaling. It is important to note that the relations (3.10)–(3.12), sometimes also called "naive scaling", are only expected to be valid if the critical point satisfies hyperscaling properties, which is true below the upper-critical dimension.

The power of the quantum–classical analogy can be nicely demonstrated by considering a correlation function at the critical point. In a classical d-dimensional system, correlations typically fall off with a power law in real space implying a momentum space behavior as

$$G(k) \propto k^{-2+\eta_d} . \tag{3.13}$$

If we are now interested in a quantum phase transition with $z = 1$ which maps onto a $(d+1)$-dimensional classical theory with known properties, we can conclude that at $T = 0$

$$G(k, i\omega_n) \propto [k^2 + \omega_n^2]^{(-2+\eta_{d+1})/2} \tag{3.14}$$

where the Matsubara frequency ω_n simply takes the role of an additional wavevector component. For real frequencies, the above translates into a retarded Green's function of the form

$$G^R(k, \omega) \propto [k^2 - (\omega + i\delta)^2]^{(-2+\eta_{d+1})/2} . \tag{3.15}$$

This examples illustrates the character of the excitation spectrum at a quantum critical point: G^R does not show a conventional quasiparticle pole, but instead a branch cut for $\omega > k$, corresponding to a critical continuum of excitations. This implies that modes become overdamped, and the system shows quantum relaxational dynamics [89].

Now, if a quantum transition seems to map generically onto a classical transition, what is different about quantum transitions? For a number of reasons it turns out that *not* all needed properties of a given quantum system can be obtained from a classical theory, the reasons being:

(A) Theories for quantum systems can have ingredients which make them qualitatively different from their classical counterparts; examples are topological Berry phase terms and long-ranged effective interactions arising from soft modes.

(B) Even if a quantum–classical mapping is possible, calculating real time dynamics of the quantum system often requires careful considerations and novel theories, as an approximation done on the imaginary time axis is in general not appropriate for real times [89].

(C) Quenched disorder leads to an extreme anisotropy in space-time in the classical theory [101].

(D) The dynamic properties of the quantum system near a quantum critical point are characterized by a fundamental new time scale, the *phase coherence time* τ_ϕ, which has no analogue at the corresponding classical transition [89]. Notably, τ_ϕ diverges as $T \to 0$ for all parameter values, i.e., the quantum system has perfect phase coherence even in the disordered phase – this seems to be peculiar considering that all correlations decay exponentially in the high-temperature phase of the corresponding $(d+z)$-dimensional classical model. Technically, this peculiarity is related to the nature of the analytic continuation between imaginary and real times. In many models for quantum phase transitions, it is found that $\tau_\phi \propto 1/T$ as $T \to 0$ at the critical coupling, but τ_ϕ diverges faster in the stable phases.

3.4 Probing QPT and Rosch-Si Scaling

As been discussed in the previous sections the presence of Quantum Critical Points (QCPs) occurs in systems where a continuous quantum phase transition (QPT) at $T = 0$ is induced by tuning some control parameter like pressure p, doping or magnetic field H. Such zero-temperature critical points can determine the properties of materials in a wide range of temperatures. In general, quantum critical points are more difficult to characterize compared to their classical counterparts. At a classical critical point, thermodynamic quantities typically diverge; the associated critical exponents historically played a central role in our eventual understanding of scaling and universality. Some of these divergences, however, have to disappear at a QCP: there are constraints placed by the third law of thermodynamics due to the very fact that the transition takes place at zero temperature. Recently Rosch and Si [102] have showed that at the quantum critical point, reached by varying presssure[1], the thermal expansion α is more singular than the specific heat c_p. Consequently, the "*Gruneisen ratio*" $\Gamma = \alpha/c_p$, *diverges.* When scaling applies, $\Gamma \sim T^{-1/(\nu z)}$ at the critical pressure $p = p_c$, providing a means to measure the scaling dimension of the most relevant operator that pressure couples to; in the alternative limit $T \to 0$ and $p \neq p_c$ with a prefactor that is, up to the molar volume, a simple *universal* combination of critical exponents. Since we are going to use the same idea to address the presence of QCP in cuprates superconductors as well as 2D electron system let us look at this analysis in more detail

We define the Grüneisen ratio Γ [103, 104] in terms of the molar specific heat[2]

$$c_p = \frac{T}{N}\frac{\partial S}{\partial T}\bigg|_p, \tag{3.16}$$

and the thermal expansion

$$\alpha = \frac{1}{V}\frac{\partial V}{\partial T}\bigg|_{p,N} = -\frac{1}{V}\frac{\partial S}{\partial p}\bigg|_{T,N} \tag{3.17}$$

as

$$\Gamma = \frac{\alpha}{c_p} = -\frac{1}{V_m T}\frac{\partial S/\partial p}{\partial S/\partial T} \tag{3.18}$$

where S is the entropy and $V_m = V/N$ the molar volume. In ordinary situations, pressure dependences are regular and a finite Grüneisen ratio is expected as is

[1] In cuprates pressure seems quite irrelivant as coupling constant.

[2] The "effective" or "thermodynamic" Grüneisen parameter is usually defined as $\Gamma c_p/(\kappa_T c_V)$, where κ_T is the compressibility. We think that our Γ is both experimentally and theoretically the more convenient quantity. Furthermore, neither κ_T nor c_p/c_V is expected to be singular at a magnetic QCP.

indeed observed in all previous measurements of this quantity in the literature. Such a regular dependence is typically described by assuming that the system is dominated by a single energy scale E^* (e.g. the Fermi energy in a metal or the Debye energy if acoustic phonons dominate) so the molar entropy takes the form

$$S/N = f(T/E^*). \tag{3.19}$$

The Grüneisen ratio is then temperature independent and given by[103, 104, 105, 106]

$$\Gamma = \frac{1}{V_m E^*} \frac{\partial E^*}{\partial p}. \tag{3.20}$$

However, this formula already suggests that a diverging Γ can be expected when some energy scale E^* vanishes as it happens at a QCP.

A quantum critical point is reached in a singular fashion by tuning some external parameter and, in general, this external parameter is thermodynamically coupled to pressure (exceptions are high-T_c superconductors, magnets, and the superconductor-insulator transition) [3]. In the low temperature limit, the singular terms of S and T in Eq. (3.18) cancel out leaving Γ to depend only on singularities associated with the pressure p. As the pressure controls the QPT, such a singularity always exists and the Grüneisen ratio diverges at any QCP. This divergence is entirely determined by the scaling dimension of the control parameter, which is the most relevant operator to which the pressure couples. As shown below, this leads to a T dependence, $\Gamma \sim 1/T^{1/\nu z}$ [see Eq. (3.26)]. *In other words, the temperature exponent of the Grüneisen ratio provides a direct means to measure* νz *and, as a result, characterize a QCP.* Put in a slightly different way, the thermal expansion contains valuable information complementary to that obtained from the specific heat: while c_p measures the response to T, α describes the response to the tuning parameter of the QPT, the second relevant variable at a QCP. This has to be contrasted with a classical phase transition. There, generically, only one relevant operator exists to which both T and p couple. Accordingly, Γ will be constant close to a classical transition.

To observe the singular behavior of Γ or the thermal expansion the pressure has to couple sufficiently strongly to the critical dynamics. This is for example the case in heavy fermion compounds where the intricate competition between magnetic interactions and the Kondo effect can be tuned by pressure, doping or magnetic field to yield a QPT, typically from a metallic antiferromagnet to a metallic paramagnet. The high sensitivity, e.g. to pressure, arises from the exponential dependence of the Kondo temperature on system parameters. Whether the transitions in these systems conform to the Gaussian picture associated with $T = 0$ spin-density wave (SDW) transitions[107, 108] or are non-Gaussian as in a locally quantum critical point[109] is a question of great current interest.

[3] The "universality"of pressure is overemphasized in the HF community.

3.4.1 Rosch-Si Scaling Analysis

Close to any QCP, the correlation length ξ diverges as a function of a control parameter r, $\xi \sim |r|^{-\nu}$, where e.g. $r = (p - p_c)/p_c$ or $r = (H - H_c)/H_c$. Correspondingly, a typical correlation (imaginary) time, $\xi_\tau \sim \xi^z$, diverges as the QCP is approached. The "dynamical critical exponent" z depends on the dynamics of the order parameter and relates time and length scales.

If one assumes that the critical behavior is governed by ξ and ξ_τ (a more careful discussion of this assumption is given below), the critical contribution to the free energy per mole, $F_{cr} = F - F_{\text{reg}}$, can be cast into the following standard scaling *Ansatz* (using hyperscaling)

$$\begin{aligned} \frac{F_{cr}}{N} &= -\rho_0 r^{\nu(d+z)} \tilde{f}\left(\frac{T}{T_0 r^{\nu z}}\right) \\ &= -\rho_0 \left(\frac{T}{T_0}\right)^{(d+z)/z} f\left(\frac{r}{(T/T_0)^{1/(\nu z)}}\right), \end{aligned} \tag{3.21}$$

where ρ_0 and T_0 are non-universal constants, while $f(x)$ and $\tilde{f}(x)$ are universal scaling functions. Obviously, $f(x \to 0) \approx f(0) + x f'(0) + \ldots$ is regular as there is no phase transition at $r = 0, T > 0$. The limit $\tilde{f}(x \to 0) = \tilde{f}(0) + c x^{y_0+1}$ describes the low temperature behavior of the phases to the left or right side of the QCP (in general different for $r > 0$ and $r < 0$). Note that the exponent $y_0 > 0$ has to be positive due to the third law of thermodynamics. It characterizes the power-law behavior of the specific heat $c_p \sim T^{y_0}$, e.g. $y_0 = 1$ for a Fermi liquid, $y_0 = 2$ for a d-wave superconductor in $d = 2$, or $y_0 = d$ and $d/2$ for an insulating antiferromagnet and ferromagnet, respectively.

Thermodynamical quantities are easily obtained from (3.21). The critical contribution c_{cr} to the specific heat at $r = 0$ is given by

$$c_{cr}(T, r = 0) = \frac{(d+z)d}{z^2} \frac{\rho_0}{T_0} f(0) \left(\frac{T}{T_0}\right)^{d/z} \tag{3.22}$$

and for $T \to 0$, $r \neq 0$

$$c_{cr}(T \to 0, r) = \frac{\rho_0 c y_0 (y_0 + 1)}{T_0} \left(\frac{T}{T_0}\right)^{y_0} r^{\nu(d - y_0 z)}. \tag{3.23}$$

Similarly, in the case of a pressure tuned QCP with $r = (p - p_c)/p_c$ the critical contribution α_{cr} to the thermal expansion reads

$$\alpha_{cr}(T, r = 0) = -\frac{d + z - \frac{1}{\nu}}{z} \frac{\rho_0 f'(0)}{T_0 p_c V_m} \left(\frac{T}{T_0}\right)^{(d - \frac{1}{\nu})/z} \tag{3.24}$$

and for $r \neq 0$

$$\alpha_{cr}(T \to 0) = -\frac{\rho_0 (y_0 + 1) c \nu (d - y_0 z)}{T_0 V_m} \frac{r^{\nu(d - y_0 z)}}{p_c r} \left(\frac{T}{T_0}\right)^{y_0}. \tag{3.25}$$

The thermal expansion is more singular than the specific heat leading to a Grüneisen ratio

$$\Gamma_{cr}(T, r=0) = \frac{\alpha_{cr}}{c_{cr}} = -G_T T^{-1/(\nu z)}, \tag{3.26}$$

where the prefactor

$$G_T = \frac{(d+z-1/\nu) z f'(0)}{(d+z) d f(0)} \frac{T_0^{1/(\nu z)}}{p_c V_m} \tag{3.27}$$

contains some non-universal parameters (p_c and T_0). *We reach the important conclusion that the temperature exponent of the Grüneisen ratio is equal to* $1/\nu z$. In the other limit $T \to 0$, $r \neq 0$ we obtain the universal result

$$\Gamma_{cr}(T \to 0, r) = -G_r \frac{1}{V_m (p - p_c)}. \tag{3.28}$$

Remarkably, even the (generally unknown) scaling functions cancel out in the amplitude G_r, leaving only a combination of critical exponents and the dimensionality:

$$G_r = \frac{\nu(d - y_0 z)}{y_0}. \tag{3.29}$$

Note that the universality of this prefactor is connected to the third law of thermodynamics – a finite residual entropy per volume ($y_0 = 0$) would spoil this result.

It is rather difficult to measure thermal expansion inside a pressure cell. However, in many systems doping acts like "chemical pressure". If doping x and pressure p can be quantitatively related, $p - p_c = c(x - x_c)$, a measurement of Γ for different samples at ambient pressure can be used to check the prediction (3.28) quantitatively. For generic tuning parameters, we need to substitute $(\partial r/\partial p)$ for $1/p_c$ in Eqs. (3.24,3.25) and modify Eqs. (3.26,3.28) accordingly.

3.5 Thermodynamics and Quantum Criticality in Cuprate Superconductors

The first successful application of the above analysis has been applied in several heavy fermion intermetallic systems. Here we are going to address the universality of QTP in cuprate superconductors. The possible existence of quantum phase transitions (QPT's) in a variety of condensed matter systems is attracting much interest[85, 89]. The cuprate high Tc superconductors have played a prominent role in this development since it has been suspected for a long time[84, 110, 111, 112, 113] that the state realized at the doping where the superconducting transition temperature is maximal (x_{opt}) is controlled by a continuous QPT. This suspicion is mainly motivated by the observation of the 'wedge' in the doping (x)-temperature (T) plane set by the 'pseudogap' ($T_{PG}(x)$) and 'coherence'[114, 115]

($T_{coh}(x)$) crossover temperatures, bordering a 'quantum-critical' (QC) region characterized by power law behaviors. It is believed that this signals a QPT from a poorly understood 'pseudo-gap' phase at low dopings to a Fermi-liquid at high dopings. Although direct evidences appeared for the presence of scale invariance of the quantum dynamics in the QC regime[116], it is unclear if this 'critical state' is truely critical in the sense that it is characterized by universality and hyperscaling[89]. Given that apparently fermionic degrees of freedom are involved, this is from a theoretical point of view far from obvious because the fermion signs obscure the analogy with thermal phase transitions[117]. One would like to establish emperically the presence of scaling laws, revealing universality. Such evidences are lacking in the cuprates.

Thermodynamics is of course in the first instance associated with temperature. A classical phase transition is driven by temperature, but this is profoundly different for a quantum phase transition (See first section). The QPT is driven by a zero temperature control parameter r, and the path integral formalism shows that temperature takes the role of a finite size[89], as the compactification radius of the imaginary time dimension $L_\tau = \hbar/(k_B T)$. As we saw in the previous section, the essence of the Rosch-Si scaling analysis[102] is that one has to determine the dependence of the free energy relative to variations of the coupling constant to learn about the quantum singularity. However, standard thermodynamics associated with variations of temperature gives additional information of the finite size scaling variety. Their combination yields a powerful phenomenological scaling tool box.

Following Si-Rosch[102], our analysis rests on a single theoretical assumption. It is assumed that the QPT is associated with an unstable fixed point at zero temperature, reached by tuning a single zero temperature variable y such that $r = (y - y_c)/y_c$ measures the distance from the critical point residing at y_c. Since temperature T corresponds with L_τ it enters the singular part of the free energy density F_s as a finite size under a scale transformation $x \to bx$,

$$F_s(r,T) = b^{-(d+z)} F_s(b^{y_r} r, b^z T) \,, \tag{3.30}$$

where d is the space dimensionality and z the dynamical exponent, while hyperscaling is assumed in order to relate the finite size to the scaling dimension y_r of the coupling constant ($y_r = 1/\nu$ where ν is the correlation length exponent). Eq. (3.30) is equivalent to the Eq. (3.21) of scaling forms for the free energy density, Since there is no singularity at $r = 0, T > 0$, $f(x \to 0) \simeq f(0) + xf'(0) + (1/2)x^2 f''(0) + \cdots$ while $\tilde{f}(x) = \tilde{f}(0) + g(x)$ where $g(x)$ describes the low temperature thermodynamics of the phases to the left- or right side of the QPT. When the phase is fully gapped $g(x) \sim e^{-1/x}$ while for a massless phase $g(x) = cx^{y_0+1}$ such that y_0 corresponds with its specific heat exponent ($y_0 = 1$ for a Fermi-liquid, and 2 for a 'nodal liquid' characterized by d-wave like 'Dirac cones').

We find it convenient to parametrize the exponents in terms of d, z and the zero-temperature analogue of the specific heat exponent α characterizing a thermal

phase transition,

$$\alpha_r = 2 - \frac{d+z}{y_r} \tag{3.31}$$

In analogy with classical criticality, we expect this exponent to be a fraction of unity. We will consider the specific heat $C = -T(\partial^2 F/\partial T^2)$ and the quantity $\eta_r = (\partial^2 F/\partial r \partial T)$, revealing the dependence of the entropy on the coupling constant. However, we will extend the analysis by also including the 'coupling constant susceptibility' $\chi_r = \partial^2 F/\partial r^2$, being the quantity which is actually most sensitive to the zero-temperature singularity.

From the scaling forms Eq. (3.21) and the above definitions it follows that the singular parts of various measurable quantities have the following temperature dependence in the quantum critical state ($r = 0$),

$$\begin{aligned}
C_{cr}(T, r=0) &= \rho_0 f(0) \frac{(d+z)d}{z^2} \left(\frac{T}{T_0}\right)^{d/z}, \\
\eta_{r,cr}(T, r=0) &= -\frac{\rho_0 f'(0)}{T_0} \frac{1-\alpha_r}{2-\alpha_r} \frac{d+z}{z} \\
&\quad \times \left(\frac{T}{T_0}\right)^{(d(1-\alpha_r)-z)/(z(2-\alpha_r))}, \\
\chi_{r,cr}(T, r=0) &= -\rho_0 f''(0) \left(\frac{T}{T_0}\right)^{-((d+z)\alpha_r)/(z(2-\alpha_r))}.
\end{aligned} \tag{3.32}$$

On the other hand, in the massless phase characterized by a specific heat exponent y_0, at low temperatures in the vicinity of the QPT,

$$\begin{aligned}
C_{cr}(T \to 0, r) &= \frac{\rho_0 c}{T_0} y_0 (y_0 + 1)\, r^{(2-\alpha_r)(d-y_0 z)/(d+z)} \\
&\quad \times \left(\frac{T}{T_0}\right)^{y_0}, \\
\eta_{r,cr}(T \to 0, r) &= -\frac{\rho_0 c}{T_0} \frac{(d-y_0 z)}{d+z} (y_0+1)(2-\alpha_r) \\
&\quad \times r^{(2-\alpha_r)(d-y_0 z)/(d+z)-1} \left(\frac{T}{T_0}\right)^{y_0}, \\
\chi_{r,cr}(T \to 0, r) &= -\rho_0 \tilde{f}(0)(d+z-1)(2-\alpha_r) r^{-\alpha_r} \\
&\quad -c\rho_0 \frac{(d-y_0 z)}{d+z}((2-\alpha_r)\left(\frac{d-y_0 z}{d+z}\right) - 1) \\
&\quad \times r^{(2-\alpha_r)(d-y_0 z)/(d+z)-2} \left(\frac{T}{T_0}\right)^{y_0+1}.
\end{aligned} \tag{3.33}$$

From the above equations one directly infers the main results from Rosch-Si[102]: the 'Grüneisen ratio' $\Gamma_r = \eta_r/C \sim T^{-y_r/z}$ in the quantum critical state while in the massless phase it becomes exactly $(d-y_0 z)/(y_0 y_r) r^{-1}$, i.e. it acquires an universal amplitude expressed entirely in terms of the exponents. The significance of the

coupling constant susceptibility χ_r is immediately clear from Eq.'s (3.32,3.33). Its temperature dependence reveals that it is more singular than η_r, which is in turn more singular than C. In addition, its temperature independent part diverges in the approach to the critical point with the exponent α_r, in direct analogy with the divergence of the specific heat with α in the approach to a thermal phase transition.

Let us now apply the above scaling laws to the specific context encountered in the cuprates. By restricting ourselves to thermodynamics we have to assume very little in addition to Eq. (3.30): (i) In the cuprates the relevant zero-temperature direction is the electron density varied by the doping p. The reduced coupling constant corresponds therefore with $x = (p - p_c)/p_c$ (ii) Recently, evidences have been accumulating showing that the overdoped state is a Fermi-liquid, characterized by $y_0 = 1$[114, 115, 118]. (iii) We rely on the specific heat as measured by Loram, Tallon and coworkers[121]. Since the superconductivity appears to hide the critical behavior, the regime of interest is at high temperature.

Given the assumption that electron density is the zero temperature control parameter it follows from elementary thermodynamics that the quantities η_r and χ_r relate to μ,

$$\begin{aligned} \eta_{cr,x} &= \frac{\partial S_{cr}}{\partial x}|_{\mu} = -\frac{\partial \mu}{\partial T}|_{x} , \\ \chi_{cr,x} &= \frac{\partial^2 F_{cr}}{\partial x^2}|_{\mu} = \frac{\partial \mu}{\partial x} = \frac{1}{n^2 \kappa} , \end{aligned} \tag{3.34}$$

where κ is just the electronic compressibility and n the total electron density. Notice that when pressure is the control parameter, $\chi \sim \partial^2 F/\partial p^2 \sim \partial V/\partial p$ refers to the total compressibility.

Let us now turn to the measured electronic specific heat of the cuprates[121]. In fact, the remarkable property of the measured specific heat is its uninteresting appearance. In the overdoped regime it is indistinguishable from the specific heat of a conventional BCS superconductor. At high temperatures, $C = \gamma T$ with a temperature independent γ like in a Fermi-liquid, and at the superconducting transition the specific heat shows a BCS-like anomaly. Upon decreasing doping, all what happens is that the pseudogap scale manifests itself quite clearly in the form of a decreasing γ, a fact exploited by Loram *et al.* to study the doping dependence of the pseudo gap temperature T_{PG}. Above T_{PG} γ is temperature independent and connected smoothly with the specific heat in the overdoped regime, showing no noticable doping dependence.

It seems to be a reflex to assume that the 'metallic' appearance of the γ above T_{PG} is just revealing that a Fermi-liquid state is re-established at high temperatures, but this is actually quite unreasonable. Recently, evidences has been accumulating that on the overdoped side a 'coherence' crossover occurs: one can identify a temperature T_{coh} below which transport shows Fermi-liquid signatures[115, 118] while photoemission reveals that the quasiparticles become underdamped[114]. T_{coh} emerges at optimal doping and increases with increasing doping in the overdoped regime. It is no wonder that the low temperature specific heat in this Fermi-

liquid regime is conventional, but why is it so that it remains conventional above T_{coh}? Stronger, why is it unaltered at temperatures $> T_{PG}$ even in the strongly underdoped regime?

Let us reconsider the scaling of the specific heat in the QC regime, Eq. (3.32). The remarkable fact is that its temperature dependence is predicted to be uninteresting! Its temperature exponent is just given by the ratio of the number of space- (d) and effective time (z) dimensions. In the quantum critical regime of the cuprates $C \sim T$ and this means that $d = z$, the number of space dimensions equals the number of time dimensions! At these high temperatures, it seems reasonable to assume that $d = 2$, with the implication that $z = 2$, signalling diffusion.

There is a non-trivial consistency with the observation that the specific heat is not sensitive to the crossover from the quantum critical- to the Fermi-liquid regime at T_{coh}. From Eq. (3.33) it follows that the specific heat in a massless state knows about the proximity of the QPT via the factor $r^{(2-\alpha_r)(d-y_0 z)/(d+z)}$, governing the divergence of the quasiparticle mass. The exponent contains the combination of the dimensions $d - y_0 z$ and when $d = z$ and $y_0 = 1$ as in the Fermi-liquid the exponent vanishes and the specific heat becomes insensitive to the zero temperature singularity! The specific heat is expected to be just the same at all temperatures and dopings as long as $T > T_{PG}$ despite the fact that other properties demonstrate large scale changes in the physics.

To further stress this point, let us consider what happens in the pseudo-gap regime $T < T_{PG}$. The measured specific heat shows that in between the superconducting T_c and T_{PG} $C \sim T^2$ and thermodynamically it can be viewed as a 'nodal liquid' characterized by $y_0 = 2$. Insisting that $d = z$ it follows from Eq. (3.33) that $C \sim r^{-(2-\alpha_r)/2} T^2$. From Eq.(3.21) it follows immediately that the pseudogap scale $T_{PG} \sim r^{z/y_r} = r^{(2-\alpha_r)/2}$; it just means that $C \sim T^2/T_{PG}$ which is consistent with experiment. Notice that this would fail when $d \neq z$.

Because α_r is expected to be small, T_{PG} is expected to be weakly sublinear in x when $d = z$. In a recent paper[122], the behavior of T_{PG} for small x has been determined in 123 samples where the superconductivity has been surpressed by Zn doping. T_{PG} turns out to be indeed weakly sublinear in x, suggesting that α_r is in the range $0.2 - 0.3$, i.e. a reasonable value for a strongly interacting unstable fixed point.

Up to this point we have presented the case that if a QPT is present at optimal doping, the quantum singularity is largely hidden from the specific heat for specific reasons ($d = z$, the Fermi-liquid). To establish the presence of this singularity one has to look elsewhere and the remedy is obvious: the thermodynamic potential. Assuming $d = z$ one finds an interesting collection of scaling behaviors for $\partial\mu/\partial T$ and the inverse compressibility χ_x.

Omitting non-universal factors and including the specific heat for completeness, these become in the quantum-critical regime,

$$C_{cr} \sim T; \ \left(\frac{\partial \mu}{\partial T}\right)_{cr} \sim -T^{-\frac{\alpha_r}{1-\alpha_r}}; \ \chi_{cr,x} \sim -T^{-\frac{2\alpha_r}{1-\alpha_r}}. \qquad (3.35)$$

Hence, by measuring the temperature dependences of the chemical potential and the compressibility in the *high temperature* quantum critical regime one obtains directly the 'quantum alpha' characterizing the nature of the quantum singularity. Notice that the incompressibility should be precisely twice as singular as the temperature derivative of μ.

The Fermi-liquid regime ($y_0 = 1$) is not particularly revealing,

$$C_{FL,cr} \sim T; \ \frac{\partial \mu}{\partial T}_{FL,cr} = 0; \ \chi_{FL,cr,x} \sim x^{-\alpha_r} . \tag{3.36}$$

The critical part of $\partial\mu/\partial T$ vanishes because the prefactor contains $d - y_0 z$ as does the temperature dependent part of χ_n. Only the temperature independent part of the inverse compressibility reveals directly the quantum singularity.

In the pseudogap regime ($y_0 = 2$) this changes drastically. Parametrizing matters in terms of the pseudogap scale $T_{PG}(n) \sim x^{(2-\alpha_r)/2}$,

$$\begin{aligned} C_{PG,cr} &\sim \frac{T^2}{T_{PG}(x)}; \ \frac{\partial \mu}{\partial T}_{PG,cr} = \frac{C_{PG,cr}}{x}; \\ \chi_{SP,cr,x} &\sim Ax^{-\alpha_r} - B\frac{C_{SP,cr}T}{x^2} . \end{aligned} \tag{3.37}$$

The second and third line reflect the workings of the 'generalized Grueneisen parameters'. $\partial\mu/\partial T$ is clearly 'one order more singular' in n than the specific heat, but the temperature dependent part of the incompressibility is actually 'twice as singular'.

3.6 Conclusions

Using elementary power counting arguments we have discovered an emperical strategy which should make it possible to decide if the 'quantum criticality' of the cuprates has to do with universality. We have found that the temperature- and density dependences of the chemical potential in the various regimes should obey *six* scaling laws, which are all governed by a single fundamental scaling dimension (α_r). As an input, we have used the specific heat data to argue that the effective number of space (d) and time (z) dimensions characterizing the critical state have to be the same.

We are not aware of chemical potential- and electronic compressibility measurements of the cuprates having the required accuracy. However, this does not appear to represent a problem of principle. In the experimental literature one finds a variety of methods to measure these quantities[123, 124], a prime example being the vibrating Kelvin probe method allowing for high accuracy measurements of the chemical potential, used by van der Marel and coworkers sometime ago to determine the density dependence of the superconducting T_c[125]. Using these experimental methods one might establish once and for all the presence or absence of a genuine quantum phase transition in the cuprates.

It is very interesting to note that scaling relations derived in this chapter might also be put to the test in the context of the metal-insulator transition in the two dimensional electron gas, which is going to be the main focus for the rest of this thesis. In the last chapter we are going to adress the properties of the electronic compressibility which has been measured in the proximity of this quantum phase transition.

Chapter 4

Thermodynamics of The 2D Electron System

As emphasized earlier in the introduction chapter, almost all of what understood regarding electron/hole dynamics of 2D electron systems in MOSFETs or GaAs/AlGaAs comes from transport experiments. In this chapter we will review what is known experimentally on the thermodynamic properties of 2D electron system near MIT, in more detail. This is very important because theoretically, within the framework of Fermi Liquid, one does not expect any qualitative change in the thermodynamic properties [1]. However, recent theories for strong interacting systems have predicted that there should be profound consequences in thermodynamic measurements [36, 34]. Therefore it is very clear that if a zero temperature phase transition exist, it may reveal itself also in the such quantities like the electronic compressibility, $\kappa^{-1} = n^2 \frac{\delta\mu}{\delta n}$ (where μ is the chemical potential). In fact, very little is known regarding the thermodynamic behavior of 2D electron and hole systems. Compressibility reflects the ability of an electronic system to screen external charges. When an external charge is brought in close proximity of the surface of a metal, mobile electrons accumulate in its vicinity to screen it. A system that cannot redistribute its charge, is therefore incompressible. The electronic compressibility, therefore, provides a primary tool to understand the type of the ground state that is formed by electrons. Recent experiments show that thermodynamic behavior, namely electronic compressibility, exhibit an unusual change cross the MIT point in 2D electron/hole systems. In this chapter we are going to address the results of these experiments in more detail.

4.1 Bulk compressibility

The compressibility or $\frac{\delta\mu}{\delta n}$ of a system reflects how its chemical potential varies with density. For non-interacting fermions it simply amounts to the single-particle Density of States (DOS), which in 2D is density independent ($\delta\mu/\delta n = \pi\hbar^2/m$). This picture, however, changes drastically when interactions are included. Ex-

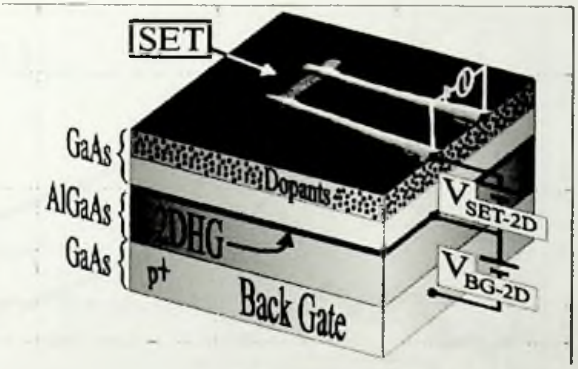


Figure 4.1: An SEM micro-graph of the SET, placed on top of a schematic structure along with the measurement circuit (from Ilani *et al.* [14])

change and correlation effects weaken the repulsion between the electrons, thereby reducing the energy cost, thus leading to negative and singular corrections to $\frac{\delta\mu}{\delta n}$. Within the Hartree-Fock (HF) theory, which includes both the DOS and exchange terms, one finds:

$$\frac{\delta\mu}{\delta n} = \frac{\pi\hbar^2}{m} - \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{e^2}{4\pi\varepsilon} n^{-1/2} \tag{4.1}$$

with the compressibility becoming negative at low enough densities. Measurements of the macroscopic compressibility of 2D electron/hole gases, in the metallic regime, have indeed confirmed this behavior [130, 129]. To date there are two measurements of compressibility have been reported by two different experimental groups.

Ilani *et al.* [14] expanded the study of local chemical potential, $\mu(n)$, of the 2D hole gas (2DHG) into the MIT regime using Single Electron Transistors (SET). Their measurements utilize several Single Electron Transistors (SETs), situated directly above a Two-Dimensional Hole Gas (2DHG). This technique allows them to determine the local behavior of $\mu(n)$ as well as its spatial variations. Simultaneous macroscopic transport measurements were conducted to ensure a precise determination of the MIT critical density in the same sample. Their measurements indicate a clear thermodynamic change in $\mu(n)$ at the MIT. In addition, they find that the behavior of $\mu(n)$ in the metallic phase follows the HF model and is spatially homogeneous. The insulating phase, on the other hand, is found to be spatially inhomogeneous. In this regime, $\frac{\delta\mu}{\delta n}$ deviates by more than an order of magnitude from the predictions of the HF theory.

A local measurement of μ requires the ability to measure electrostatic potentials with high sensitivity and good spatial resolution. It has been previously shown that both can be achieved using an SET [133, 132, 131]. Ilani *et al.* have therefore deposited several aluminum SETs on top of the inverted GaAs/AlGaAs hetero-

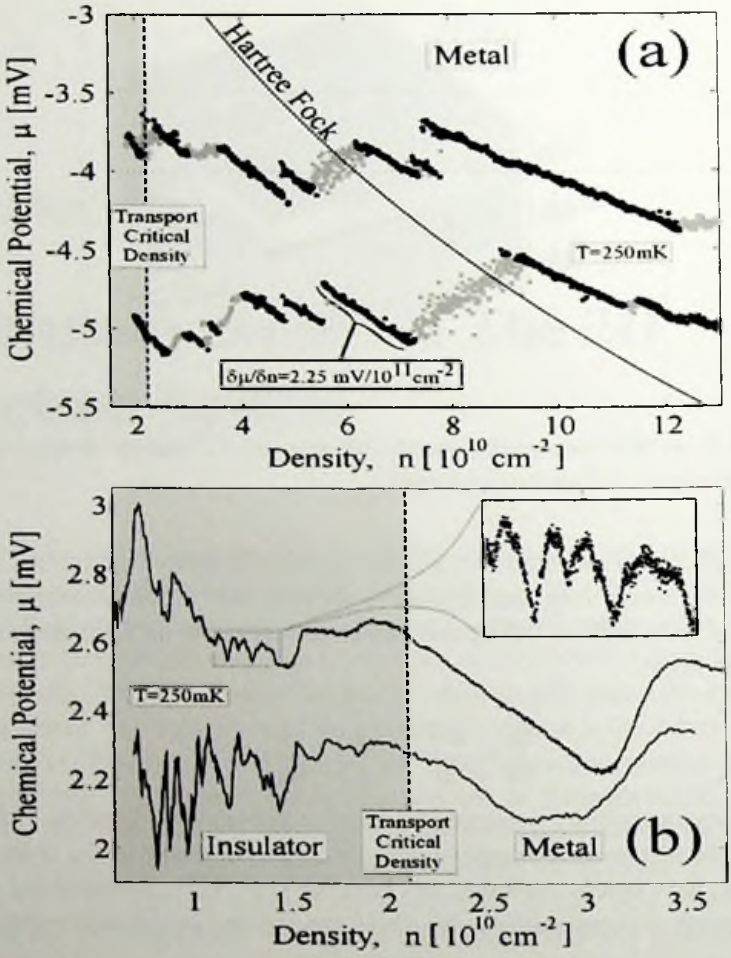


Figure 4.2: (a) $\mu(n)$ in the metallic regime. The HF predictions are depicted by the solid line. The negative-slopes are high lighted (dark symbols) to demonstrate their resemblance to HF model. (b) $\mu(n)$ across the MIT and in the insulating region. Inset: A closer look at the data in the insulating regime. Each slope is composed of many data points allowing for accurate determination of slopes (from Ilani *et al.* [14]).

structure (see Fig. 4.1). This configuration allows them to study simultaneously the local thermodynamic behavior and the macroscopic transport properties and, thereby, trace possible correlation between them. This technique relies on the fact that μ of the 2DHG changes as its density is varied. At equilibrium, the Fermi energy is constant across the sample, and therefore, a change in μ induces a change in the electrostatic potential, which is directly measured by the SET. The size of the SET and its distance from the 2DHG determine its spatial resolution, estimated to be $0.1 \times 0.5 \mu m^2$. They could measure μ by measuring the local electrostatic

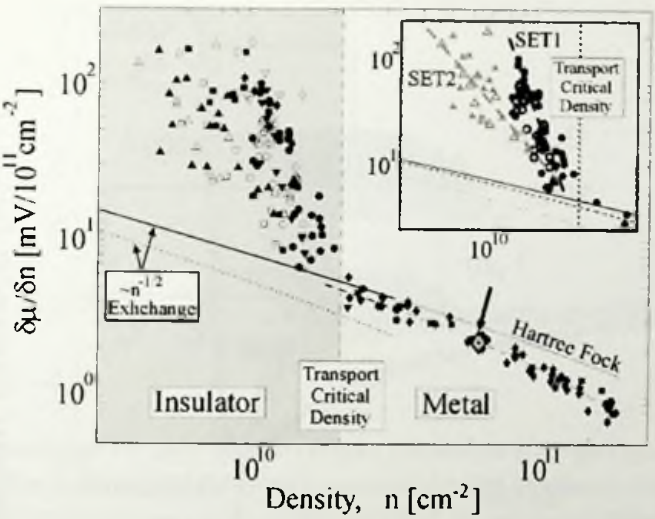


Figure 4.3: $\delta\mu/\delta n$ collected from five SETs on three different hall bars from two different wafers. In the insulating regime, both negative and positive slopes are shown (closed and open symbols respectively). Each point corresponds to a well-defined segment in the $\mu(n)$ trace. The points marked by arrows correspond to the marked segment in Fig.4.2 (a). Inset: Results from two different SETs on the same device demonstrating the spatial dependence 'of $\delta\mu/\delta n$ in the insulating side (from Ilani *et al.* [14]).

potential with a very high sensitivity and good spatial resolution. The results of the experiment are shown in Fig. 4.2 and Fig. 4.3. Several typical traces of μ vs. n are shown in metallic, (Fig. 4.2a), as well as insulating, (Fig. 4.2b), regime across the MIT. Instead of the expected monotonic dependence, μ exhibits a rich structure of oscillations. The qualitative change in the oscillation pattern of $\mu(n)$ near the critical density suggests that the system experiences a thermodynamic change at the MIT.

In Fig. 4.2a the measured oscillations has been compared with the expected monotonic behavior of the HF theory. In contrast to the theory, the measured μ has a density independent average which suggests that some kind of screening mechanism is present in the system. While the negative slopes of $\mu(n)$ (black symbols) follow the HF theory, there are apparent additional drops (some of which are extremely sharp) between them (gray symbols). This saw-tooth profile is reminiscent of the behavior of the chemical potential of a quantum dot as a function of its density [134] and suggests the existence of discrete charging events in the dopant layer, situated between the 2DHG and the SET. Thus, the measured μ of the 2DHG varies undisturbed along the negative slopes, until a certain bias is built between the 2DHG and the SET that makes the charging of an intermediate localized state energetically favorable. This causes a sharp drop in the measured electrostatic potential, after which μ continues to vary undisturbed until the next screening event occurs.

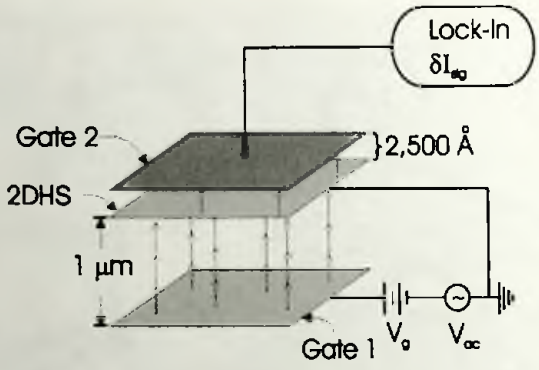


Figure 4.4: (a) Diagram of experimental setup of Dultz and Jiang. (b) Typical trace of C_q which is proportional to κ, and (c) R_s which is a measure of dissipation as a function of applied DC voltage at three different AC frequencies: 43.2, 100, and 200 Hz. All data was taken at $B = 0$ and $T = 4.2\ K$ (from Dultz and Jiang [13]).

Because screening occurs only at discrete points one can reconstruct the underlying $\mu(n)$ from the unscreened segments in the measurement. This has been shown in Fig. 4.3. Each single point in this graph represent the slope of a well-defined segment, like the ones emphasized in Fig 4.2a. In the metallic side, the data collapse on a single curve in a good agreement with Hartree-Fock (HF) theory within a factor of two over more than an order of magnitude in density. On the other hand $\partial\mu/\partial n$ starts deviating from HF in the close proximity to the transport-measured critical density, signifying a change in the screening properties of the 2DHG at the transition. They also performed different measurements with different location of the SETs on the sample and they found out that in the metallic regime the results are essentially the same suggesting that this phase is *homogeneous in space.*

The sudden appearance of rapid oscillations as the system crosses to the insulating side already signifies that something dramatic happens at the transition. One can see from Fig. 4.3 that $\partial\mu/\partial n$ starts deviating from the theoretical curve in close proximity to the transport-measured critical density, signifying a change in the screening properties of the 2DHG at the transition. Another intriguing result is the non-universal behavior of the slopes on the insulating side. It should be emphasized that each slope is determined from a complete segment in the $\mu(n)$ curve (see inset to Fig. 4.2b) and is, therefore, determined very accurately. The fluctuations in the slopes, seen even in a measurement done using a single SET, are completely reproducible and suggest that mesoscopic effects are present. Furthermore, the average behavior of $\partial\mu/\partial n$ in this fluctuating regime is seen to be position dependent. An example of that is shown in the inset in Fig. 4.3 where we plot $\partial\mu/\partial n$ measured by two separate SETs on the same Hall bar. Such dependence on position indicates that once the system crosses into the insulating phase it *ceases to be spatially homogeneous.*

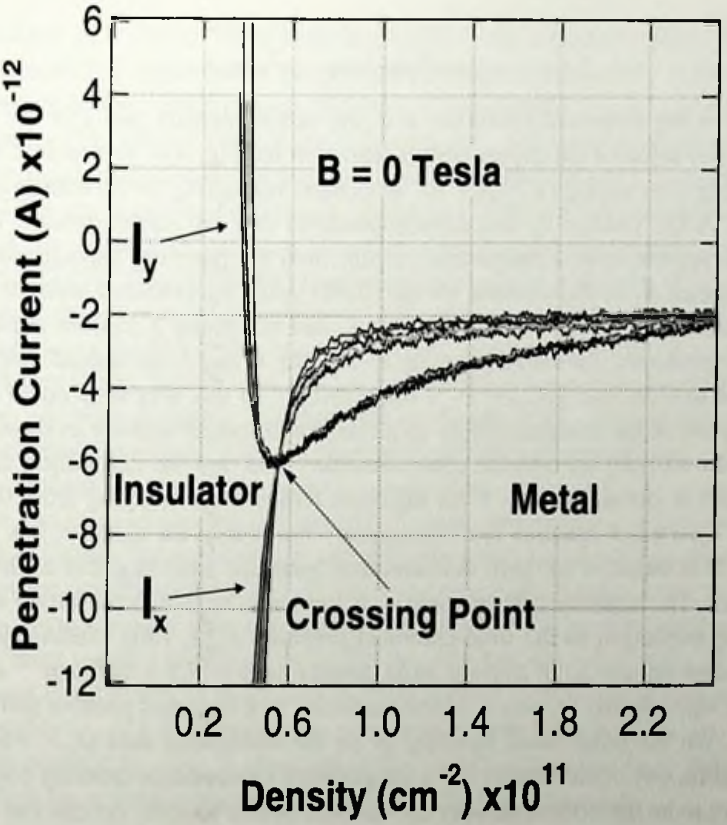


Figure 4.5: $I_x \propto -R_s$ (dissipation of the 2DHS) and $I_y \propto \kappa^{-1}$(inverse compressibility of 2DHS) vs density for five temperature at an excitation frequency of 100 Hz: blue- .33 K, green- 0.56 K, black- 0.82 K, orange- 1.02 K, red- 1.28 K. The crossing point of the fie dissipation channel curves corresponds to the metal-insulator phase transition at $B = 0$. The minimum in the inverse compressibility channel occurs at the same density of $p = 5.5 \times 10^{10}\ cm^2$ (from Dultz and Jiang [13]).

Second experiment comes performed by Dultz and Jiang [13]. They took a different method to map the dependence of the thermodynamic density of states (or equivalently, the compressibility) of a strongly interacting two-dimensional hole system (2DHS). To do this the capacitance between the the 2D electrons and the gate is measured. This capacitance can be modeled as the geometrical capacitance in series with a quantum capacitance. The quantum capacitance per unit area c_q is related to the density of states (DOS), $\delta\mu/\delta n$, or compressibility, κ, by

$$c_q = e^2(\delta\mu/\delta n) = n^2e^2\kappa \tag{4.2}$$

where μ is the chemical potential, n is the carrier density and e is the electron charge. The setup of the experiment is shown in the Fig. 4.4. To measure the compressibility, they applied a 10 mV AC excitation voltage V_{ac} to the bottom electrode (Gate1). A DC voltage V_g was superimposed to vary the carrier density. By modeling the system to as a distributed circuit, both the quantum capacitance C_q and the resistance R_s of the channel for the 2DHS could be extracted separately based on the measured values for the in-phase I_x and 90^o phase I_y current components. In this experiment, for frequencies up to 200 Hz, I_x and I_y are indeed directly proportional to $-R_s$ and $1/C_q(\propto \kappa^{-1})$ respectively. In this way they could map out the behavior of the compressibility as a function of carrier density as shown in Fig. 4.5. It shows both the inverse compressibility and and the dissipation signals as a function of density at $B = 0$ for different temperatures ranging from 0.33 K to 1.28 K. Two main features are immediately noticeable for the $I_y \propto 1/\kappa$ channel. First, $1/\kappa$ is negative for high densities and becomes more negative as the density decreases. The negative compressibility is known to be due to the strong exchange energy contribution to the total chemical potential [129, 130]. Secondly, a sharp turn-around occurs at an average hole density of $p = 5.5 \times 10^{10}\ cm^{-2}$ as $\delta\kappa/\delta p$ changes sign. As the density is further reduced, $1/\kappa$ becomes positive and diverges rapidly. On the other hand focusing in on the dissipation data ($I_x \propto -R_s$) in the same figure, one could clearly see a temperature independent crossing point which is believe to be the critical density for the MIT in this sample. As one can see from Fig. 4.5, Dultz and Jiang argued that since the dissipation ($I_x \propto R_s$) is becoming more negative as temperature increases, the 2DHS become more resistive which is a signature of metallic behavior. The opposite is seen on the low density side of the crossing where the characteristic temperature dependence is that of an insulator. The crossing point occurs at the point where $\partial\kappa/\partial p$ changes sign precisely at the minimum of $1/\kappa$. Therefore, they concluded that there is a clear signature of metal-insulator phase transition at $B = 0$ in their thermodynamic measurement as the holes are localized right at the point where $1/\kappa$ is at minimum. They also concluded that they can not explain their observation theoretically based on non-interacting models.

As in the hole system of previous experiment the Coulomb interaction is very strong, Dultz *et al.* [137] also measured thermodynamic compressibility of moderately interacting two-dimensional electron system (2DES), using exactly the same field-penetration method. In previous (above) mentioned experiment, the 2DHS has a rather strong correlation due to the heavy effective mass. The dimensionless parameter r_s which measures the ratio of the correlation energy to the Fermi energy ranges between 5 to 25 and the qualitative change in the compressibility was observed at r_s=13. To understand whether this behavior is specific to strongly correlated systems and how it is linked to the 2D metal-insulator transition, they studied another two-dimensional electron system (2DES) where interactions are

much weaker, r_s = 0.5-3. There is no transition in the transport. They found that the compressibility behavior in the present system is qualitatively different from that observed in the 2DHS system. Its density dependence can be well described by the theory [31, 138] of percolating droplet states, discussed in chapter 5.

They found important differences between their measurements in weakly interacting 2DES and those in the strongly interacting 2DHS. First, the resistive component of the signal does not have a temperature independent crossing point like that for the 2DHS. This should be expected from a sample with only weak to strong localization crossover behavior as shown in the transport measurements. The transport properties of the 2DES can be understood in terms of the more "conventional" scaling theory [139]. Secondly, dμ/dn for the 2DES gas is positive everywhere (see Fig. 4.6). At high densities, this dependence is expected. As one can see from the quantitative agreement with the Hartree-Folk (HF) approximation calculation for low r_s values [4]. Since the effective mass of electrons in GaAs is only 0.067 times the rest mass of an electron (roughly 5 times less than the effective mass of the holes in GaAs), the maximum r_s value of this particular system is only about 3 for the lowest density studied here. Positive compressibility means that the kinetic energy is always greater than the exchange energy for the present system. Third, although there is still a minimum in $\partial\mu/\partial n$, just like that observed in the 2DHS, the turn-around is not nearly as sharp as is the case for that system and occurs at a density of roughly (r_s = 1.4). Finally, the temperature dependence of $\partial\mu/\partial n$ is rather strong, in contrast to what is observed in the 2DHS where $\partial\mu/\partial n$ is temperature independent for a large range of temperatures from 0.3 K to 4.2 K.

4.2 Local Compressibility

As has been discussed before the compressibility reflects the ability of an electronic system to screen external charges. When an external charge is brought in close proximity to the surface of a metal, mobile electrons accumulate in its vicinity to screen it. A system that cannot redistribute its charge is, hence, incompressible. The electronic compressibility, therefore, provides a primary tool to understand the type of ground state that is formed by electrons. Ilani *et. al.* used a single electron transistor (SET) to measure the local compressibility of a two-dimensional hole gas (2DHG) across the MIT and looked into the microscopic structures of this quantity in more detail. The transport properties of these structures have been studied extensively, and a clear MIT phenomenology is observed. An aluminum SET positioned on the surface of the structure is used to measure the local chemical potential, μ, as well as the local compressibility, $\kappa = n^2[\delta n/\delta\mu]$ (n being the average density of holes), at dilution fridge temperatures (50 mK).

Looking at the behavior of μ as a function of the hole density (Fig. 4.7A), it seems that starting from the metallic side, μ increases as the density is lowered–a behavior that corresponds to negative compressibility [130, 129]. This behavior persists down to the critical density, n_c, which is determined by transport measure-

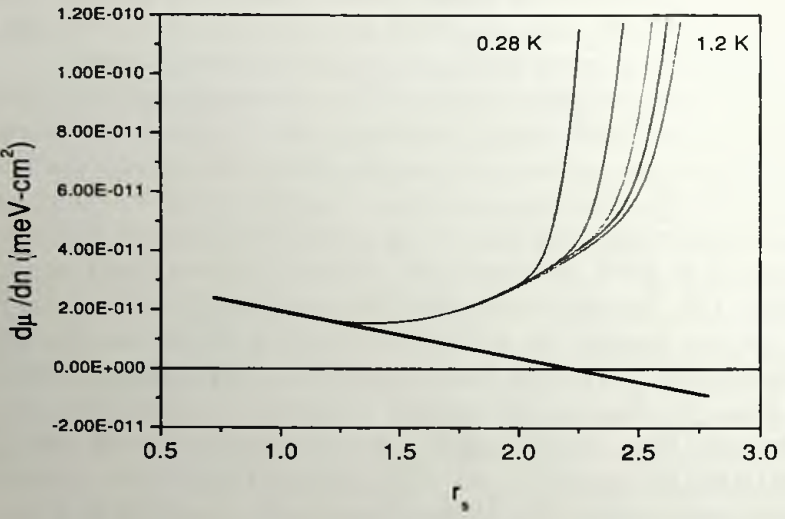


Figure 4.6: $d\mu/dn$is plotted against r_s. The solid line is calculated using HF approximation.(from Dultz *et al.* [137])

ments. When the density is further lowered, μ starts to decrease rapidly, indicating a positive and vanishing compressibility. This change of sign of the compressibility constitutes the macroscopic thermodynamic signature of the MIT.

The microscopic behavior emerges as an intricate fine structure in the compressibility that is superimposed on the average macroscopic behavior. Two generic types of microscopic structures are apparent in the measured signal (Fig. 4.7). The larger structure is a staircase decrease of μ on the insulator side (arrows in Fig. 4.7A). However, a much finer structure appears when we directly measure the derivative of μ with respect to the back-gate voltage, $\delta\mu/\delta V_{BG}$ (Fig. 4.7B). This structure assumes the form of very sharp spikes that are superimposed on a smooth background. Because the back-gate voltage controls the average density, $\delta\mu/\delta V_{BG}$ is a direct measure of the local inverse compressibility, $n^2\delta\mu/\delta n$. The emergent spikes in $\delta\mu/\delta V_{BG}$ correspond to microscopic processes that happen in the 2DHG near the MIT and their spectrum captures the fingerprint of these processes. These spikes commence already in the metallic phase, where a relatively sparse spectrum of negative spikes is observed (Fig. 4.7C). As the density is lowered across the transition, this spectrum becomes richer–both the number of spikes and their amplitudes increase substantially (Fig. 4.7D). Once deep in the insulating side, the spectrum is made up of many positive and negative overlapping spikes. The ap-

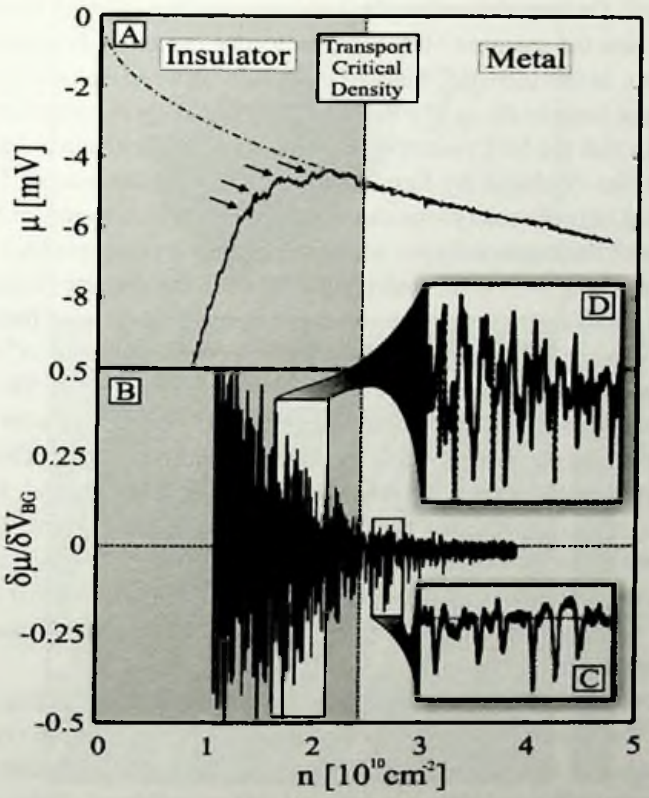


Figure 4.7: (A) Density dependence of the local chemical potential, μ, across the MIT. By monitoring the current through the SET, one can deduce the local electrostatic potential, which directly determines the local μ, given that the sample is in equilibrium (4). A transport measurement on the same device determines the MIT critical density, nc (dashed line). Above this density, $\mu(n)$follows the prediction of the Hartree-Fock theory multiplied by 1.8 (dashed-dotted line). A rich reproducible microscopic structure is observed when we measure the derivative with respect to the back-gate voltage, $\delta\mu/\delta V_{BG}$ (B). This structure assumes the form of sharp spikes (a prototype spike is marked in gray), which emerge already on the metallic side (C) but become more dense and intense as the system crosses to the insulator side (D). Identical features of the spikes spectrum have been observed in measurements of several other devices. In each of these devices, a different spike spectrum was observed, being a fingerprint of the specific local disorder.(from Ilani *et al.* [15])

pearance of spikes superimposed on the continuous $\delta\mu/\delta V_{BG}$ signal already hints at the presence of a new phase that develops as we approach the MIT.

A spike in the local compressibility signifies a discrete screening event. Such a process is rather unexpected if one assumes that the 2DHG can smoothly redistribute its charge. However, within various theories of the MIT, there are several fundamental processes that might break the smooth screening capability on the mi-

croscopic level. One set of theories [44, 140] suggests that strongly localized states must appear near the apparent MIT. Each such state can either be empty or hold a single electron. In this scenario, the sharp transition between the two situations will cause an abrupt jump in the local μ and a spike in $\delta\mu/\delta V_{BG}$. Another set of theories [31, 32] claim that the MIT results from percolation in the electron liquid. In this scenario, puddles of charge are formed in the percolation landscape. The puddles can be charged only discretely–one electron at a time–which results in having only a discrete set of back-gate voltages where the system is compressible. Because in both scenarios the system is incompressible between the discrete charging events, the spikes in $\delta\mu/\delta V_{BG}$ are expected to be superimposed on an incompressible background ($\delta\mu/\delta V_{BG} = 1$). This is not observed experimentally (Fig. 1C), where the observed background has very high compressibility ($\delta\mu/\delta V_{BG} \ll 1$). Ilani *et al.* suggested that there might be the coexistence of two phases: one with continuous response producing the compressible background and one with discrete response producing the spikes–namely, a two-liquid behavior. One can see (Fig. 4.7D) that as the system deepens into the insulating regime, the background $\delta\mu/\delta V_{BG}$ increases toward the incompressible value ($\delta\mu/\delta V_{BG}$= 1), whereas the number of spikes and their amplitudes increase. This behavior demonstrates that as the density of holes is reduced, the discrete phase grows at the expense of the coexisting compressible phase.

The above models are only suggested candidates for the nature of the discrete phase, and clearly, more complicated models, such as electron crystallization [36, 48, 4], should also be considered. However, the essential feature that all the possible descriptions would share is the local reorganization of charge, a necessary condition for the appearance of spikes. This locality implies that the onset of spikes must be accompanied by a spatial structure.

In their experiment Ilani *et. al.* demonstrated how this spatial structure is resolved experimentally even with a detector that is fixed at a single point. The idea is to use the SET not only as an electrometer but also as a top gate. In contrast to the back gate, which affects the density uniformly in space, the top gate influences the density in a manner proportional to the distance from its edge. A spike in Fig. 4.7 occurs when a charging event crosses the Fermi energy. This happens at a certain local density. Changing the back-gate voltage by V_{BG} shifts this event from the Fermi energy. However, one can restore the local density and hence shift the event back to the Fermi energy by applying a compensating top-gate voltage V_{TG}. The further away the event is from the SET, the larger is the V_{TG} that is required in order to shift it back. Each charging event is, therefore, characterized by a "charging line" in the plane of $[V_{BG}, V_{TG}]$ (Fig. 4.8). The slope of the charging line, $s = (\delta V_{TG}/\delta V_{BG})$, determines the position of the event in the 2DHG: An event happening just below the SET would have the smallest slope, s_{min}. The further away the event is from the SET, the more vertical its slope will be. The value of s_{min} can be estimated from the ratio of capacitances between the 2DHG and the top and back gates to be $s_{min} = 2.3$.

Figure 4.8 shows how the overlapping spectrum of Fig. 4.7D disentangles

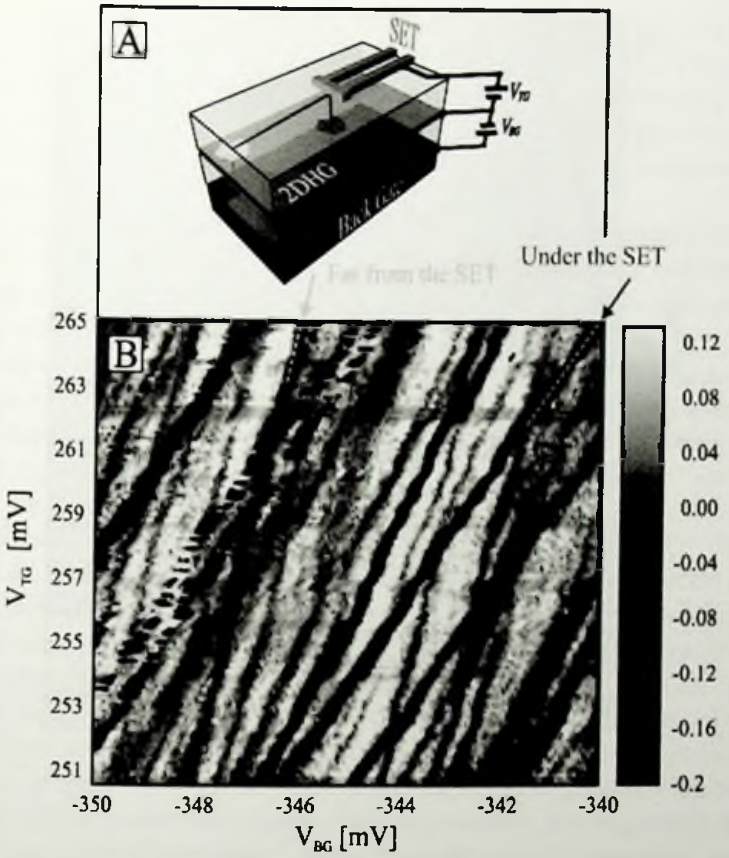


Figure 4.8: (A) Scheme of the measurement circuit, which is used to identify the position of events within the 2DHG. (B) Color map representing the measured value of the derivative $\delta\mu/\delta V_{BG}$ as function of the back-gate voltage, V_{BG}, and top-gate voltage, V_{TG}. Each discrete event in the 2DHG gives rise to a line in this plot whose slope reflects the distance of the event from the SET . Arrows mark the lines that correspond to two typical events, one that occurs directly under the SET and one that occurs far from the SET (from Ilani *et al.* [15]).

into separate events, each having a different slope and amplitude. The span of slopes reflects the different positions of the events in the 2DHG. The smallest slope observed is $s_{min}^{exp} = 2.26$, which corresponds to an event that occurs just under the SET. The largest slope observed, $s_{max}^{exp} = 6.9$, corresponds to an event occurring $1.1\mu m$ away from the SET and defines the measurement horizon. This horizon is in good agreement with what is expected for a SET situated 400 *nm* above the 2DHG. A clear correlation can be seen between the position of an event and its amplitude–the closer it is to the SET, the stronger it appears. As the slopes can be measured rather accurately, we are able to position the events within the 2DHG

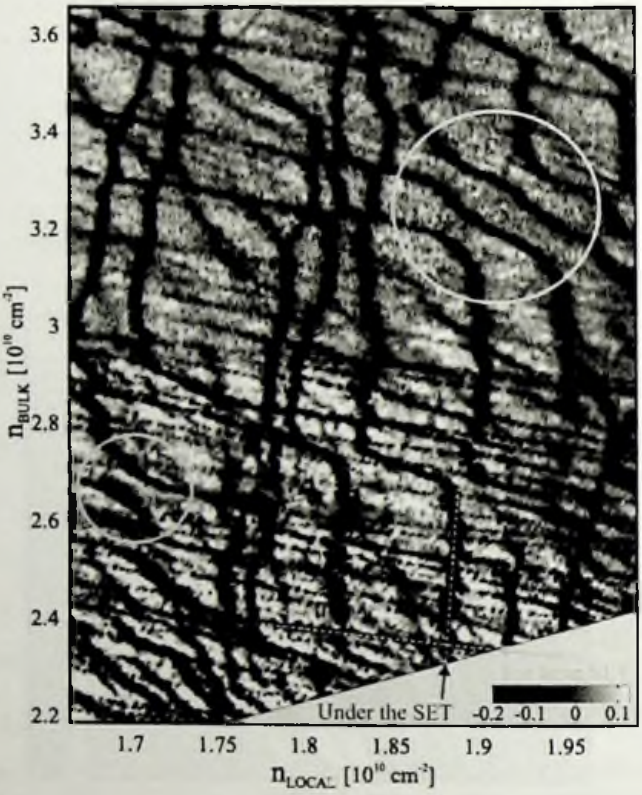


Figure 4.9: Evolution of discrete events in response to an induced density gradient. Here, as in Fig. 4.8, the derivative μ/V_{BG} is measured as function of V_{BG} and V_{TG}; however, now substantially higher voltages are applied so the density profile in the 2DHG is modified. The coordinate system is rotated with respect to Fig. 4.8: The horizontal axis corresponds to the density just below the SET, n_{local}, and the vertical axis corresponds to the density far away from the SET, n_{bulk}. In this coordinate system, vertical (horizontal) charging lines correspond to local (distant) charging events. The circles mark avoided crossing between the charging lines that demonstrate the existence of tunneling amplitude between relatively distant charge configurations (from Ilani *et al.* [15]).

with an accuracy of up to ~20 *nm*.

This microscopic measurements, therefore, reveal that as the density of holes approaches the critical density from the metallic side, the 2DHG fragments into localized charge configurations that are distributed in space. The rather abrupt appearance of these events near n_c suggests that this fragmentation is the microscopic signature of the MIT. However, their simultaneous appearance in large numbers masks the ability to determine their size, mutual interactions, and the quantum mechanical probability for charge tunneling among the different configurations. To unravel these properties, Ilani *et al.* use the SET (top gate) to form a density

gradient in its vicinity. As we increase the bulk density using the back gate, we reduce it locally using the SET top gate. This maintains the local density under the SET at a constant value and generates a density gradient surrounding the SET. Because the spikes appear only at densities that are comparable to or smaller than n_c, we can continuously reduce the area from which the spikes appear and hence their number.

The result of such a measurement is shown in Fig. 4.9. The result is ploted in a coordinate system that is rotated to reflect the independent control over the density under the SET, n_{local}, and far away from the SET, n_{bulk}. In this plot, horizontal charging lines correspond to events that occur very far from the SET, and vertical charging lines correspond to local events, occurring directly under the SET. As we go from the bottom of the figure to its top, we continuously evolve from a homogeneous system, namely, $n_{bulk} \approx n_{local}$, to a system with a large density gradient, $n_{bulk} >> n_{local}$. Noticeably, distant events (horizontal lines) are affected by n_{bulk}, and as this density exceeds n_c, most of them disappear rather abruptly, reflecting once more the fragmentation of the 2D system at the MIT. At the top of the scan, we are, therefore, left with only a few local charging events (vertical lines), which allows to focus on the nature of the interactions between them. The coupling among the different charge configurations is apparent where the corresponding charging lines cross. Many encounters of avoided crossings are apparent, implying the existence of tunneling between the corresponding events. Encircled in this figure are large tunneling gaps that connect configurations whose spatial separation exceeds 1 μm. As only nearly overlapping configurations are mixed by tunneling (which is otherwise exponentially suppressed), this suggests that certain configurations extend more than 1 μm in space–an order of magnitude more than the electron wavelength at this density.

4.3 Conclusions

Almost all our understanding of MIT in 2D electron/hole systems in MOSFETs comes from transport properties. Therefore it is very crucial to see whether this phenomena is the generic characteristic of 2D electrons/holes or related to the special detailed properties of such systems. To address this question fundamentally, one can probe the thermodynamic behavior of such system near MIT critical point. One of the most important quantities which reflects the universality of the $T = 0$ quantum phase transition is the 2D electron systems is the electronic compressibility. This quantity reflects directly the ability of electrons/holes to screen external charges, makes it possible to distinguish between two fundamentally different phases either side of MIT point . Recent experimental developments have already made it possible to measure such important quantity both in bulk and local scales. As a results of such experiments, which we revealed in this chapter, it is now understood that 2D electron systems undergoes a clear thermodynamical change across MIT critical point. This reflects itself in an unusual change in the

behavior of inverse compressibility, κ^{-1}.

Electric field penetration experiments show that by decreasing the density, inverse compressibility, κ^{-1}, reaches a minimum around $r_s^c \sim 13$ in highly mobile hole systems. This happens around the same critical density coming from transport measurement in these samples. Further decrease in density shows that this quantity (κ^{-1}) is turning upward toward positive values, crossing the zero point (singular compressibility) and diverges very strongly. Same kind of experiments on low mobility electron systems show qualitatively the same behavior but not as sharp and as strong as has been seen in hole systems. Different set of experiments based on Single Electron Transistor (SET) could probe the local properties on electronic compressibility in 2D hole systems. Such experiments determine the local behavior of chemical potential $\mu(n)$ as well as it's spatial variations. It has been shown that the behavior of $\mu(n)$ in the metallic phase follows the theoretical predictions based on HF theory and the system in spatially homogeneous in this regime. The insulating phase, on the other hand, is found to be spatially inhomogeneous. In this regime, the quantity $\partial\mu/\partial n$ which is proportional to inverse compressibility, deviates very clearly from the predictions of HF theory.

A very fundamental reason for such behavior of the electronic compressibility and its relationship with the local properties of electron/hole system is, in fact, a central focus of this thesis. Later in next chapters, by adapting a correct physics of unscreened Coulomb interaction at low densities within density functional theory we will try to explain some the experimental observations summarized here.

Chapter 5

DFT-LDA and The Droplet Picture

The experimental findings discussed in the previous chapter are very important, at least, in two respects : first, they show that the existence of metal-insulator transition in 2D electron system in MOSFETs or GaAs/AlGaAs heterostructures is clearly linked to the very fundamental changes in thermodynamics of such system across MIT critical point. This reveals itself very clearly in the behavior of the chemical potential, μ, and its derivative with respect to the density, corresponding with the inverse of electronic compressibility κ^{-1}. Second, they strongly challenge theoretical attempts for explaining such unusual behavior as most of the current theoretical explanations for MIT phenomena in 2D systems are preoccupied entirely with transport properties.

Focussing on the compressibility, it is very clear that any complete theory aimed at explaining the behavior of this quantity has to incorporate the nature of the two phases at both side of the transition and the physical process linking these two phases. So far there is no such theory ia available in the theoretical market of MIT phenomena. Although Hartree-Fock theory can easily explain the incompressible phase of Wigner crystals[71], it fails to produce, even qualitatively, anything like the incompressibility data seen in the experiments. According to this theory the electronic compressibility of 2D system is always negative and has to decrease monotonically by decreasing the density of charge carriers[129, 130, 4]. On the other hand, although Finkelstein theory (for two valley system) has a metal insulator transition in transport properties (RG calculation of conductance)[47], it does not address anything related to the observed thermodynamical changes near the transition point. In fact in this theory the compressibility, related to the interaction in the singlet channel is always assumed to be constant. We are going to address this in more detail in the next chapter where we introduce the Si-Varma physics[34].

There also have been suggestions that the observed MIT in 2D system might be related to percolation physics [31]. Based on this school of thought, the metallic

phase is an inhomogeneous state consisting of a conducting liquid phase separated by regions of an insulating, and low-density "vapor" phase. These insulating islands (droplets) percolate through conducting phase when the density of charge carriers is decreased. Accordingly, metal-insulator transition occurs at the percolation threshold. Explicit calculations based on this theory [32] (droplet picture) show that the drop of metallic resistance, $\rho(T)$, is no more than a factor of two, while experimentally it exceeds a factor of ten in Si-MOSFETs. In fact, these percolation pictures of MIT requires the metallic state to be inhomogeneous on some scale. As discussed in previous chapter, the experimental results of Ilani *et al.* [14] seem to indicate an inhomogeneous insulating phase, but a homogeneous metallic one. Resting on a local density approximation (LDA) treatment of the effects of the electron-electron interactions, Shi and Xie implemented the idea of droplet picture and looked into the behavior of the electronic compressibility[138]. Although pure LDA fails to explain the sharp turn-up in the compressibility behavior near the MIT (see sections 5.3 and 5.4), but by employing the somewhat artificial device of a strongly correlated potential, Shi and Xie showed that the inverse compressibility of 2D electron system exhibits a minimum and turns upward around the percolation threshold density.

In this chapter we will summarize, very briefly, the general formalism of density functional theory (DFT) and the LDA for interacting 2D electrons/holes subjected to a disorder potential. Next we will present necessary numerical tools for simulating 2D electron system in an external disorder potential and then will give the results for compressibility coming from droplet picture by Shi ans Xie. We reproduced their calculation for both (strongly) correlated and uncorrelated (or weakly correlated) disorder potential and this will be discussed in the next Chapter. Finally in th last section (in this chapter) we will compare the droplet-picture results of Shi and Xie with the experiments and will emphasize the shortcoming of this formalism. We will try to convince the reader that there is an important piece of physics missing in this point of view (droplet picture) and will set the stage for introducing Si-Varma physics in the next chapter.

5.1 Density Functional Theory (DFT) and LDA

The density functional theory (DFT) approach expresses ground state properties -such as total energy, equilibrium positions and magnetic moments- in terms of the electronic density $n(\mathbf{r})$ or spin density, and provides a scheme for calculating them. The method avoids the problem of calculating the ground state wavefunction. *Thomas* and *Fermi* [142, 141] had attempted to formulate such an approach much earlier; their work suffered, however, from inaccuracies in the treatment of the kinetic energy and failed to allow for exchange and correlation effects. Nonetheless, this idea served as an starting point for the development of more advanced method by *Hohenberg, Kohn* and *sham* [143, 144].

For the application to real systems certain exact relations, derivable by dens-

ity functional theory, must be replace by approximations. By far the most popular one is the local density approximation (LDA) to the density functional theory. It provides us with very simple and at the same time very successful computational scheme. As a consequence, the computation of the ground state energy or the electronic density distribution is reduced to solving a single-particle Schrödinger equation with a local self-consistent potential. This reduction is given unambiguously by the theory, rendering the calculations on an ab-initio level, a significant economy of computational expenses compared with a calculation of many-body wave function. At the same time, however, it suggests that only little insight will be gained into the electron correlation problem.

5.1.1 Hohenberg-Kohn-Sham Theory

The developments and surceases of density functional theory are based on two theorems due to Hohenberg and Kohn. The first one states that the ground state energy E of a many-electron system in the presence of an external potential $V(\mathbf{r})$ is a functional of the electronic density $n(\mathbf{r})$, and that it can be written (in two dimensions) as

$$E[n(\mathbf{r})] = \int d^2r V(\mathbf{r}) n(\mathbf{r}) + F[n(\mathbf{r})]. \tag{5.1}$$

Here $F[n]$ is a unique -though unknown- functional of density $n(\mathbf{r})$ only, and does not depend on $V(\mathbf{r})$. The second theorem states that $E[n(\mathbf{r})]$ is minimized by the ground state density. To apply the theory, then, approximations must be made. Before describing them let us divide F in its different parts:[1]

$$F[n] = \frac{e^2}{2} \int d^2r d^2r' \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{kin}[n] + E_{xc}[n]. \tag{5.2}$$

The first term describes the Coulomb repulsion of the electrons. From the remaining we single out the kinetic energy of the system of the *noninteracting* electrons $E_{kin}[n]$. What remains is $E_{xc}[n]$, usually called the exchange and correlation energy. It should be noted that $E_{kin}[n]$ is not the true kinetic energy of the system, which would be hard to calculate owing to the many-body effects; instead, it is the kinetic kinetic energy of a fictitious, noninteracting system with the ground state density $n(\mathbf{r})$. The part of the kinetic energy which is difficult to calculate is contained in $E_{xc}[n]$, also including the exchange and remaining correlation energy. In order for $E[n]$ to be minimized, the density must satisfy the variational equation

$$\delta E = \int d^2r \delta n(\mathbf{r}) \left\{ V(\mathbf{r}) + e^2 \int d^2r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{kin}[n]}{\delta n(\mathbf{r})} + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \right\} = 0 \tag{5.3}$$

The variation $\delta n(\mathbf{r})$ is subjected to the subsidiary condition

[1] For the moment we give all the formulation in continuous form and later we discretize it on a 2D square lattice.

$$\int d^2r\delta n(\mathbf{r}) = 0 \tag{5.4}$$

in order for the total electron number to be conserved ($\int d^2r n(\mathbf{r}) = N$). The important observation is that Eq.5.3 is precisely the same as for noninteracting system in which electrons move in an effective external potential

$$V_{eff}(\mathbf{r}) = V(\mathbf{r}) + e^2 \int d^2r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}(\mathbf{r}). \tag{5.5}$$

Here an exchange-correlation potential has been defined through

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}. \tag{5.6}$$

One should note that in practice it is sometimes difficult to determine the functional derivative even when $E_{xc}[n]$ is known. The equivalence to a noninteracting electron system has become possible because of the way the kinetic energy $E_{kin}[n]$ has been extracted. This implies that $n(\mathbf{r})$ can formally be obtained if we first solve a Schrödinger equation of the form

$$\Big(-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(\mathbf{r})\Big)\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \tag{5.7}$$

and then calculate from it

$$n(\mathbf{r}) = 2\sum_i^{N/2} |\psi_i(\mathbf{r})|^2. \tag{5.8}$$

The sum is over the eigenfunctions with the lowest eigenvalues. From the above it becomes clear that $V_{eff}(\mathbf{r})$ is a function of the density $n(\mathbf{r})$. The set of self-consistence equations 5.7 and 5.8 are often called Kohn-Sham equations.

5.1.2 Local Density Approximation (LDA)

The eigenvalues ε_i in Eq. 5.7 has the following physical significance

$$2\sum_i^{N/2} \varepsilon_i = E_{kin}[n] + \int d^2r V_{eff}(\mathbf{r}) n(\mathbf{r}). \tag{5.9}$$

Now from Eq.5.1and Eq.5.2 it follows that the total energy in given by

$$E[n] = 2\sum_i^{N/2} \varepsilon_i - \frac{e^2}{2}\int d^2r d^2r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n] - \int d^2r v_{xc}(\mathbf{r})n(\mathbf{r}). \tag{5.10}$$

The real eigenvalues ε_i do not describe electronic excitation energies, which are in fact generally understoodt o be complex due to finite lifetimes of the excitations.

However, it turns out that for the infinite system with extended stats the energy of the highest occupied level (i.e. $\varepsilon_{N/2}$) is equal to the chemical potential μ. The Fermi energy is therefore correctly given by density functional theory. This does not hold true for the form of the Fermi surface, for which a corresponding proof neither exists nor should be expected. The complexity of real many-body problem is contained in the unknown exchange-correlation potential $v_{xc}(\mathbf{r})$. Nevertheless, making simple approximations, we can hope to circumvent the complexity of the problem. Indeed, the simplest possible approximation, i.e., local-density approximation (LDA), has proven very successful.

The local density approximation (LDA) consists of replacing the exchange-correlation energy $E_{xc}[n]$ by

$$E_{xc}[n] = \int d^2 r n(\mathbf{r}) \varepsilon_{xc}(n(\mathbf{r})), \tag{5.11}$$

e where $\varepsilon_{xc}(n(\mathbf{r}))$ is the exchange and the correlation energy per electron of a homogeneous electron gas of density n which is considered to be known[2]. Then

$$v_{xc}(\mathbf{r}) = \frac{d(n(\mathbf{r})\varepsilon_{xc}(n(\mathbf{r})))}{dn(\mathbf{r})} \tag{5.12}$$

depends only on $n(\mathbf{r})$ and Eq.5.7 takes a simple form. The effective potential $V_{eff}(\mathbf{r})$ depends only on $n(\mathbf{r})$, and the Schrödinger equation become much easier to solve than the Hartree-Fock equation with the nonlocal exchange potential. At the same time Eq.5.7 goes beyond the independent-electron approximation because of the correlation effects contained in $v_{xc}(\mathbf{r})$.

5.2 Droplet Picture and DFT-LDA Formulation for 2D Electron System

As mentioned earlier, there has been suggestions that MIT seen in the experiments might be a percolation type physics. He and Xie [31] have argued that there is a new liquid phase in the two-dimensional electron system in Si MOSFETs at low enough average electron densities. The percolation transition of this liquid through the disorder landscape in the classical limit causes the metal-insulator crossover behavior observed in recent -transport- experiments. A droplet state of the electron system resulted from the phase separation of the electrons into this new liquid phase and a low density "vapor" phase. The low density phase is called "vapor" purely for the reason that its density is low. In the presence of impurities, the "vapor" phase is a disordered Wigner crystal. Although electron-electron interaction plays important role in this picture, the type and scale of disorder potential fluctuations are very crucial in the formation of droplets at least on the level of LDA. Following the idea the "vapor" phase percolation, Shi and Xie implemented

[2]Accurate determinations these energies are available due to Quantum Monte Carlo (QMC) calculations [4].

droplet picture within LDA and looked into the behavior of inverse compressibility at low densities and its possible relationship to the droplet percolation near MIT. By choosing a special form for the disorder potential -to create (somewhat artificially) an inhomogeneous liquid needed for this formalism- they showed that the inverse compressibility, $1/\kappa \propto \partial\mu/\partial n$, turns upward near the percolation transition density. However, the strength of the κ^{-1} divergence is much weaker than it has been seen in the experiments. These calculations cannot -even qualitatively- explain the inverse compressibility data in highly mobile hole system in the experiment (see section 5.3). But due to the importance of their approach in connection with the materials we will present in the next three chapters, we found it very necessary to present and discuss these results here in more detail (section 5.2.3). Therefore in this section we are going to give DFT-LDA formulation for disordered interacting 2D electron system and present the relevant numerical techniques and approximations necessary to study the thermodynamics (chemical potential and inverse compressibility) as well as local properties (electron density distribution) of such system. We will reproduce the results from Shi and Xie calculations and discuss it in relation with the experimental data for $1/\kappa$. We will make the case that an important piece of physics is missing in this formalism. In turn this LDA viewpoint does form the fundament for the future developments in the next chapters.

5.2.1 Local Density Approximation for 2D Disordered Electrons

As has been already discussed in the previous section, in density function theory, for the system of 2D electrons subjected to a disorder potential, the functional form of the total energy is going to be

$$E_{tot}[n] = E_{kin}[n] + E_{e-e}[n] + E_{xc}[n] + E_d[n]. \tag{5.13}$$

Here $E_{kin}[n]$ is the functional of the kinetic energy, $E_{e-e}[n]$ is the direct Coulomb energy due to the charge inhomogeneity and $E_d[n]$ is the potential energy due to the disorder. The strong correlation effects caused by the electron-electron interaction is included in the exchange-correlation functional, $E_{xc}[n]$. The ground state density distribution can be obtained by minimizing the total energy functional with respect to the density. In the following we will discuss this expression in more detail to introduce approximations for the various combinations to the energy functionals.

- Kinetic Energy and The Gradient Expansion:

we notice that in $E_{tot}[n]$ only the kinetic energy depends on the electron wave functions explicitly:

$$E_{kin} = \sum_i \int_V d\mathbf{r}\psi_i^*(\mathbf{r})(-\frac{\hbar^2}{2m^*}\nabla^2)\psi_i(\mathbf{r}), \tag{5.14}$$

thus, the computational effort in minimizing the energy functional will be drastically reduced if E_{kin} is also expressed in terms of the local electron density $n(\mathbf{r})$.

This is the so called *gradient expansion* of the kinetic energy functional and for most solid state functional gives very good results for structural and electronic properties compared with both full wave function calculations and experimental measurements. In the following, a brief derivation of such an expanded kinetic-energy functional form is given for the 2D electron gas. For this derivation let us put $\hbar^2/m^* = 1$ for simplicity and then we will restore its original value at the end of calculation.

Following the procedure given in Ref. [146], the kinetic energy can be written in the form,

$$E_{kin} = \int\int d\mathbf{r}d\mathbf{r}' n(\mathbf{r}) f(k_F|\mathbf{r}-\mathbf{r}'|) n(\mathbf{r}'), \tag{5.15}$$

where $k_F$is the Fermi momentum for average electron density n_0 of the system. Considering a noninteracting system with a density having only a very small deviation from the homogeneous gas caused by a small potential perturbation $v(\mathbf{r})$,the total energy of system can be written as

$$E = \int\int d\mathbf{r}d\mathbf{r}' n(\mathbf{r}) f(k_F|\mathbf{r}-\mathbf{r}'|) n(\mathbf{r}')$$

$$+ \int d\mathbf{r} v(\mathbf{r}) n(\mathbf{r}) - E_f \int d\mathbf{r} n(\mathbf{r}), \tag{5.16}$$

where $v(\mathbf{r})$ is the external field and E_f is the Lagrange multiplier required to keep the total number of the electron fixed. Minimizing the energy functional, we get

$$2\int d\mathbf{r}d\mathbf{r}' f(k_F|\mathbf{r}-\mathbf{r}'|) n(\mathbf{r}') + v(\mathbf{r}') - E_f = 0. \tag{5.17}$$

Assuming samll perturbation form the uniform density, i.e., $n(\mathbf{r}) = n_0 + \delta n(\mathbf{r})$,we have the following two equations:

$$2n_0 \int d\mathbf{r}' f(k_F|\mathbf{r}-\mathbf{r}'|) - E_f = 0, \tag{5.18}$$

$$2\int d\mathbf{r}' f(k_f|\mathbf{r}-\mathbf{r}'|)\delta n(\mathbf{r}') + v(\mathbf{r}') = 0, \tag{5.19}$$

or in k space,

$$2F(k)\Delta N(k) = -V(k), \tag{5.20}$$

where $F(k)$, $\Delta N(k)$, and $V(k)$ are the Fourier transforms of f, δn, and v, respectively. Comparing with the linear response theory, we have

$$F(k) = \frac{1}{2G(k)}, \tag{5.21}$$

where $G(k)$ is the Green's function and for large k,

$$G(k) = \frac{1}{\pi}[1 - \sqrt{1 - (2k_F/k^{2)}}] \tag{5.22}$$

$$\sim \frac{1}{\pi}[\frac{2k_F^2}{k^2} + \frac{2k_F^4}{k^4} + \frac{4k_F^6}{k^6}]. \tag{5.23}$$

On the other hand when k is zero, $G^{-1}(0) = \pi$ [155]. Now we introduce the function $W(k)$,

$$W(k) \equiv G^{-1}(k) - \frac{\pi}{2}\frac{k^2}{k_F^2} + \frac{\pi}{2}. \tag{5.24}$$

Thus, $F(k)$ rewritten as

$$F(k) = \frac{1}{2G(k)} = \frac{1}{2}W(k) + \frac{\pi}{4}\frac{k^2}{k_F^2} - \frac{\pi}{4}. \tag{5.25}$$

Using the above expression, the kinetic energy can be written as

$$E_{kin} = \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' n(\mathbf{r})w(|\mathbf{r}-\mathbf{r}'|)n(\mathbf{r}')$$

$$-\frac{\pi}{4}\int d\mathbf{r}n^2(\mathbf{r}) + \frac{\pi}{4}\int dk|n(k)|^2\frac{k^2}{k_F^2}. \tag{5.26}$$

Since for the 2D homogeneous electron gas $n_0 = k_F^2/2\pi$, we have

$$\frac{\pi}{4}\int dk|n(k)|^2\frac{k^2}{k_F^2} = \frac{1}{8n_0}\int d\mathbf{r}|\nabla n(\mathbf{r})|^2 \tag{5.27}$$

$$\approx \frac{1}{8}\int d\mathbf{r}\frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})}, \tag{5.28}$$

where the approximation is valid up to quadratic order. Finally, we arrive at

$$E_{kin} = \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' n(\mathbf{r})w(|\mathbf{r}-\mathbf{r}'|)n(\mathbf{r}')$$

$$-\frac{\pi}{4}\int d\mathbf{r}n^2(\mathbf{r}) + \frac{1}{8}\int d\mathbf{r}\frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})}, \tag{5.29}$$

where $W(k)$is given by Eq.5.24. When the variation of the density is slow, one has

$$\int d\mathbf{r}' n(\mathbf{r}')w(|\mathbf{r}-\mathbf{r}'|) \approx n(\mathbf{r})\int d\mathbf{r}' w(|\mathbf{r}-\mathbf{r}'|)$$

$$= n(\mathbf{r})W(k=0) = \frac{3\pi}{2}n(\mathbf{r}). \tag{5.30}$$

In summary the first two term of Eq.5.29 correspond with the Thomas-Fermi term for 2D electron gas and the last term is just von Weizsäcker correction [148]; finally we arrive at

$$E_{kin}[n] = \int d\mathbf{r}[\frac{\pi}{2}n^2(\mathbf{r}) + \lambda\frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})}, \tag{5.31}$$

where $\lambda = 1/8$.

- Electron-electron interaction:

Part of the interaction effects between electrons(holes) can be expressed in form of direct Coulomb interaction, $E_{e-e}[n]$, which is

$$E_{e-e}[n] = \frac{e^2}{\varepsilon}\int d^2r \int d^2r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{5.32}$$

where ε is the dielectric constant of the sample. Here it is assumed that electrons are moving in a fixed positive background which does not contribute to their dynamics. In this case, it is better to write the Coulomb part as

$$E_{e-e}[n] = \frac{e^2}{\varepsilon}\int d^2r \int d^2r' \frac{\bar{n}(\mathbf{r})\bar{n}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{5.33}$$

where $\bar{n}(\mathbf{r}) = n(\mathbf{r}) - n$, in which $n = N/V$ is the average density of the positive background.

- Exchange and correlation effect:

The exchange-correlation energy functional, E_{xc}, can be divided into two parts

$$E_{xc}[n] = \int d^2r n(\mathbf{r})[\varepsilon_{exch}(n(\mathbf{r})) + \varepsilon_{corr}(n(\mathbf{r}))], \tag{5.34}$$

where ε_{exch} and ε_{corr} are the exchange and correlation energies per electron, correspondingly. For these energies we use the local form as derived from accurate Monte Carlo (MC) calculations by Ceperely and Tanatar [4]. According to their calculations, for normal (unpolarized) electron fluid, one can write the following expression for the exchange energy per electron (in Rydberg units)

$$\varepsilon_{exch}(n(\mathbf{r})) = -\frac{8}{3}\sqrt{\frac{2}{\pi}}n^{1/2}(\mathbf{r}), \tag{5.35}$$

and in order to fit the MC calculated -ground state- energy to a functional form, they assume that the correlation energy, ε_{corr}, can be approximated by (in Rydberg units)

$$\varepsilon_{corr}(n(\mathbf{r})) \approx a_0\frac{1+a_1x}{1+a_1x+a_2x^2+a_3x^3}, \tag{5.36}$$

where $x = (r_s)^{1/2}$ and the parameters a_i (for normal electron liquid) in the above Padé approximation are determined by a nonlinear least-squares fit to the MC data at different densities of electrons. These parameters for the system of spinless electrons are: $a_0 = -0.3568$, $a_1 = 1.1300$, $a_2 = 0.9052$, $a_3 = 0.4165$. This Padé form behaves like $a + br_s$ as $r_s \to 0$, which has the correct high-density expansion (short of the $r_s \ln r_s$ term), and admits the asymptotic form

$$E_{corr} = \frac{a}{r_s} + \frac{b}{r_s^{3/2}} + \frac{c}{r_s^2} + \ldots \tag{5.37}$$

as $r_s \to \infty$. It is clear that for very low densities (large r_s) one can easily neglect the correlation energy.

- External Disorder Potential:

We assume that 2D electron system is subjected to an -external- disorder potential $V_d(\mathbf{r})$. This implies that the last term in energy functional Eq.6.11 takes the form

$$E_d[n] = \int d^2 r n(\mathbf{r}) V_d(\mathbf{r}). \tag{5.38}$$

The disorder potential can be a uncorrelated quench disorder of strength V_s

$$\langle V_d(\mathbf{r}) V_d(\mathbf{r}') \rangle = V_s^2 \delta(\mathbf{r} - \mathbf{r}'), \tag{5.39}$$

or a smooth and correlated disorder charactarized by -for example- an exponential fall-off

$$\langle V_d(\mathbf{r}) V_d(\mathbf{r}') \rangle = V_s^2 \exp\Big(- \frac{|\mathbf{r} - \mathbf{r}'|}{\xi} \Big), \tag{5.40}$$

where V_s is the amplitude of the potential fluctuation, and ξ is the correlation length of the disorder. For a slowly varying potential, ξ is roughly the average size of valleys in a disorder landscape. Another interesting example of disorder is an off-plane charge impurity potential of the form

$$V_d(\mathbf{r}) = -\sum_i \frac{n_i}{\sqrt{|\mathbf{r} - \mathbf{r}_i|^2 + d^2}}, \tag{5.41}$$

where d is the distance between electrons and the impurity planes, and the impurities are randomly distributed with a density n_i.

5.2.2 Ground State Energy and Density

The density distribution of the ground state can be obtained by minimizing the energy functional under the constraint of a constant total electron number

$$\int n(\mathbf{r}) d^2 r = N. \tag{5.42}$$

Since $n(\mathbf{r})$ is non-negative it is often convenient to use the new variable χ with

$$n(\mathbf{r}) \equiv \frac{N\chi(\mathbf{r})^2}{\int d\mathbf{r}\chi(\mathbf{r}')^2}, \tag{5.43}$$

where N is the total number of the electrons in the system. The constraint that the total electron number is a constant is automatically satisfied with new variable. Minimization of the energy functional with respect to the density requires

$$\frac{\delta E_{tot}[\chi(\mathbf{r})]}{\delta\chi(\mathbf{r})} = 0, \tag{5.44}$$

which leads to the following nonlinear equation:

$$[-\nabla^2 + V_{eff}[n(\mathbf{r})] - E_0]\chi(\mathbf{r}) = 0, \tag{5.45}$$

where

$$V_{eff}[n(\mathbf{r})] = 2\pi n(\mathbf{r}) + V_d(\mathbf{r})$$

$$+\frac{\delta}{\delta n(\mathbf{r})}[(\varepsilon_{exch} + \varepsilon_{corr})n(\mathbf{r})], \tag{5.46}$$

and

$$E_0 = \frac{\int d\mathbf{r}\chi(\mathbf{r})[-\nabla^2 + V_{eff}[n(\mathbf{r})]\chi(\mathbf{r})}{\int d\mathbf{r}\chi(\mathbf{r}')^2}. \tag{5.47}$$

Solving the nonlinear partial differential equation Eq.5.45 is difficult and numerically inefficient. Instead of solving this equation we directly minimize $E_{tot}[n(\mathbf{r})]$ using the conjugate gradient method [156]. By using this method, we can directly minimize $E_{tot}[n]$ by iterating the local function $\chi(\mathbf{r})$ until the total energy is minimized. This is achieved by the following equation subjected to the constraint Eq.5.42:

$$\chi^{m+1}(\mathbf{r}) = \chi^m(\mathbf{r}) - \Delta\left\{\frac{\delta E[\chi(\mathbf{r})]}{\delta\chi(\mathbf{r})}\right\}^{conj}_{\chi=\chi^m}, \tag{5.48}$$

where the function $\chi(\mathbf{r})$ of the $(m+1)$th iteration is obtained from it's values at mth iteration. $\{...\}^{conj}$ means that we have to evaluate the gradient of the total energy with respect to the local function $\chi(\mathbf{r})$ in conjugate direction. Δ is "time" step in the iteration procedure which is chosen to be small enough to maintain the stability of the iteration process, but large enough for fast convergence. This is basically equal to solving $\delta E[\chi]/\delta\chi = 0$ under the constraint 5.42.

5.2.3 Compressibility in Droplet Picture

Shi and Xie applied the above technique of minimizing the total energy functional to the 2D electron system. To create an inhomogeneous charge distribution in liquid phase, necessary for any percolation scenario to work, it turns out to be necessary to include slowly varying, strongly correlated disorder potential. The calculation is carried out in a 128×128 discrete space. The size of the system is set as $L = 256a_B^*$, where a_B^* is the effective Bohr's radius, $a_B^* = \varepsilon\hbar^2/m^*e^2$, with ε being the dielectric constant and m the effective mass of an electron. The periodic boundary condition and the Ewald sum (see Appendix B) for the Coulomb interaction are applied to minimize the finite size effect. The electron density is adjusted by changing the total electron number N. The density distribution and the total energy of the system are calculated for different densities, and the chemical potential is calculated by using the formula,

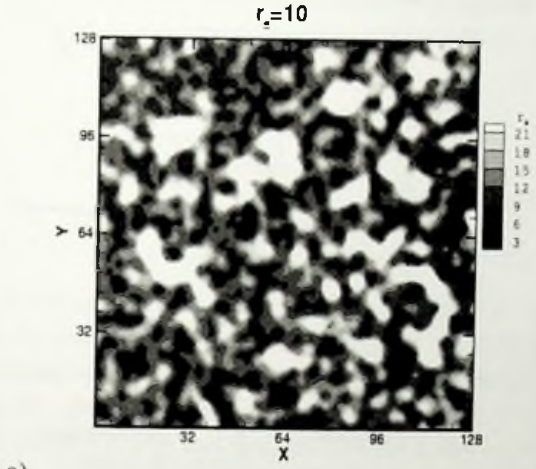
$$\mu(N) = \langle E(N+1) - E(N) \rangle_{disorder}. \tag{5.49}$$

The compressibility of the system can be calculated by

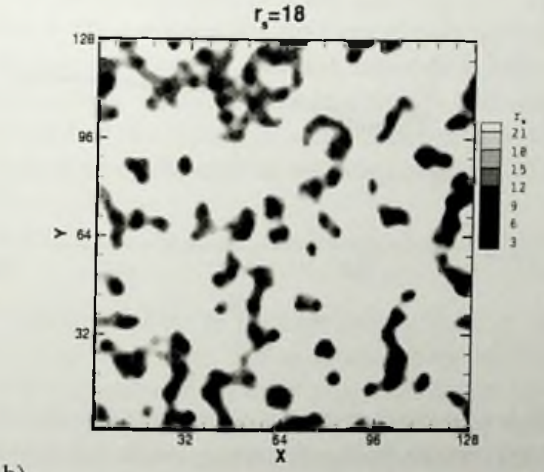
$$\frac{1}{\kappa} = \langle \frac{N^2}{V} \frac{\partial \mu}{\partial N} \rangle_{disorder}, \tag{5.50}$$

where $V = L^2$ is the total area of the system and $\langle ... \rangle_{disorder}$ means average over disorders. Figure 5.1 shows the density distribution of the system. It can be clearly seen that the electrons form some high density regions, while the density of other regions are essentially zero (the "vapor" phase). Depending on the average density of the system, the high density regions may connect each other together ($r_s = 10$), or form some isolated regions ($r_s = 18$). There exists a certain density in between where the high density regions starts to percolate through the system, and form a conducting channel. Based on these pictures Shi and Xie argued that the metal-insulator transition observed in the 2D electron systems is associated with a percolation transition in the electron density. The electron-electron interaction is important for the conducting behavior of a dilute electron system for the reason that it makes the density distribution more extended because of the Coulomb repulsion.

Figure 5.2 shows the compressibility of the systems. Experimentally the quantity $\delta\mu/\delta N \propto \kappa^{-1}$ is direct measure of the compressibility. It is well known that the compressibility of a uniform electron gas is negative in the low density region due to the effect of the exchange and correlation energies, as shown by the solid line in Fig. 5.2(a). However, when disorder is present, the behavior changes greatly. In the low density, the electrons tend to occupy the valleys of the disorder landscape, and the local density, instead of the average density, determines the compressibility of the system. On the other hand, at higher densities, all of the valleys are filled, and one can expect the compressibility of the system to resume the behavior of a uniform electron gas. One can see a non-monotonic behavior for $\delta\mu/\delta N$, as shown by the dots in Fig. 5.2(a), which are in good agreement, qualitatively, with some of the experimental measurement [129]. Accordingly, the turning point of



a)



b)

Figure 5.1: The density distributions for the different electron densities. The contour plot has been used for the local density parameter $r_s = 1/\sqrt{\pi n}$. The density of the white area decreases rapidly to zero when r_s increases. The amplitude and the correlation length of the disorder potential are $V_s = 0.2Ry$, $\xi = 0.2L$ respectively.(From Shi and Xie[138])

the compressibility ($N \approx 50$, $r_s \approx 14$) coincides with the percolation threshold of the system. At low densities, the data points in the plot show strong fluctuation, indicating the effect of the local fluctuation of the disorder potential.

Based on the droplet state scenario, Shi and Xie argued that this behavior of the compressibility can be understood by the following simple (but physically unclear) theory. By definition, the chemical potential μ is the energy needed to add an

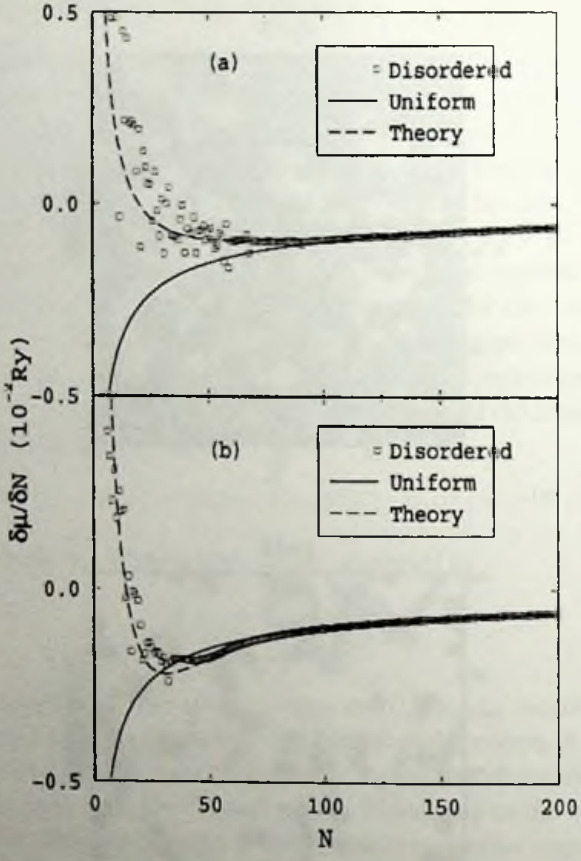


Figure 5.2: $\delta\mu/\delta N$ as a function of the electron density. Solid lines are for the uniform electron gas, squares are the data points for the disordered system, and the dashed-lines are the results from the theory discussed in the text. Two kinds of the disorder are used in the calculation: (a) The same disorder landscape as in Figure 5.1; (b) Off-plane charge impurities with $d = 10a_B^*$, $n_i = 2.5 \times 10^{-3}/a_B^{*2}$. The parameters in the dashed-lines: (a) $n_0 = 10^{-3}/a_B^{*2}$, $\alpha = 1.5$; (b) $n_0 = 0.5 \times 10^{-3}/a_B^{*2}$, $\alpha = 2.3$. N is the total number of the electrons in the simulation box. (From Shi and Xie [138])

electron into the system,

$$\begin{aligned} \mu(N) &= E(N+1) - E(N) \\ &\approx \varepsilon_0[n_{eff}(N+1)](N+1) - \varepsilon_0[n_{eff}(N)]N \\ &= \frac{\delta}{\delta N}\{\varepsilon_0[n_{eff}(N)]N\}, \end{aligned} \tag{5.51}$$

where we suppose that the electron energy is determined by the local density of the electrons. $E(N)$ is the total energy of the system, $\varepsilon_0(n)$ is the energy per electron for the uniform electron gas, and n_{eff} is effective local density. For the inhomo-

geneous system as shown in Fig. 5.1, the effective local density can be estimated by $n_{eff}(n) \approx n/f(n)$, where $f(n)$ is the fraction of high density region. After some algebra, one finds

$$\mu(n) = \mu_0\left(\frac{n}{f}\right)\left[1 - \frac{d\ln f}{d\ln n}\right] + \varepsilon_0\left(\frac{n}{f}\right)\frac{d\ln f}{d\ln n}, \tag{5.52}$$

$$\frac{\delta\mu}{\delta n} = \mu_0'\left(\frac{n}{f}\right)\frac{1}{f}\left[1 - \frac{d\ln f}{d\ln n}\right]^2 - \varepsilon_0'\left(\frac{n}{f}\right)\left(\frac{n}{f}\right)^2\frac{d^2 f}{dn^2}, \tag{5.53}$$

where μ_0 is the chemical potential for a uniform electron gas. In the low density limit, $f(n) \to 0$, the local effective density is greatly different from the average density of the system. As a consequence, the density dependence of the chemical potential, $\delta\mu/\delta n$, changes greatly. In general, supposing $f(n) \sim n^\alpha$ in the low density limit, the analysis shows that $\delta\mu/\delta n$ will have a non-monotonic behavior if $\alpha > 1$. The behavior of $f(n)$ is determined by the local disorder potential profile. In a 2D system, the infinite harmonic potential has $f(n) \sim n$. So the requirement $\alpha > 1$ is equivalent to the condition that the local disorder potential has a weaker confinement effect than the harmonic potential. However it is not really clear what is the relationship between f and disorder potential V_d and whether this condition can be easily satisfied in a typical experimental system. In Fig.5.2, the above equation has been used for $\delta\mu/\delta n$ with the following relation of $f(n)$ to fit the data,

$$f(n) = \frac{1}{1 + (n_0/n)^\alpha}. \tag{5.54}$$

This form has a correct behavior in the high density limit, $f \to 1$, and the low density behavior is controlled by α. Before turning to the next section where we compare the LDA-droplet calculations to the experimental data for compressibility, discussed in the previous chapter, we would like to note that although this theory would explain the data from the numerical simulations for inverse compressibility, the physical meaning of the function f is rather obscure. The effective density defined above becomes unphysical at low densities where $f \to 0$. Furthermore, with a very small change in any of the above fitting parameters, it is impossible to fit the theory to the numerical data even qualitatively.

5.3 Comparison with Experiment

Let us now compare the above results with experiment. We consider two experimental results by Dultz *et al.* [13, 137], namely the inverse compressibility measurement for both two-dimensional electron and hole system (2DES, and 2DHS). Over the years, it has been suggested that the interacting 2D electrons in the presence of disorder can be understood by the droplet, or puddle, model as discussed in above [31, 32]. The electrons are separated into "liquid" droplets with local density higher than the average density and "vapor" islands with lower density. For a

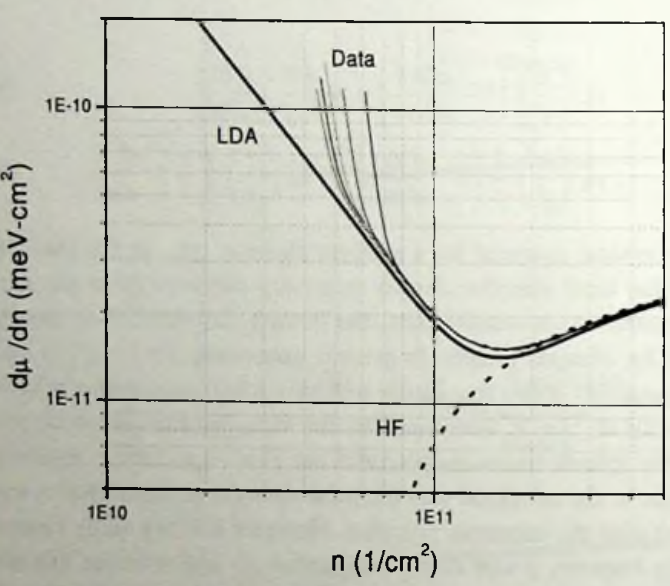


Figure 5.3: The experimatal data for the low-mobility electron system for different temperatures, T, from 0.28K to 1.2K. This data is compared with the theoretical prediction of compressibility based on LDA-droplet picture . The LDA approximation describes the data well in the high densities with n_0=1.6 × 10^{11} cm^{-2} and α=2.0 as fitting parameters. For comparison the theoretical curve based the Hartree-Fock (HF) theory of free electron gas (neglecting the correlation energie) is also included (data from Dultz *et al* [137]).

given disorder, due to the potential fluctuation of the donors, as the Fermi energy is decreased, the "vapor" region increases. In this model, a crossover from metallic behavior to insulating behavior should be seen around the percolation threshold.

Following the suggestion in the previous sections, the compressibility can be calculated numerically using LDA for the experimental parameters for 2DES and found that the basic characteristics of the electron data can be well described by the formalism of the LDA. As shown also by Dultz *et al.* , the LDA approximation describes the data rather well in the low mobility electron system (see figure5.3). To fit the data based on the LDA-droplet picture one can use the relation Eq. 5.53 for inverse compressibility and make use of the definition of effective density n_{eff} with the relation Eq. 5.54 for $f(n)$. The best fit can be produced adopting the following values for the parameters: $n_0 = 1.6 \times 10^{11}\ cm^2$ and $\alpha = 2.0$. Due to the special relationship between f with α and n_0, it turns out that the numerical fit is very sensitive to these parameters. Any small deviations of from $n_0 = 1.6 \times 10^{11}\ cm^2$ or α=2.0 (even as small as 10%) dramatically effect the theoretical curve

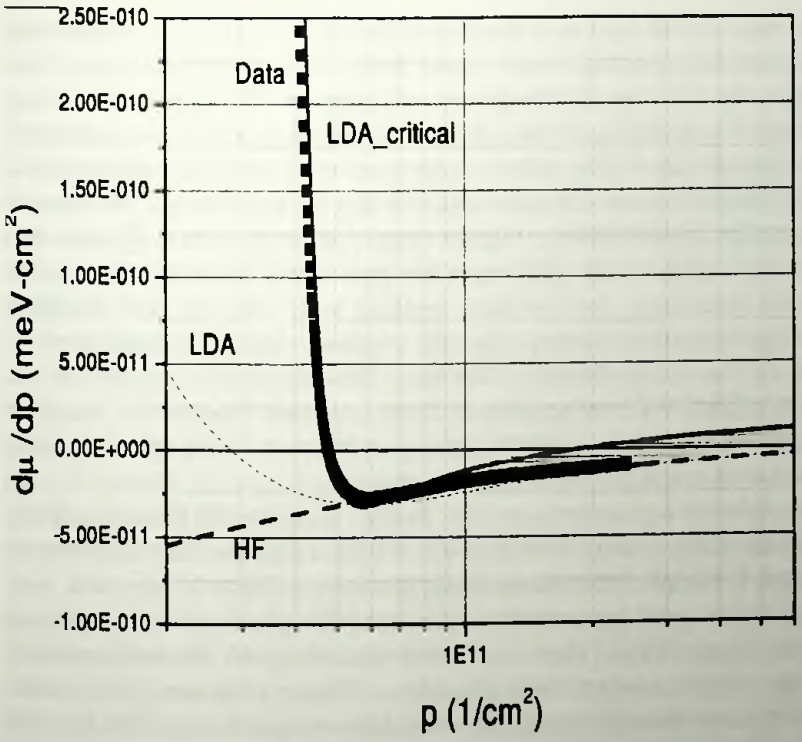


Figure 5.4: As discussed in the text, the LDA approximation of droplet picture cannot describe the $d\mu/dp$ vs. p curve obtained in the 2D hole system, since the theoretical curve varies much slower than the experimental data with p (p is the hole density). An example of reasonable fit for low densities has been also shown based on the singular $f(n)$ (LDA_critical curve), as described in the text. HF curve is chosen for comparison, derived from the Hartree-Fock theory of the free electron gas (neglecting correlation energy).

and it makes it impossible to fit the data even qualitatively. Therefore one may conclude that the incidentally matching values for $\alpha = 2.0$ to that of 2.3 used to fit the numerical simulation by Shi and Xie for off-plane charge impurities (see figure 5.2). This parameter should depend on the form of the disorder potential which has not been addresses in the droplet theory. It is very clear from figure 5.3 that the data (for compressibility) at lower densities deviate from the theoretical curve. This obviously shows that LDA physics would fail as we get closer to the region of low densities where the electron-electron interaction is enhanced.

On also should note that the qualitatively good data fit in Figure 5.3 to the theoretical curve (for high densities) should not give the impression that the droplet model explains all experimental data for different systems. In fact, Figure 5.4 shows that the LDA does not fit the data for the 2DHS density dependence of $d\mu/dp$ at all. We remind the reader that in these systems r_s is large, due to the fact that the effective mass of holes is large (roughly 5 times larger than that of elec-

trons). To compare with the data in the experiment of 2DHS [137], it is impossible to find parameters of α and n_0 which would fit the curve even qualitatively. This clearly shows that it is more into the physics of highly interacting hole system than the (LDA-based) percolation physics. It is interesting to note that, one can match the data , in low density region, with an expression which contains a mathematical singularity (which of course is inconsistent with that for percolation). An example of such curve has been shown in Fig.5.4 using $f(n) = (1 - (n_0/p)^{\beta})$ with n_0= $3.5 \times 10^{10}\ cm^{-2}$ and $\beta = 0.06$. This curve has been shown as LDA_critical in the figure and the terminilogy does not have anything to do with any kind of critical theory; it is just introduced to emphasize the proposed singularity which leads to a criticality of free energy through f-function. This observation implies that the physics of the 2DHS which appears in the large r_s regime, "cannot" be described by the droplet model and one needs to develope a theory to incorporate the strong interactions between charge carriers in these systems.

The necessity of a singularity in $f(n)$ further suggests that there is a phase transition in the 2DHS system. This observation further supports the notion that the non-analytical behavior of the thermodynamic compressibility is consistent with the unusual signature of the metal-insulator transition observed in the transport for the same sample. Indeed there are theories describing the 2D metal-insulator transition as a phase transition from a liquid to a Wigner glass state [36] or from a paramagnetic state to a spontaneously ordered ferromagnetic sate [70]. It is difficult to determine the existence of these potential states from these measurements due to the absence of a quantitative analysis of the compressibility of the Wigner glass state or the ordered ferromagnetic state. For example, the compressibility of a Wigner crystal, in the absence of disorder, is known theoretically [149] to be only twice that for a liquid in the HF theory. However, the compressibility for a glass state, in the presence of realistic disorder, has not yet been calculated.

It is reasonable to speculate about why the droplet model is more appropriate to describe the smaller r_s system. It is now a rather common view that the interaction can suppress the effect of disorder. As will be discussed later, the numerical simulation of the correct physics of the 2D electron(hole) system shows that, as r_s becomes more important the system tends to keep the electron density as uniform as possible to minimize the Coulomb interaction. So for the strongly interacting heavy mass hole system, it is possible that the tending toward droplet formation is largely suppressed.

5.4 Conclusions

One can get no information about the thermodynamics of 2D disordered electron system within Finkelstein theory, as it always assumes the compressibility of the system (related to the singlet interactions) is fixed and unrenormalized. On the other hand, due to the lack of information on electron correlations, Hartree-Fock theory fails to give any explanation on the compressibility behavior seen in the

experiments. Although it may describe the Wigner crystal very well, the critical density arising in this theory is very small. Over the years, it has been suggested that the interacting 2D electrons in the presence of disorder can be understood by the droplet, or puddle, model. The electrons are separated into "liquid" droplets with local density higher than the average density and "vapor" islands with lower density. For a given disorder, due to the potential fluctuation of the donors, as the Fermi energy is decreased, the "vapor" region increases. In this model, a crossover from metallic behavior to insulating behavior should be seen around the percolation threshold.

It seems that density functional theory, within local density approximation, can provide a useful tools for study disordered 2D electron liquid near MIT critical point. Within this framework at least using a special type of correlated disorder potentials with slowly varying amplitude, it has been shown that for strong enough disorder the electronic inverse compressibility deviates from free-electron gas prediction (HF) and goes upward. Although this approach might be able to describe the compressibility data of low-mobility electron system in higher densities, it clearly fails to give anything -even qualitatively- like the behavior of this quantity in the low density 2D hole system, where the anomaly is much stronger.

We believe that there is an important piece of physics missing in this formalism which is very crucial for electron system at low densities. In this regime where the interactions become more and more important, it would be very difficult for electron system to screen themself and the system tends to keep the electron density as uniform as possible to minimize the Coulomb interaction. In the next chapter we are going to show how to formulate this idea within DFT-LDA formalism.

Chapter 6

Si-Varma Theory and Correction to LDA

As has been discussed in the previous chapter, in on hand, there is less than little hope in Hartree-Fock (HF) approximation for correct description of thermodynamical and transport properties of electron (hole) system in MOSFET approaching from the metallic side. On the other hand, although DFT formulation of the free energy within LDA framework yields a reasonable description of - the experimentally seen - inhomogeneity in the spatial distribution of the electronic density near the insulating phase, however, the change in the behavior of the electronic compressibility across the metal-insulator transition point is very smooth compared to the experimental observations. HF theory underestimates the metallic behavior and gives a good description of Wigner crystal deep into the insulating phase, while LDA overestimates the electronic screening in the liquid phase at low densities near MIT. Clearly, both theories are disconnected from either side of MIT transition point. Moreover, any percolation physics of metal-insulator transition needs the metallic state to be inhomogeneous in some scale. But the experimental evidence [14] discussed in chapter 4 seems to indicate otherwise: a homogeneous metallic phase. In this chapter we will introduce a theory that tries to address the correct physics near the MIT point approaching it from metallic side; and it is based on the notion that near the transition, Coulomb interactions in the electron system cannot be screened easily any longer when the MIT is approached. This become important when $r_s > 1$, due to the strong long-range Coulomb interaction the electrons have little room to move in order to screen the interaction. In this situation one can expect that the screening length, s, might become larger than mean free path of the electrons, ℓ, and the crucial question is: is there new physics emerging from this regarding screening properties in the presence of disorder. This is the focus point of the Si-Varma theory of the MIT. In this chapter, we will introduce this theory and then we also will develop a DFT formulation of this theory as a first order correction to LDA which is the main result of this thesis: *density functional theory (DFT), applicable to the metallic side of the MIT transition, incorporating*

the physics of highly unscreened electrons (Si-Varma Physics).

6.1 Screening length and compressibility renormalization near MIT

A feature of the RG approach of Finkelstein, discussed in chapter 2, is that the compressibility κ (proportional to the inverse screening length $1/s$) is unrenormalized. This is valid at high enough density so that the screening is good and s is less than the mean free path ℓ. Si and Varma pointed out that at low electron density (large r_s) of the experiment, one might expect $s > \ell$ and thus a renormalization of the compressibility. They argued that for large r_s, the interaction becomes unscreened at low temperature and the compressibility approaches zero, as might be expected as one nears a metal-insulator transition from the metallic side. They also calculated a large suppression of the conductivity which overcomes the weak increase found in the earlier RG analysis of Finkelstein [22, 23] and Castellani *et al.* [24, 56]. This leads to a metal-insulator transition (at which both conductivity and compressibility go to zero) controlled by whether s is larger or smaller than ℓ.

Now let us look at this idea in more detail. It has been already known that for the length scales larger than the mean-free path, ℓ, the interactions may be parametrized by just two dimensionless parameters: interaction in the singlet channel γ_s and interaction in the triplet channel γ_t. The singlet interaction parameter depends on the range of the interaction, but for Coulomb interaction it is always assumed to be completely screened. The lowest order of γ_s is independent of density while γ_t varies from zero (for vanishing r_s) to order of unity (for large r_s). In chapter 2 we saw that, in the Finkelstein theory for weakly-disordered $2D$ system, the interaction parameter γ_t scales to infinitely large values before zero temperature is reached. To understand the situation better consider a disordered system of size L in $2D$. The one-loop correction (including a correction due to Finkelstein [22, 23]) to conductance, g, can be written as

$$\frac{\delta g}{g} = -\frac{1}{g}(\gamma_s - \frac{3}{2}\gamma_t)\log(L/\ell). \tag{6.1}$$

It is clear that γ_t tends to increase the conductivity. In the Finkelstein theory γ_t has also a logarithmic singularity and scales upward as temperature is decreased. There is no γ_s renormalization in this theory. The value of γ_s in this theory is about 2. As a result, the renormalization flow is always toward a state of finite conductivity at low temperature, i.e. a metallic state for any density. Unfortunately the RG procedure fails as soon as γ_t exceeds unity. However, as the temperature is lowered, the resistivity first increases slightly and then begins to decrease (as often seen experimentally), just as the coupling becomes too large. This suggests that a low-temperature metallic state might be achieved. As we discussed in chapter 2, this theory does not contain a metal-insulator transition (in its original form), nor is the nature of the possible metallic state revealed.

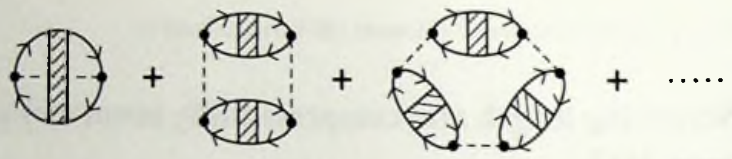


Figure 6.1: Series of ring contribution to the ground state energy with disorder. The hatched lines represent the t matrix for impurity scattering, and the dashed lines represent the Coulomb interaction.

There is no renormalization of compressibility at all in the Finkelstein theory since the γ_s parameter coming from the exchange contribution is scale-independent. Si and Varma [34] considered the possibility of scaling of γ_s within Finkelstein theory. This is quite reasonable for two obvious reasons: first of all, there is a metal-insulator transition in the pure limit, i.e., a Wigner transition. As r_s increases, either in two or three dimensions, a first-order transition to the Wigner crystal is expected to occur due to long-range nature of the Coulomb interaction even when spin is left out. The Wigner transition appears to satisfy conditions in which disorder turns a first-order transition to a continuous transition [145]. So it seems surprising that the singlet interaction would become irrelevant on the metallic side in the presence of disorder. Secondly, In any Fermi-liquid like theory the renormalization of the spin symmetric interaction parameter is related to the renormalization of the compressibility. The effective electron-electron interaction can be represented exactly as

$$U_{eff}(q,\omega) = \frac{U(q)}{1+U(q)\pi(q,\omega)} \tag{6.2}$$

where $U(q) = 2\pi e^2/\varepsilon q$ and $\pi(q,\omega)$ is the irreducible polarizability, which in the limit of $\omega \to 0$, $q \to 0$ is the compressibility κ. The screening length therefore is proportional to κ^{-1}. In the insulating state one expect unscreened interactions. If the metal-insulator transition is continuous, the screening length on the metallic side must diverge and the compressibility tend to zero.

The compressibility (and the correction to it) can be calculated from the calculation of the ground state energy E_g (and correction to energy, δE_g), per electron, using the following relation:

$$\kappa^{-1} = \frac{d^2 E_g}{dn^2}. \tag{6.3}$$

The interaction contribution to the ground state energy E_g, associated with the presence of disorder, can be calculated from the correction to the exchange and correlation contribution to this energy. This includes essentially integrals over momentum, q, and frequency, ω, of functions of the dynamic structure factor [151]. The low-momentum transfer part comes from the so called ring diagrams with disorder

$$E_{ring} = \frac{i}{2}\int d\omega \int d^2q \log[1 + U(q)\pi(q,\omega)] \tag{6.4}$$

where $U(q) = 2\pi e^2/\varepsilon q$ in 2D and the proper polarization $\pi(q,\omega)$ is the same as in diffusive regime. It can be expressed as

$$\pi(q,\omega) = \kappa \frac{Dq^2}{i\omega + Dq^2}, \quad q \ll \ell^{-1} and\, \omega \ll \tau^{-1} \tag{6.5}$$

The diagrammatic picture of this process is shown in Fig.6.1. For large momentum transfers, processes beyond the ring diagrams, representing exchange corrections at short distance, become important. The contribution from these additional processes can be calculated [34] following the Hubbard interpolation scheme, in which the bare Coulomb interaction in the susceptibility in multiplied by a factor $[1 - F(q)]$, where $F(q) = q^2/2(q^2 + k_F^2)$. As a result the leading order-disorder correction to the ground state coming from all above processes, is [34]

$$\delta E_g = -A\frac{s_0}{\ell}. \tag{6.6}$$

Here ℓ is the mean-free path, and

$$s_0 = (2\pi e^2 \nu/\varepsilon)^{-1} \tag{6.7}$$

is the screening length in Thomas-Fermi approximation, in which ν is the density of states for free electrons. The constant $A \approx 4/\pi$. Using Eq.6.3, the corresponding additive correction to the inverse compressibility is

$$\delta(\frac{\kappa_0}{\kappa}) \approx 0.16 r_s \frac{s_0}{\ell(\tau_0)}, \tag{6.8}$$

Where $\kappa_0 = \nu$ is the contribution to the compressibility of the kinetic energy. Here τ_0 is assume to be independent of density, as is appropriate for the experimental systems in the immediate vicinity of MIT.

The actual screening length s is related to the compressibility through

$$s/s_0 = \kappa_0/\kappa. \tag{6.9}$$

Thus the screening length increases as ℓ^{-1}. Using Eq.6.8, one can look for the condition that $s \gg \ell_0$, the value of ℓ at temperatures of order of the Fermi energy. The dominant contribution for the correction to the compressibility comes for $r_s \gg 1$ and gives the condition

$$r_s \geq 3(\omega_0\tau_0)^{1/2}, \tag{6.10}$$

where τ_0 is the scattering time and $\omega_0 = \hbar/ma_B^2$; m is the electron mass and a_B is the Bohr radius.

The most important prediction of the theory is vanishing of the compressibility as the transition is approached from metallic side, as have been seen in the experiments [13, 129, 14]. Another prediction of the theory is the condition Eq.6.10 for the metal-insulator transition. The existing experimental results [7, 12, 11] are consistent with it in the τ-dependence while the order of magnitude of the critical r_s according to Si-Varma is larger than that in the experiments. The metal-insulator transition occurs at $\omega_0\tau_0$ of about 100 and r_s of about 20; while, r_s at the transition showing slight increase as sample quality is improved.

6.2 DFT-LDA formulation of Si-Varma

We have seen in the last chapter that upon using the density functional theory, one can formulate the total energy, E_{tot}, of the interacting electron system within LDA. Since in the LDA the effect of electronic screening near metal-insulator transition is over-looked, we have seen that within this framework one does not obtain an accurate description of the change in the behavior of the electronic compressibility near the metal-insulator transition point, when the r_s become large and the disorder $(1/\omega_0\tau_0)$ is weak. This should be specially important for highly mobile electron/hole systems. So, it seems very reasonable to implement the Si-Varma idea within DFT as a first-order correction to LDA, caused by the presence of the inhomogeneous disorder potential. I emphasize here that this correction is far from being the theory of the metal-insulator transition of electron systems under consideration. The theory does not contain any information regarding the Wigner phase on the other side of the transition point.

Before giving the explicit form of the energy functional, we also want to emphasize here that the Si-Varma results Eqs.6.4-6.6 relate ground state energy to density fluctuations in a local way. This is very important as this will allow us to improve the density functional theory, keeping the improved functional local in the charge density. This will save us of any complications arises from, for example, the gradient correction to exchange-correlation energy.

Let us again express the total energy functional of the interacting electron system in DFT-LDA formalism, including the correction from Si-Varma:

$$E_{tot}[n] = E_{kin}[n] + E_{e-e}[n] + E_{xc}[n] + E_d[n] + E_{s-v}[n]. \tag{6.11}$$

The description of each term is as follows: the first three terms, E_{kin}, E_{e-e}, E_{xc}, are exactly the same as for the standard LDA of the 2D electron system (see chapter 5). The fourth term in the above energy functional is a disorder energy

$$E_d = \int d^2r n(\mathbf{r}) V_d(\mathbf{r}), \tag{6.12}$$

where $V_d(\mathbf{r})$ is the uncorrelated disorder potential. In the Si-Varma theory the strength of disorder in the system is parametrized by the scattering time, τ_0, or equivalently the quantity $\omega_0\tau_0$. Using the Fermi's golden rule we can relate this

quantity ($\omega_0\tau_0$) to the amplitude of the disorder potential (V_d). The transition (scattering) rate according to the Fermi's golden rule is

$$\frac{1}{\tau_0} = \frac{2\pi}{\hbar}|V_d|^2\nu, \tag{6.13}$$

or (in Rydberg units) we can rewrite it as

$$|V_d| = \left(\frac{\hbar^2\varepsilon}{me^2}\right)\sqrt{\frac{2}{\tau_0\omega_0}}. \tag{6.14}$$

Here, τ_0 is the scattering time associated with the disorders, $\omega_0 = \hbar/ma_B^2$ and $\nu = m/\pi\hbar$ is the density of states of free electrons in 2D. The relationship between ω_0 and the more conventional disorder parameter $(k_F\ell)^{-1}$is as follow:

$$(k_F\ell)^{-1} = \frac{r_s^2}{2\tau_0\omega_0} \tag{6.15}$$

The last term comes from the correction to the exchange energy due to the ring diagrams, by Si and Varma, discussed in previous section. This part can be written as (in Rydberg units)

$$E_{s-v} = \int d^2r n(\mathbf{r})\varepsilon_{s-v}(n(\mathbf{r})). \tag{6.16}$$

As expressed in Eq.6.6, the correction to the exchange-correlation energy is a function of s_0, the Thomas-Fermi screening length, and ℓ, the mean free path, of the electron system. Fortunately, the ratio of these two quantities can be written as a function of the density of the charge carries, n, and this is in fact the key to the argument that the Si-Varma correction can easily be considered as the correction to the LDA approximation in DFT formalism. The mean free path of the electrons, ℓ, can be expressed as follows:

$$\ell = \tau_0\hbar k_F/m \tag{6.17}$$

in which $\hbar k_F$ is the momentum of electrons near Fermi surface. In two dimension k_F can be re-expressed in terms of r_s as follows:

$$k_F r_s a_B = \sqrt{2} \tag{6.18}$$

Using the above formula, Eq.6.7 and Eq.6.17, the correction part to the exchange-correlation energy, or Si-Varma correction, can be written in terms of the density of the charge carriers and the disorder parameter as

$$\varepsilon_{s-v}(n(\mathbf{r})) = -\frac{A}{2\sqrt{2}}\left(\frac{me^2}{\varepsilon\hbar^2}\right)\frac{(\pi n(\mathbf{r}))^{-1/2}}{\omega_0\tau_0}. \tag{6.19}$$

To obtain the density distribution of the ground state the total energy functional is minimized, with respect to $n(\mathbf{r})$. Before discussing the numerical results for probing the compressibility of the disordered 2D electron system (see chapter 7), let me conclude with some general observations:

- In the clean limit the local-density approximation (LDA), discussed in previous chapter, is expected to hold when the density varies sufficiently slowly in space, i.e., as long as $k_F^{-1}|\nabla n/n| \ll 1$. Although the LDA results are exact for the uniform electron gas, very often this condition is not fulfilled as exemplified by, for instance, metallic Cu. But even in these cases, the LDA often gives good results when used in electronic-structure calculations. the Si-Varma correction to LDA still keeps the DFT local as it is the case for normal DFT in the local density approximation.

- For disordered electron system in the regime of relatively high density (diffusive regime) one expects that LDA gives reasonable results. In this regime, due to the fact that the mean free path of the electrons, ℓ, is much larger than the screening length, s, the interaction is not very strong and the electronic wave function is spread out over many disorder sites. As one goes to the low density region the electron-electron interaction, as well as the effect of disorder, become more and more important. In this regime $\ell \ll s$ so the electrons interact very strongly on the scale of impurity-impurity distance. In this situation *LDA+(strong)disorder* within the DFT framework is bound to fail.

- Eq.6.19 represents the correction to the exchange-correlation energy density considering both the effect of disorder and strong interactions when electron density is low (Si-Varma correction). It is clear that in the limit of zero disorder, the Si-Varma correction is vanishing and one recovers the normal LDA results for the clean system. At the same time for high electron density the effect of the Si-Varma is much smaller than the exchange energy and one recovers the regime of diffusive electrons. Therefore, it seems quite reasonable to consider the Si-Varma term as a correction to the LDA for the case of weakly disordered electron system in low density regions.

- Although this might be the correct physics of low density and weakly disordered systems residing at the liquid side of the MIT transition point, one has to keep in mind that it will eventually fail sufficiently close to the transition to the Wigner glass like state.

- In the low density region where the Si-Varma term takes over, as we will see in the chapter 8, density variations are actually suppressed and the electron system tends to become more homogeneous than in LDA. As a result the quantity $k_F^{-1}|\nabla n/n|$ is still small. In a remarkable self-consistent manner the theory protects its own integrity in the metallic state up to a point very close to the MIT where the experiments show the onset of density fluctuations.

6.3 Conclusions

At low enough density, the screening length of a disordered interacting electron gas may become larger than its mean free path. The compressibility, being proportional to the inverse screening length, is expected to vanish as the transition to the insulating state is approached. This extra physics gives rises to the corrections to the exchange-correlation energy of 2D electron system. Due to the fact that this correction is local with respect to the density of the charge carriers, it is possible to develop a local density functional formalism (DFT) to study the properties of such incompressible liquid near MIT phase transition. This in turn serves as a correction in the LDA physics becoming important at low densities and weak disorder potentials. In the next chapter we will explore the ramifications of this theory in a quantitative detail and will discuss the thermodynamics of such a system by looking at the behavior of electronic compressibility in the vicinity of the MIT.

Chapter 7

Si-Varma Compressibility near the Metal-insulator Transition

To test the Si-Varma theory on disordered interacting electron system, we consider a system of N interacting electrons on a two dimensional square lattice of size L, subjected to an external disorder potential. One can start from the metallic side and increase r_s to look into the evolution of chemical potential, μ, and its derivative, which is the inverse compressibility, $\partial\mu/\partial n$. The density functional description [143, 144] has the benefit that one has to find just the relationship between the ground state energy and the ground state density distribution. This facilitates matters especially considering thermodynamical properties because these are typically expressed in terms of local density at the ground state. An example is the electronic inverse compressibility associated with the second derivative of the ground state energy to the density. In this way one avoids intricate issues related to the coherence of the wave function as of importance to e.g. transport properties. The LDA is based on accurate results for the homogeneous electron gas in the absence of any perturbing potential. This can be regarded as a good description of the metallic state of the 2DES away from the MIT, since the measured compressibility is consistent with the LDA result. In the Chapter 5 we discussed the outcome of such an approach (LDA-droplet picture), focused on the behavior of compressibility at low densities. Although this approach may be a good description of 2D electron system at relatively high densities in the presence of strong disorder potential, it fails to give any reasonable prediction at low densities and weak disorder. In this regime, as we have already addressed in chapter 6, due to the strong electron-electron interactions, the screening length, (s) is much larger that the electron mean free path, (ℓ). This gives rise to an extra physics at low density regime which, as shown in previous chapter, it manifest itself in the correction to be added to the LDA exchange-correlation potential which becomes discernible upon approaching the MIT. This is not very surprising because, as discussed in previous chapter, the first order transition to the Wigner crystal is already there in clean system. So one could argue that the introduction of disorder, while staying on the metallic side,

actually turns the first order transition into a continuous one. In this chapter we investigate the consequences of the Si-Varma physics, within DFT approach, for the behavior of compressibility of disordered 2D electron system.

7.1 The DFT Energy Functional and Energy Minimization

In previous the chapter we derived the correction to DFT for disordered interacting electron system in the low density regime coming from the consideration of the fact that in this regime local screening might have been suppressed due to the presence of the strong Coulomb interaction: Si-Varma correction. To obtain the electron density distribution of the ground state, the total energy functional $E[n(\mathbf{r})]$ is minimized, $\delta E[n] = 0$, with respect to $n(\mathbf{r})$ under the constrain that the number of electrons is kept constant. Having the total energy and the local density distribution of the ground state, we obtain information on the bulk (compressibility and chemical potential) as well as local properties (charge distribution and the scale of inhomogeneity) of the 2D electron system.

7.2 Numerical Results for μ and κ

After discretization of the total energy functional Eq.6.11, the simulation is carried out on a system of N electrons on a square lattice of size 50×50. The interactions are fixed by the parameter $U/2$, hence r_s is only depends on the density of the particles. Ewald sum (see appendix B) and periodic boundary conditions are applied to minimize the finite size and surface effects. The output of the simulation is the density distribution and the total energy for different densities and different disorder strength. Two different type of disorder landscape has been used, uncorrelated disorder $< V_d(\mathbf{i})V(\mathbf{j}) > \sim \delta_{\mathbf{ij}}$ and correlated disorder with exponential fall-off $< V_d(\mathbf{i})V(\mathbf{j}) > \sim exp(-\frac{|\mathbf{i}-\mathbf{j}|}{\xi})$, where ξ is the correlation length of the disorder potential. For better understanding, in this chapter we present the results for thermodynamical bulk properties like chemical potential, μ, and compressibility, κ, while in the next chapter we discuss the local properties like local density distribution ($n(\mathbf{r})$), and other properties like the size and the number of electron (hole) droplets in the system.

The chemical potential μ and compressibility κ are calculated by finite differences in terms of the total particle number N using

$$\mu(N) = < E(N+1) - E(N) > \tag{7.1}$$

and

$$\kappa^{-1} = \frac{N^2}{V} < \frac{\partial^2 E(N)}{\partial N^2} >, \tag{7.2}$$

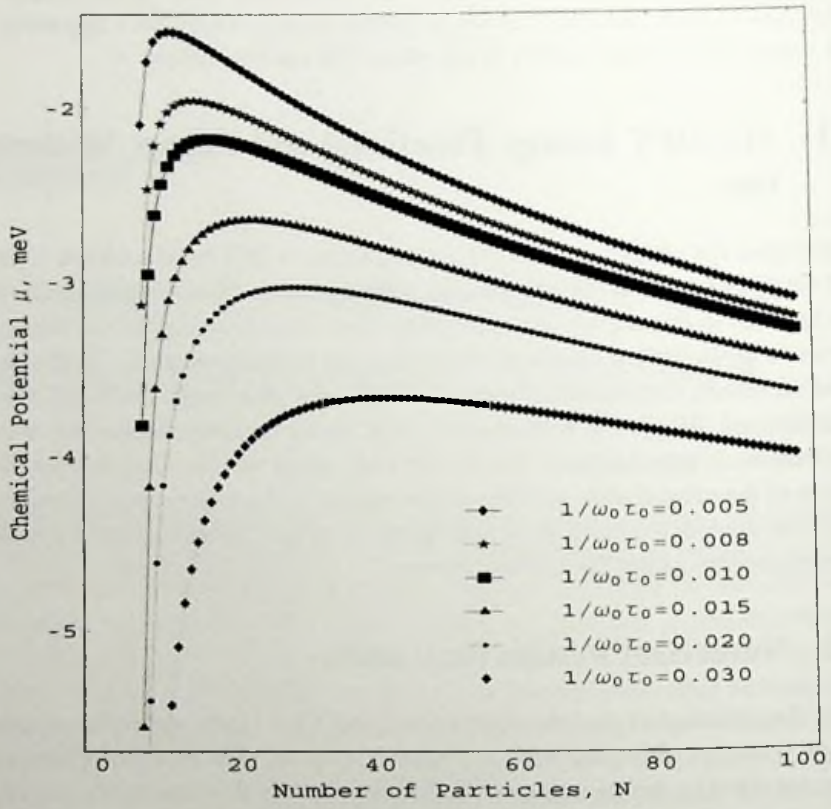


Figure 7.1: Numerical results for the chemical potential of 2D electron system for six different (uncorrelated) disorder strength. The maximums in this graphs correspond to the density where the compressibility of electron system diverges.

or more precisely

$$\kappa^{-1} = \frac{N^2}{L^2}\langle E(N+1) + E(N-1) - 2E(N)\rangle \tag{7.3}$$

where $V = L^2$ is the volume of the system and the brackets denoting the average over disordered samples.

In Fig.7.1 the simulation results for the chemical potential for six different value of disorder strength is presented. It is very crucial to mention that, these results are for uncorrelated disorder potential. Looking at the behavior of μ as a function of the density of charge carriers it is clear that - for typical disorder strength, starting from the metallic side, μ increases as density is lowered - a behavior that corresponds to negative compressibility. This behavior persists down to a

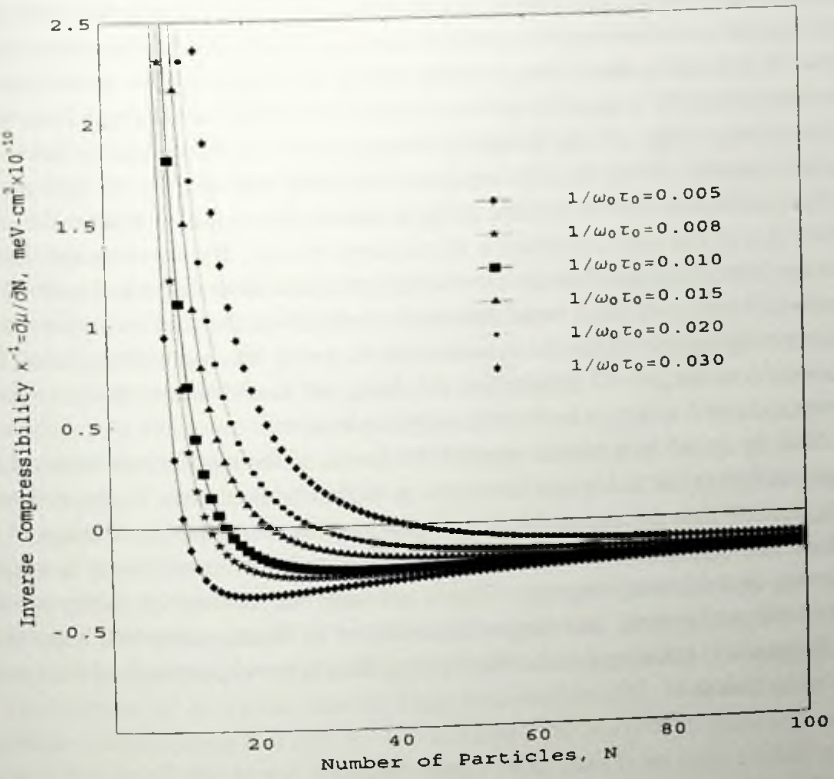


Figure 7.2: Numerical results for inverse compressibility, $\partial\mu/\partial n \propto \kappa^{-1}$, as a function of density of charge carriers. Six different results for different (uncorrelated) disorder strength are presented. The point where $\delta(\kappa^{-1})/\delta n = 0$, is the point in which the electron-electron interaction is dominated and Si-Varma become more important than exchange.

critical density $n_{1c}(1/\omega_0\tau_0)$. This is the critical density usually corresponds to the critical density coming from transport measurements. When the density is further lowered, μ starts to decrease rapidly, indicating positive and vanishing compressibility. Such behavior of chemical potential as a function of density is in very good agreement with the experimental data [15, 150]. Fig.7.2 shows the corresponding behavior of the inverse compressibility (or incompressibility), κ^{-1}. As discussed in previous chapters, compressibility reflects the ability of an electronic system to screen external charges. When an external charges is brought to the proximity of electronic system, mobile electrons accumulate in its vicinity to screen it. Hence a system that cannot redistribute its charge is incompressible. Now looking at the behavior of incompressibility in Fig7.2 two main features are immediately notice-

able. First, for certain values of disorder strength, $1/\omega_0\tau_0$, $1/\kappa$ is negative for high densities and becomes more negative as the density is decreased as has been seen experimentally in dilute electron and hole systems. The negative compressibility is known to to be due to the strong exchange energy contribution to the total chemical potential. Secondly, a sharp turn-around occurs at a critical density $n_{2c}(1/\omega_0\tau_0)$ as $d\kappa/dn$ changes sign. As the density is further reduced, $1/\kappa$ becomes positive and diverges rapidly. From this behavior, one can argue that starting at high densities the disordered electron system can be considered as a quite compressible fluid mostly due to the large contribution of exchange energy. By lowering the density both the interaction and disorder effects become much more important and for the electrons it becomes much more difficult to redistribute themselves to screen the Coulomb interaction. Thus the system turns to a very incompressible fluid. This is basically is the general idea behind the theory of Si and Varma as discussed in previous chapter, reflected in these quantitative results.

Now let us see how we can explain the results of the simulations within LDA theory including the Si-Varma correction. In the LDA calculation, the local density (n_{local}) rather than the average density, is used to determine the total energy of the system. Let us assume that for the electron fluid system this local density is roughly the same as the average density. This is not very bad assumption since we will find in the next chapter that the inclusion of the Si-Varma correction suppresses inhomogeneity. Ignoring density fluctuations, the chemical potential of the system can be written as

$$\mu(n) = d(\varepsilon(n)n)/dn,$$

where ε is the energy density for uniform electrons incorporates also the fact that in the low density regime the screening length is very large: Si-Varma correction. This energy can be expressed as

$$\varepsilon(n) = \varepsilon_0(n) + \varepsilon_{Si-Varma}(n), \tag{7.4}$$

where

$$\varepsilon_0(n) = \frac{\pi\hbar^2 n}{2m^*} - \frac{4}{3}\sqrt{\frac{2}{\pi}}\frac{e^2}{4\pi\varepsilon}n^{1/2} + \varepsilon_{corr}(n), \tag{7.5}$$

and $\varepsilon_{Si-Varma}$ is the Si-Varma correction derived in previous chapter

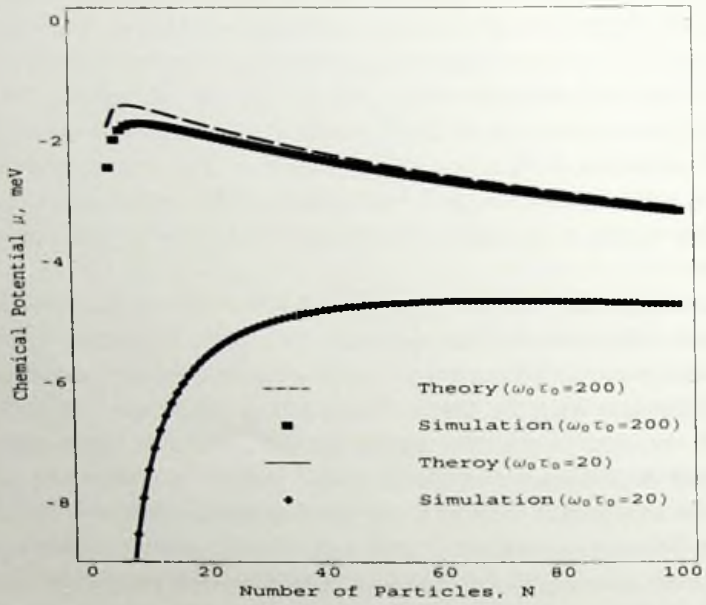
$$\varepsilon_{Si-Varma}(n) = -\frac{m^* e^2}{\varepsilon\hbar^2}\sqrt{\frac{2}{\pi}}\frac{n^{-1/2}}{\omega_0\tau_0}, \tag{7.6}$$

where $\omega_0\tau_0$ controls the amount of disorder in the system in a presence of strong Coulomb interaction, controlling the strength of the Si-Varma term. For comparison, we calculate the chemical potential, μ, and its derivative, $d\mu/dn$ with respect to the average density numerically. The results for inverse compressibility, $\kappa^{-1} \sim d\mu/dn$, as follows from the simplified analytical considration is shown in Fig.7.3 in comparison to the data obtained from the simulation. In this figure, the

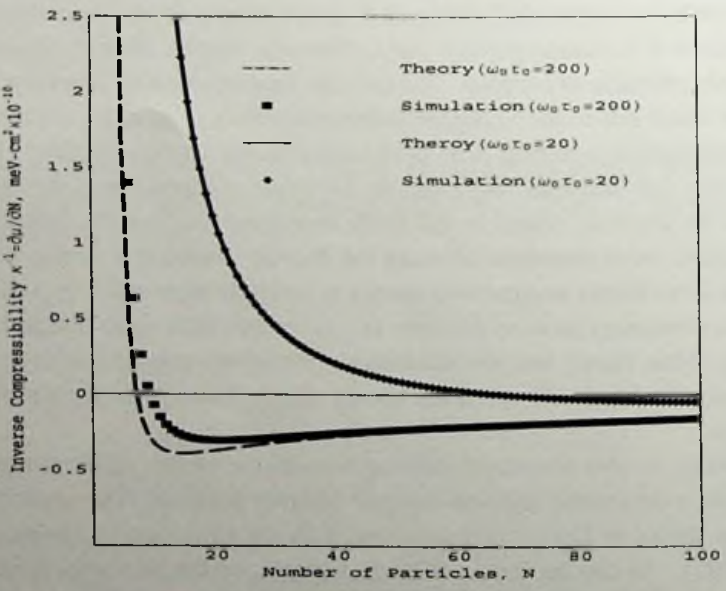
incompressibility as a function of charged carrier number is shown for two different cases: low mobile electrons (high disorder) and highly mobile electrons (low disorder). Although there is a good agreement between theory and simulation in both regime (strong and weak disorder), one can easily see small differences between theory and simulation in the case of highly mobile electrons. This deviation from the theoretical prediction, in fact, is not very surprising. The above calculation is based on the fact that the electron gas is homogeneous which is, of course, not the case for electron system subjected by disorder potential, even in moderately low density regime.

Starting from high densities (low r_s) and low disorder strength, electrons do not see the potential valleys and the most dominant force is the long-range Coulomb force between electrons. So the electron system tends to be quite homogeneous and there is no contribution from the kinetic energy adding the behavior of the compressibility. At the same time in this regime the effect of the Si-Varma potential is very small and the dominant contribution comes from the exchange and correlation energies. Consequently there is a very good agreement between the theory and simulation. Moving to moderate densities the disorder potential starts to play a role and one finds an interplay between long-range Coulomb energy and the disorder potential. In this case the electrons are getting trapped into disorder potential valleys and the system starts to become inhomogeneous. Both in the theoretical and in the simulation curves the turning up of the incompressibility is coming from the contribution of Si-Varma physics (large screening length). However there is a deviation between these two curves. The fact that there is a density inhomogeneity (even is it is small due to the Si-Varma energy) give rises to an extra contribution into the compressibility coming from both Coulomb and kinetic energies. Finally in case of very low densities (high r_s), the Si-Varma contribution is the dominant one and the electron system is still fairly homogeneous. For the case of low mobile electrons (high disorder), although the disorder potential is strong, but the contribution of Si-Varma into the total energy is much stronger and it tries to keep the electrons as homogeneous as possible and as a result there is no disagreement between the above theory and the simulation. Somewhat counterintuitively, the greatest error associated with using the energy density takes place at "intermediate" densities.

Furthermore we also performed separate simulations for two different types of disorder, namely *correlated* and *uncorrelated* disorder potential. The result of the simulation is shown in Fig.7.4 in comparison with the LDA data (= ignoring the Si-Varma term). As can be seen, not much happens on the Si-Varma level and both data are basically the same for different disorder potential, while for LDA for uncorrelated disorder the system behaves very similar to non-interacting electrons. This difference in the behavior of the compressibility for LDA physics is very important because the turning up of the incompressibility in this case, as has been seen also in chapter 5, coming from the very special nature of disorder potential. In order to reproduce the upturn in the incompressibility, which is in fact associated with droplet formations, this potential has to be weakly confining with



a)



b)

Figure 7.3: a) Chemical potential, μ, and b) inverse compressibility (incompressibility), $\kappa^{-1} \sim d\mu/dn$, as a function of density of charge carriers. Red and blue curves correspond to the theoretical calculation explained in the text based on Si-Varma. Theoretical and simulation results are given for both highly mobile (low disorder), $\omega_0\tau_0 = 200$,and less mobile (high disorder), $\omega_0\tau_0 = 20$, electron (hole) system, for better comparison of theory and simulation.

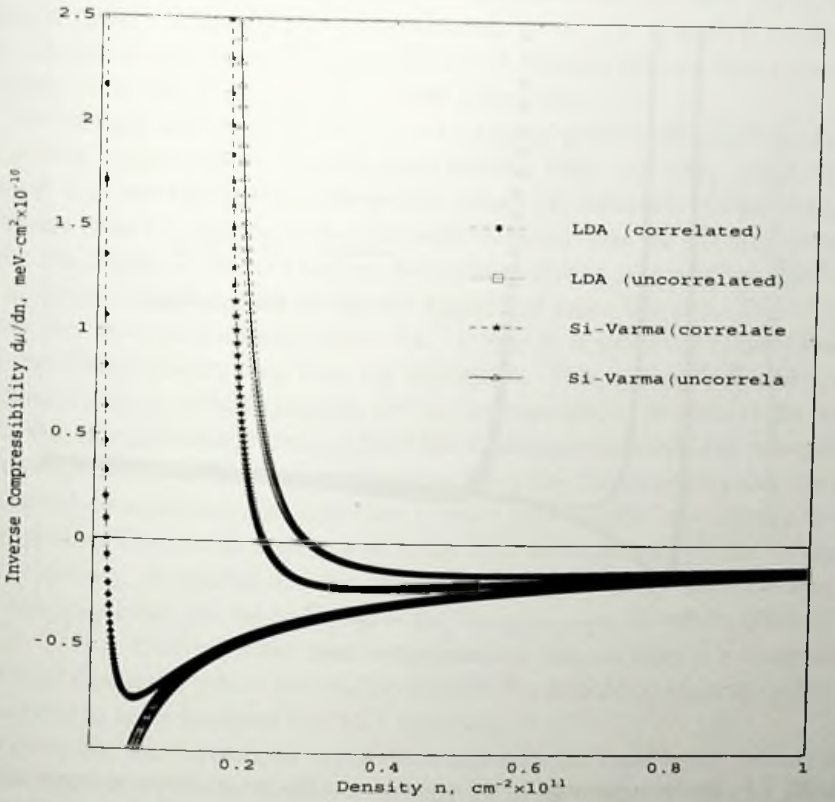


Figure 7.4: Inverse compressibility, $\kappa^{-1} \sim d\mu/dn$, as a function of density of charge carriers for typical disorder strength of $(\omega_0\tau_0)^{-1} = 0.01$ for both LDA and Si-Varma theory.

relatively large correlation length. It is clear that for the local disorder potentials with stronger confinement effect than infinite harmonic potential the pure LDA treatment is not applicable at all! This is at the same time related to the very fact that, in all percolation pictures of metal-insulator transition one needs to start from a metallic phase which is inhomogeneous on some scale. But as has been seen in the chapter 4, the experimental evidences [14] suggest that this might not be the case. This is another example of the fact that there has to be more in physics of MIT in disordered interacting electrons than just the percolation/droplet picture of the LDA treatment, which is at the end just put in by hand in the form of fine tuned correlated disorder. For uncorrelated disorder these droplets just not form on the LDA level.

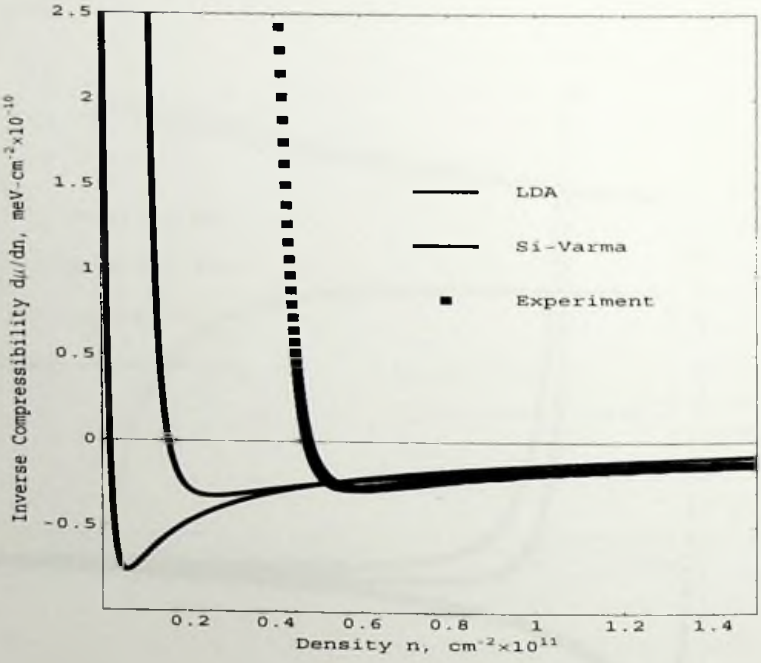


Figure 7.5: Inverse compressibility, $\kappa^{-1} \sim d\mu/dn$, as a function of density of charge carriers. Red and blue curves correspond to the prediction based on Si-Varma and (pure) LDA result respectively. Although Si-Varma gives a very good qualitative behavior of compressibility of disordered electron system but the critical density where $\delta(\kappa^{-1})/\delta n = 0$, is about three times smaller than the corresponding critical density in experiment.

7.3 Does Si-Varma Explain The Experiment?

Let us compare the results of our simulations with experiment to see to what extent the physics emerging from the Si-Varma correction can explain the experiment. For this purpose we use the experimental data for thermodynamical compressibility of highly mobile 2D hole system in $GaAs/Al_xGa_{1-x}As$, measured by Dultz and Jiang [13], which has been discussed in Chapter 4. The results are given in Fig.7.5, where also include the LDA prediction. For the best comparison we used the data corresponding to the simulation of the 2D electrons in the presence of correlated disorder potential. As we saw in previous section , the behavior of incompressibility does not depend on the type of the disorder potential using the Si-Varma exchange-correlation potential. We used a correlation length of the local disorder $\xi = 40a_B$, where a_B is the Bohr radius. The values for a_B^* and $\omega_0\tau_0$ for this fit are $a_B^* = 2.3 \times 10^{-7}$ and $\omega_0\tau_0 = 100$.

We notice that the fit to the experimental data is the best we could find using

Si-Varma potential. There in fact two parameters which one can varry: the disorder strength, $(\omega_0\tau_0)^{-1}$, and the effective interaction, $U^* = U/\varepsilon$. Where ε is the dielectric constant of the system which is different in different electron (hole) systems. The later parameter is actually fixed by the experiment.

As one can easily see, Si-Varma gives a quite good qualitative behavior of the inverse compressibility starting from metallic side. The point where $d\mu/dn$ changes sign corresponds with the density where the correction coming from the Si-Varma theory is starting to become more dominant than the exchange energy. When the density is lowered further, the electron system cannot screen itself and the screening length s starts to become larger than mean free path. Fig.7.5 also shows that the critical density where $\delta(\kappa^{-1})/\delta n = 0$, is about three times smaller that the critical density seen from the experiment. This is directly related to the fundamental form of the expression for the incompressibility in terms of the density of the charged carriers coming from the Si-Varma correction. The strength of the divergence at low densities coming from this correction is smaller than the experimental observations, although there is much improvement as compared to the LDA picture. This can be understoond easily from the expression for the Si-Varma energy density. Assuming that one can use an averaged density, the correction to the compressibility due to Si-Varma is $(\kappa^{-1})_{Si-Varma} \sim n^{-3/2}$ which diverges at $n = 0$. This is clearly not the case in experimental data, as there is a finite value of critical density n_c where for smaller densities the disordered electron system is considered to be an insulator (the MIT transition).

Given that the fluctuations of the electron density are quite weak, one can obtain accurate analytical estimates of the behavior of the compressibility by assuming a uniform density. The LDA correlation energy contributes only weakly to the compressibility and can be neglected in first instance. What remains are the contributions from the kinetic-, exchange- and Si-Varma energy,

$$\begin{aligned} \frac{\partial\mu(n)}{\partial n} &= \frac{\partial^2}{\partial^2 n}\left(E_{kin} + E_{ex} + E_{SV}\right) \\ &= 2\pi a_B^2 - \sqrt{\frac{1}{2\pi}}a_B n^{-1/2} + \frac{1}{\omega_0\tau_0}\sqrt{\frac{1}{8\pi^2}}\frac{1}{a_B}n^{-3/2}. \end{aligned} \tag{7.7}$$

The contributions from the Fermi pressure (first term), the exchange (second term) and Varma-Si (last term) are recognized and it is clear that the last one will dominate the compressibility at sufficiently small density. First of all it follows immediately that the minimum of $1/\kappa$ occurs at a 'critical' density,

$$n_c = \frac{3}{2\pi}\frac{m}{\hbar\tau_0} \tag{7.8}$$

i.e. when the mean-free path becomes of order of the Bohr radius. This obviously gives an upper limit for the critical density n_c for a given amount of disorder. Secondly it is almost impossible to fit the data based on Eq. 7.7 even when one uses unrealistic and exaggarated choice for parameters a_B and $\omega_0\tau_0$. As just mentioned,

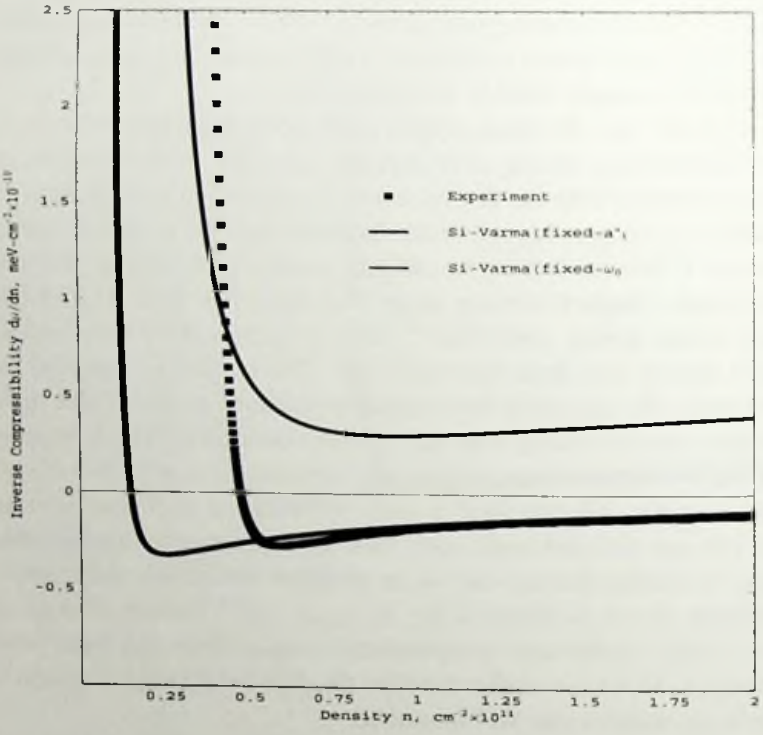


Figure 7.6: Different theoretical fits to the experimental data using different set of parameters (see the text for details).

this is related to the fundamental form of the incompressibility expression within Si-Varam. The existence of a critical point, n_c, at a finite density, rather than $n_c = 0$ is very evident from the experimental observation.

To address this problem in more detail we proceed as following: let us first start from the metallic side in high density (low r_s) region. Dultz *et al.* [13] used p-type MBE grown GaAs/$Al_xGa_{1-x}As$ single heterostructure for their compressibility measurement. The mobility of this particular sample was roughly $\mu = 1.23 \times 10^4 cm^2/V\text{-}s$ with a hole density of $p = 2.60 \times 10^{11}$ cm^{-2}. Using the relationship between the scattering time, τ_0, and hole mobility μ (not the chemical potential!)

$$\tau_0 = m_p \mu \tag{7.9}$$

where m_p is the mass of the holes in the system, one can estimate the scattering time τ_0 or the disorder strength, $1/(\omega_0\tau_0)$, in this system which is roughly $1/\omega_0\tau_0 = 7 \times 10^{-3}$. On the other hand, in this experiment the transport critical density is $p_c = 5.5 \times 10^{10} cm^{-2}$, where the dimensionless interaction strength is about $r_s \approx 13$.

This in turn yields the estimation of the effective Bohr radius in the system from the relation $r_s = 1/(a_B^*\sqrt{n\pi})$; which gives the value of effective Bohr radius as $a_B^* \approx 2 \times 10^{-7}cm\approx 0.37 a_B$. Now we can fixed one of these parameters (a_B^* and $(\omega_0\tau_0)^{-1}$), to attempt to fit the experimental data by changing the other one. The results of this attempt is shown in Fig. 7.6. It is quite clear that the fixed-a_B^* curve does not explain the data close to p_c even qualitatively. In fact it is impossible to obtain a good qualitative fit by only changing the disorder strength. The fixed-a_B^* fit shown in Fig.7.6 can be achieved only if we chose the value of effective Bohr radius as $a_B^* = 1.5 \times 10^{-8}$cm, which is about an order of magnitude smaller than the actual value of this quantity in the system. Upon fixing the quantity $\omega_0\tau_0$, it is easier to obtain reasonable fit (fixed-$\omega_0\tau_0$ curve). For densities up to and around p_c^{exp} the fit is almost exact which is not very surprising since, as we saw in the previous chapter, Si-Varma is indeed a correction to LDA for low densities (apparently no too low!). But still this curve gives a too low critical density (where compressibility changes its behavior) as compared to the experiment.

In fact this is not very surprising as if we look more carefully into Eq.7.8. This expression for the transport critical density (where $\frac{\partial}{\partial n}(\kappa^{-1}) = 0$) gives a very small critical hole density using the fixed parameter for the hole system of Dutlz *et al.* experiment [13]. Using $a_B^* = 2 \times 10^{-7}$ and $(\omega_0\tau_0)^{-1} = .007$ (coming from the experiment), we obtain critical density $p_c \approx 2 \times 10^{10}$cm^{-2} which is too small compare to the experimental critical density. More interestingly, if we use the experimental critical density $p_c = 5.5 \times 10^{10}$cm^{-2}, and a disorder strength $(\omega_0\tau_0)^{-1}_{exp} = .007$ as an input parameter, using Eq.7.8, we have $a_B^{*theory} = 1.5 \times 10^{-14}$cm. The above considrations emphasize, at least, two important facts: first, the compressibility behavior, which has been observed in the experiment, has much faster divergence than the theory and second, it might be more in the physics of 2DHS near critical point which may reveal itself in the anomaly of inverse compressibility at low densities. A very small value of a_B^* in liquid phase could be related to the very large value of hole effective mass through the relation:

$$a_B^* = \frac{\varepsilon\hbar^2}{m^* e^2} = \varepsilon(\frac{m}{m*})a_B. \tag{7.10}$$

In Chapter 9, we will address this problem in more detail in the context of the possibility of the existence of Quantum Phase Transition (QPT) in the 2D electron/hole system.

7.4 Conclusions

Physically it has to be that at low densities and small disorder strength, due to the fact that the interactions are very strong, the electron system cannot screen itself in a perfect fashion, and this give rise to an extra physics at work near MIT transition in 2D disordered electron(hole) system. This extra physics manifest itself in corrections to the exchange and correlation energies at low densities, du to

the effect of disorder: Si-Varma theory. Since this correction to the ground state energy is local, it is possible to develop a DFT formulation and we showed that in fact one has to consider it as an necessary correction to LDA at low densities. Numerical calculations of such a DFT formulation for 2D electrons indicate a substential improvement of the description of the behavior of compressibility near MIT transition point. The bad news is, however, that the "critical" density where the compressibility changes its behavior implied by these calculations is smaller than the experimentally observed critical density by a factor of three. This in turns can be related to the fact that the theory suffers from an intrinsic flaw that the divergence of the inverse compressibility takes place at $n = 0$. This it is not the case in the real system; the Wigner crystal (or glass) phase where the compressibility is vanished nearly completely takes over at finite densities. We will address this problem and a possible solution in the last chapter (Chapter 9).

Chapter 8

Local Properties of the 2D Electron System

In the previous chapter we have discussed the outcome of the DFT-Si-Varma formalism, with respect to the behavior of the electronic compressibility of the 2D electron system. In that chapter we also compared our result with the LDA-droplet picture of the MIT. As we discussed earlier (see chapter 5), the scale of density inhomogeneities in the liquid phase is crucial for the droplet picture to give any sensible result. In this scenario this is taken care off by choosing a strong and, at the same time, correlated disorder potential. On the other hand, within the Si-Varma theory, we have showed that the correct behavior of the compressibility (as compared to the experiment) near the MIT does not need a strong disorder potential and neither special kinds of correlations between the disorder sites. We believe that -at least compared to the LDA droplet picture- Si-Varma gives a qualitatively correct physical picture for the compressibility.

Therefore it would be very interesting to see what the predictions are of this theory as related to the local properties of the 2D electron system near the MIT. This is very important specially in connection with the recent experiments on the local compressibility measurement of the 2D hole system [15](see also chapter 4). In this chapter we are going to address the issues related to density inhomogeneities, their characteristics and their possible relationship to the electrostatics of 2D electron system in the expressing the local compressibility fluctuations.

8.1 Density Distribution at Ground State

The density distribution of the ground state of the 2D interacting electron system, $n(x,y)$, is a distribution of charge carriers at which the total energy functional, $E_{tot}[n]$, of the DFT formalism (explained earlier in previous chapters) is minimum. Fig.8.1 shows the local density distribution, for an average density corresponding to $r_s = 25$, from numerical simulations based on both LDA and Si-Varma theory (see chapter 5 and 6 for details). It is quite clear that the outcome for LDA is the

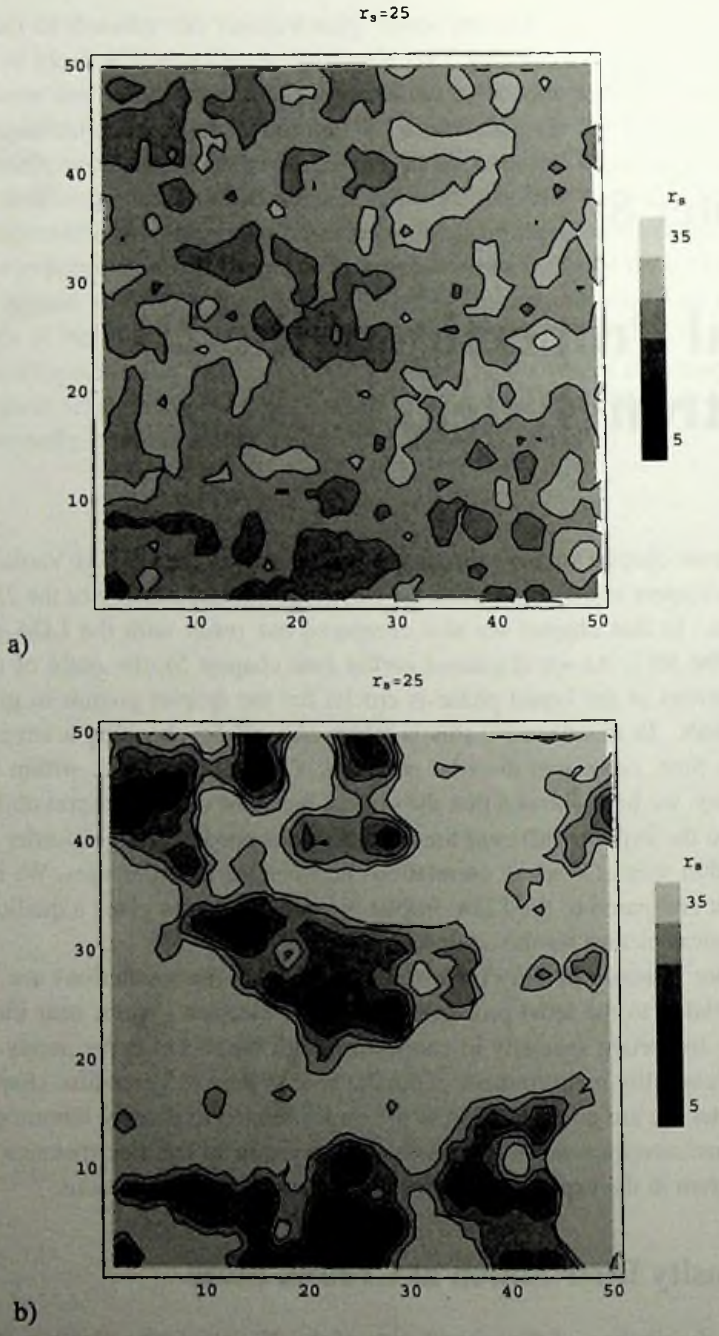


Figure 8.1: Density distribution of the charged carrier in the system for the fixed average density $r_s = 25$ for a) Si-Varma physics ($\omega_0\tau_0 = 150$) and b) LDA physics ($\omega_0\tau_0 = 1.5$).

droplet picture of Shi-Xi (explained thoroughly in chapter 6) characterized by high density ($r_s \sim 1$) metallic droplets immensed in a very low density ($r_s > 40$) background, Fig.8.1b). Including the Si-Varma potential one finds instead a quite weak density variation, Fig (8.1a), typically of the order of 25 % around the average. In part, this is due to the much smaller disorder potential required to reach the vicinity of the MIT ($\partial(\kappa^{-1})/\partial n = 0$) when the Si-Varma term is included. However, the suppression of density fluctuations is also of an intrinsic nature.

The typical energy cost due to a density fluctuation δn is given by

$$\delta E \sim (\frac{\partial^2 E}{\partial n^2})(\delta n)^2 + \cdots \sim \frac{1}{\kappa}(\delta n)^2 + \cdots \tag{8.1}$$

In the LDA case, $1/\kappa$ is negative due to the dominating exchange energy with the effect that density fluctuations are amplified; after subtracting the uniform Coulomb energy the exchange interaction acts as an attractive force. The Si-Varma term expresses that the screening length becomes large in the fluid, rendering it increasingly incompressible with the obvious effect that density fluctuations are suppressed.

In fact, it appears that experimental evidence is already available demonstrating that the typical density fluctuation in the vicinity of the MIT of a strongly correlated 2DES is consistent with the Si-Varma prediction, disproving the extreme LDA-droplet picture. We refer to the experiment by Ilani *et al.*[15] discussed in chapter 4, where the typical local compressibility fluctuations are measured directly using single electron transistors (SETs). At first sight these fluctuations appear to be quite large but one should take into account that these are most pronounced in the regime where the inverse compressibility is strongly varying as function of the density, namely very close to the MIT. Assuming that the local compressibility variation can be directly related to the local density, which is the case when the Si-Varma functional is adequate, we can estimate the typical density fluctuation. In the experiment [15] a very small change in local density (using top-gate voltage) gives rise to a very strong change in local incompressibility ($\partial\mu/\partial n$) from negative to positive values. The typical variation of local incompressibility close to the transition point is roughly $\Delta(\frac{\partial\mu}{\partial n}) \approx 7 \times 10^{-10}$meV-cm^2. Assuming that the physics of the inverse compressibility near the MIT is governed by the strong variation (upturn) of $\partial\mu/\partial n$ like the experimental curve in Fig. 7.5 (see also Ref. [13, 14]), we find that in fact the typical density variation is indeed of order 25%, consistent with our calculations (Fig. 8.1a). We notice that the droplet picture is only incorrect in the strongly correlated liquid. For r_s's of order 1, the influence of the Si-Varma term diminishes and according to our calculations the droplet picture becomes quite adequate, in agreement with recent measurements in the weakly interacting systems [137] (see also the discussion at the end of chapter 5).

8.2 Properties of Droplets

Another interesting feature of the local properties of the system under consideration is the geometrical properties of droplets of high(low) density regions: development of their size and their density as a function of the average density of charge carrier, n. To investigate this we first try to map the discretized local density distributions of figure 8.1, $n_{xy} = n(r_x, r_y)$, into a kind of binary representation as follows :

We take the density distribution at each site, n_{xy}, which we "image enhance" by taking 1 ($n_{xy} = 1$) if this density is larger than or equal to the average density, n, ($n_{xy} \geq n$); and 0 if this density is smaller than the average density ($n_{xy} < n$). Fig.8.2 shows the outcome of this conversion for the density distribution of both the LDA and the Si-Varma of Fig.8.1, for $r_s = 25$ (low density) as well as for $r_s = 10$ (high density). The white and black shows the high density (low r_s) and low density (high r_s) areas, respectively. According to the percolation picture of LDA, when the density of the charge carrier is decreasing, the low density droplets/islands is start to grow in size and these are getting connected to each other untill one reads the percolation threshold. This in turn means that the net number of these regions should also decrease. This is very evident from Fig. 8.2 going from $r_s = 10$ (Fig.8.2b) to $r_s = 25$ (Fig.8.2d) for LDA. On the other hand, looking at the same figure one can easily see that this is not the case including Si-Varma. For $r_s = 10$ there is not much difference between the distribution of low(high) density droplets coming from Si-Varma and LDA calculations, but clearly this distribution is not the same at the lower density $r_s = 25$. What is clear from this picture is that Si-Varma makes the droplets to break up and one finds a distribution chractarized by a smaller droplet size. This is very interesting because experimentally this fragmentation of 2D electron system has been seen close to the MIT critical density [15](see chapter 4).

To quantify this difference, let us calculate the average number, N_w, and average size, l_w, of these droplets as a function of density, n, using the simulation data for the density distribution, n_{xy}. Assuming an average circular shape for these regions we can proceed as follows:

The total area and the total circumference covered by droplets is given by

$$A_{total} = N_w A_w = \pi N_w R_w^2 = \frac{\pi}{4} N_w l_w^2, \tag{8.2}$$

and

$$C_{total} = N_w C_w = 2\pi N_w R_w = \pi N_w l_w. \tag{8.3}$$

where A_w and C_w are the area and the boundary of each droplet, respectively. Using the above equations we can define the following estimates for N_w and l_w

$$l_w = \langle \frac{4A_{total}}{C_{total}} \rangle, \tag{8.4}$$

and

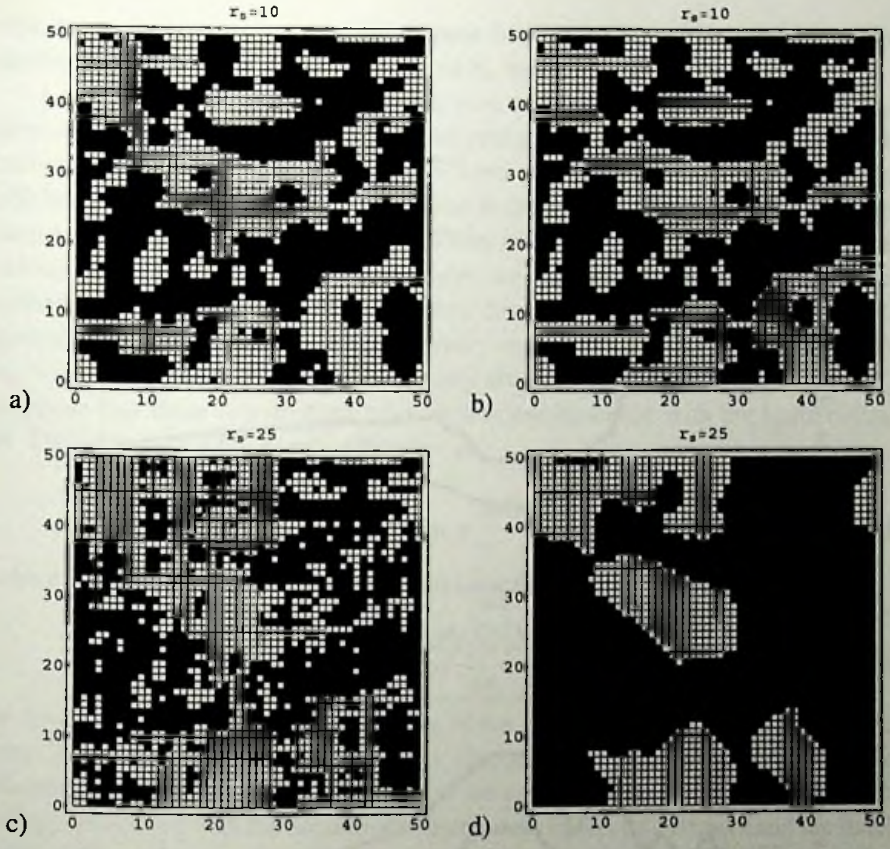
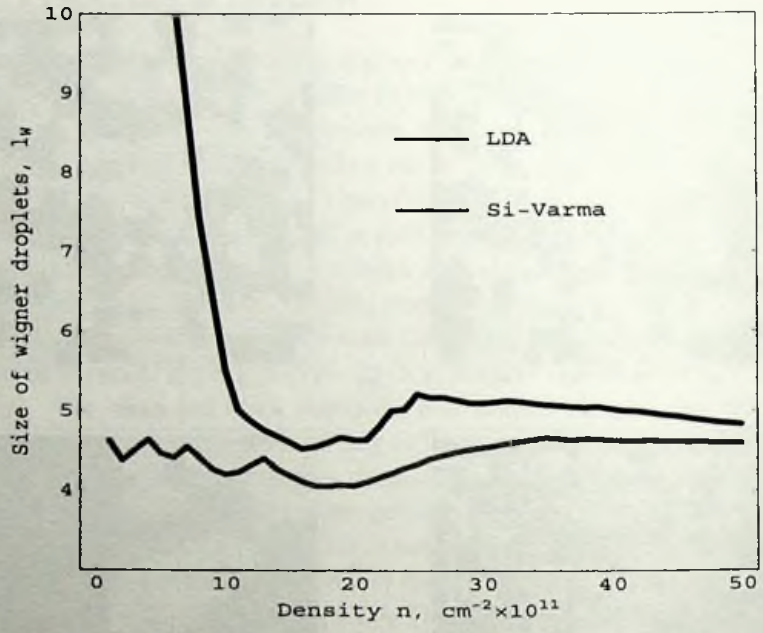


Figure 8.2: Binary representation of local density distribution of the charged carrier in the system for the fixed average density $r_s = 10$: a) Si-Varma, b) LDA and $r_s = 25$: c)Si-Varma, d) LDA.

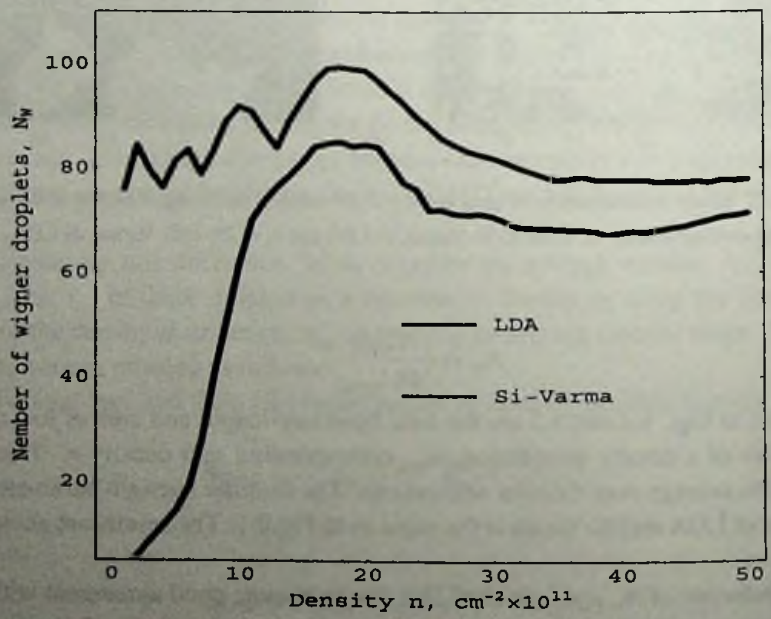
$$N_w = \langle \frac{C_{total}}{4\pi A_{total}} \rangle. \tag{8.5}$$

The inputs in Eqs. 8.4 and 8.5 are the total boundary-length and area of low density droplets of a density distribution, n_{xy}, corresponding to a density n. The $\langle ... \rangle$ indicate the average over disorder realizations. The disorder strength parameter for each case of LDA and Si-Varma is the same as in Fig.8.1. The results are shown in Fig. 8.3

The behavior of N_w and l_w in the LDA case is in quite good agreement with the percolation scenario: by decreasing the average density n the (isolated) droplets of low density area are growing in size, while upon percolating into the high density regions, their average number is decreasing, because individual drolpets merge. The situation for Si-Varma is quite different. The behavior of N_w and l_w is more ore less the same as in LDA until the densities are reached where Si-Varma physics starts to take over. After this point their evolution stays almost unchanged. It seems



a)



b)

Figure 8.3: Number and size of the low density droplets, N_w and l_w, as a function of the average density of the charge carrier in the 2D electron system under consideration. Results for both LDA and Si-Varma is shown. We used the same disorder parameters as in Fig. 8.2.

that due to the Si-Varma energy term, these droplets cannot grow any more at low densities. How to explain this behavior of N_w and l_w in the case of Si-Varma?

Let us start from the 2D electron system with an average density of n, compensated by a uniformly charged positive background with charge density en. We assume that the total average density is fixed (for the moment) and that we can adiabatically turn on the disorder potential in the system. We start with a uniform density distribution in the absence of disorder potential. Now assume that for some reason we are in a region of densities where two droplets can be embedded into the system in the clean limit. Let see whether this phase separation is possible energetically. The first droplet is the low density region with the charge concentration $n_w = n - \delta n_1$ and the second is high density droplet with the density $n_L = n + \delta n_1$. Suppose that these two droplets have circular configuration with the same radius R. The estimate for the energy difference is

$$\Delta E \sim (\mu_L - \mu_W)\pi R^2 \delta n_1 + \frac{(e\delta n_1 \pi R^2)^2}{R} + \pi R E_s, \tag{8.6}$$

where E_s is the surface energy density and μ is the chemical potential defined as

$$\mu_{L(W)} = \left(\frac{d(\varepsilon_{L(W)} n)}{dn}\right)\Big|_n \tag{8.7}$$

in which $\varepsilon(n)_{L(W)}$ is the energy density of the high(low) density area. The first term in Eq. 8.6 correspond to the energy change due to phase separation and it is linearized with respect to δn_1. The second one correspond to the positive Coulomb energy associated with the nonuniform distribution of electron density and the third term is the surface energy. Now minimizing ΔE with respect to δn_1 gives us the value of δn_1 associated with size of the droplets as $\delta n_1 = (\mu_W - \mu_L)/Re^2$ or the size of the droplet

$$R = \frac{\Delta\mu}{\delta n_1 e^2}. \tag{8.8}$$

This in turn gives the minimum energy change of

$$\Delta E_{min} = (2\pi E_s - \frac{(\Delta\mu)^2}{e^2})R. \tag{8.9}$$

Now in the clean limit $((\omega_0\tau_0)^{-1} = 0)$, $(\mu_W - \mu_L) = 0$ and $\Delta E_{min} = 2\pi E_s$, where $E_s = \sigma > 0$, which means that the energy change is always positive and there is no creation of droplet. If we turn on the disorder potential adiabatically there will be a minimum disorder strength where which this energy difference becomes zero and beyond this critical disorder strength the system favors inhomogeneities. Using LDA one can write μ in linear approximation for small δn_1. One finds that the quantity $\Delta\mu = (\mu_W - \mu_L)$ is very small and therefore the size of the droplets is governed by the correlations between the disorder sites. Now suppose we decrease the average density by a small amount of $\delta n_2 \ll \delta n_1$. Now the change in the size of droplets becomes

$$\delta R_{L(W)} = \pm(\frac{d\Delta\mu}{dn})(\frac{\delta n_2}{\delta n_1}). \tag{8.10}$$

The quantity $d(\Delta\mu)/dn$, which is also the change in the electrostatic pressure, is proportional to the inverse compressibility of the system, $\Delta\kappa^{-1}$. In the case of LDA this quantity is always negative and one can easily see that the size of the low (high) density droplets increases (decreases) by decreasing the average density n and this process is also favorable energetically as it can be seen from the Eq. 8.9. Now assume that we are close to the densities where the Si-Varma physics is matter: the electrons cannot screen themself due to large screening length with respect to the mean free path of the electrons. This extra physics in turn manifest itself in an extra (positive) correction into the incompressibility of the system, $(\kappa^{-1})_{SV} \propto (1/\omega_0\tau_0)n^{-3/2}$. For fixed value of disorder strength it is clear that this term is growing upon decreasing the average density. Around the density where Si-Varma physics is dominating, the droplets reach the maximum size and stop growing when one changes the density . This behavior is very evident from Fig. 8.3. This can also be understood by looking at the change of the minimum energy difference ΔE_{min} due to the change in average density (δn_2)

$$\delta(\Delta E_{min}) \sim (\frac{\Delta E_{min}}{R} - \Delta\mu)\Delta\kappa^{-1}\delta n_2 \tag{8.11}$$

8.3 Electrostatics and Phase Separation is 2D Electron System

Let us coclude this chapter with a brief discussion of the electrostatics of the 2D electron system realized in experiment and its relationship to the phase separation and density inhomogeneities within Si-Varma physics. The motivation comes from the interesting suggestion by Spivak [38] discussed in chapter 2. He argues that the existence of an intermediate phase between the Fermi liquid and the Wigner crystal phases is a generic property of the two-dimensional *pure* electron liquid in MOSFETs at zero temperature. The physical reason for the existence of these phases is a partial separation of the uniform phases. This phenomenon is due to a tendency for phase separation which originates from the existence of a first-order phase transition between the Fermi liquid and the Wigner crystal phases as a function of n, which is enhanced as a result of the special electrostatic geometry of the 2D electron system in the experiments. According to Spivak theory the global electrostatic neutrality in the system does not come from the presence of background positive charge (in the same plane as the 2D electron system). Instead, the electric neutrality of the system is enforced by the a metallic gate with a positive charge density $+en$, which is separated by a distance d from a metallic gate. Following this idea he showed that in 2D electron system can undergo a first order phase separation with large droplet size (See chapter 2 for details).

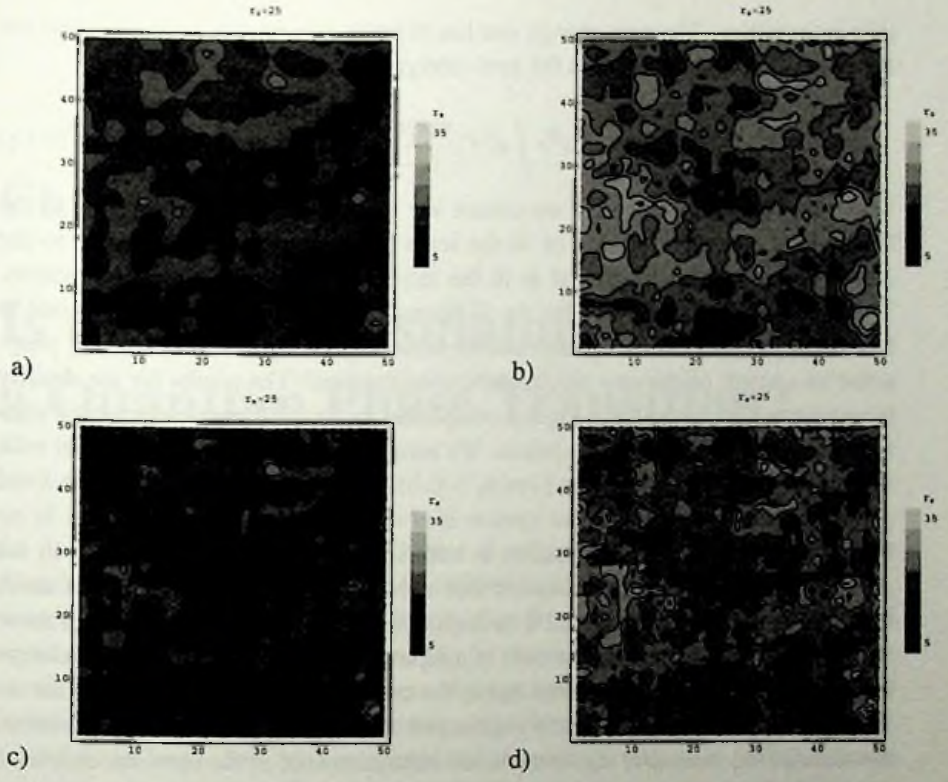


Figure 8.4: Density distribution for the 2D electron system for a typical disorder landscape within Si-Varma theory. a) and b) show this density fluctuations excluding the Spivak electrostatic configuration (see the text) for two different average densities: a)$r_s = 10$ and b)$r_s = 25$. c) and d) show the same density distribution including the Spivak suggestion: c)$r_s = 10$ and d)$r_s = 25$. The strength of disorder potential is $1/\omega_0\tau_0 = 0.01$.

We find it interesting to find out what would happen if we turn on the disorder potential in the same electrostatic configuration as in the Spivak theory. In the presence of small disorder (which is the case for highly mobile electrons in MIT experiments) we believe that the correct physics of 2D electron system (liquid) at low densities can be explained (at least compare to LDA-droplet model) by introducing the Si-Varma correction to LDA. This is due to the large screening length compare to the mean free path of the electron in this region. We want to see how much the inclusion of Spivak mechanism would affect the screening properties of 2D electron system at low densities. In other words, what is the performence of Spivak's electrostatics on the density inhomogeneity.

To test the Spivak suggestion, we move the compensating positive background to a distance d from 2D electron system and simulate the system (based on the DFT energy minimization). Now we ask what is the combined effect of this capacitor configuration of Spivak and the Si-Varma physics, specially related to the density

inhomogeneities. The only change one has to make is to rewrite the expression for the direct Coulomb interaction for new configuration.

$$E_{e-e}[n] = \frac{e^2}{\varepsilon}\int d^2r \int d^2r' [\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \frac{n(\mathbf{r})n}{|\mathbf{r}-(\mathbf{r}'+\mathbf{d})|}], \tag{8.12}$$

where $\mathbf{d} = d\hat{\mathbf{z}}$. Unfortunately we cannot see any substantial change either in the behavior of incompressibility or in the scale of inhomogeneities compare to the case when positive background is in the same plane as the 2D electron system. This is mostly due to the fact that the Si-Varma energy is the most dominant one in low density regime as the system cannot sense any change in the screening properties as a result of the new electrostatic configuration. The results for the density inhomogeneities is shown in Fig.8.4 comparing the two different electrostatic configuration for the 2D electron system. We used a correlated disorder potential with an amplitude corresponding to $1/\omega_0\tau_0 = 0.01$, and the distance between the fixed positive plane and 2D electron system is $d = 10a_B$. It is clear that there is no big difference in density fluctuation in both cases. The average difference in the scale of density fluctuation between two corresponding distribution is not more than 5% in low densities and 12% in high densities. The main difference in these two configurations is that in the case of a separated positive background the charge in homogenies are more localized due to the more screening. The results for the inverse compressibility is basically unchanged and the incompressibility data (from the simulation) is exactly the same as has been discussed in the previous chapter.

8.4 Conclusions

In this chapter we discussed the local properties of the 2D electron system within Si-Varma physics and compared our results with LDA-droplet picture discussed in chapter 5. In the LDA case the presence of dominating exchange energy has the effect that density fluctuations are amplified; after subtracting the uniform Coulomb energy the exchange interaction acts as an attractive force. The Si-Varma term expresses that the screening length becomes large in the fluid, rendering it increasingly incompressible with the obvious effect that density fluctuations are suppressed. At the same time it seems that Si-Varma energy forces the inhomogeneities, while small in amplitude, to redistribute themselves with smaller size. This is more profound at a low enough density where Si-Varma is the dominating force. Different electrostatic configurations, as suggested by Spivak, do not change the scale of the inhomogeneities in the (disordered) system where the interactions are strong rooted in the simple observation that the screening properties of the metal and the insulator in the vicinity of the MIT are barely different, a fact encoded in the Si-Varma correction.

Chapter 9

Is the Metal-insulator Transition a Quantum Phase Transition?

As discussed in the introduction (Chapter 1), when the temperature decreases, a dilute electron gas of highly mobile Si-MOSFET exhibits a strong drop in the resistance if the electron density n is higher than a certain critical one $n > n_c$, and an increase in resistance when $n < n_c$ [7]. In the vicinity of n_c the resistance possesses scaling properties as a function of temperature and electron density. This instance, from the very first caused researchers to consider the observed transition as a disorder-controlled quantum phase metal-insulator transition (MIT). This in turn gave rise to a tide of similar investigations of other systems in which any change in the sign of derivative $\frac{dR}{dT}(n)$ was taken as an evidence of the occurrence of a quantum phase MIT. If there is any quantum phase transition (QPT) occurring in these 2D electron systems, it should also reveal itself in other zero-temperature thermodynamical quantities. Zero-temperature specific heat (c_p), spin susceptibility (χ), and compressibility, κ are among candidate quantities for possible probe of quantum critical point (QCP). Although the measurement of c_p close to $T = 0$ is very difficult, but the relationship between this quantity and the effective mass, $m*$, of the quasiparticles makes it possible to probe QCP by investigating the possible scaling properties of $m*$ near QCP[1]. Indeed there are reports on the effective mass enhancement near the critical density in a metallic two-dimensional electron system. On the other hand, compressibility measurements of 2D electron/hole systems [13, 14, 15], shows that the divergence behavior of this quantity occurs not in the transport critical density n_c but rather around densities smaller than n_c. Is it possible that the divergence of inverse compressibility is a signature of a more fundamental zero-temperature metal-insulator quantum phase transition? In other words, is metal-insulator transition in 2D electron(hole) system due to the existence of QCP? In this chapter we will try to address these questations in more detail.

[1]To date there is no report on the low-temperature specific heat measurement in these systems.

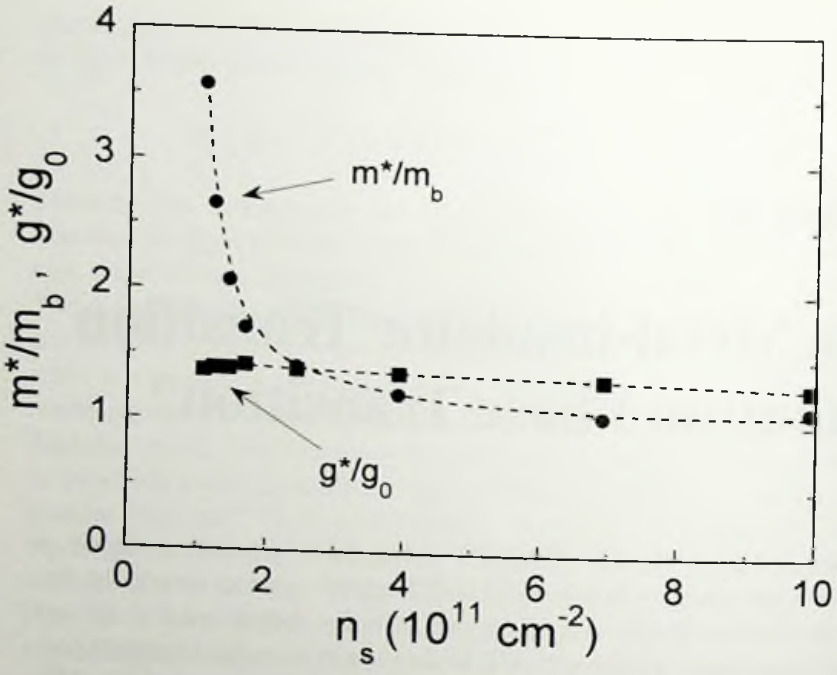


Figure 9.1: The effective mass (circles) and g factor (squares) in a silicon MOSFET, determined from the analysis of the parallel field magnetoresistance and temperature-dependent conductivity, versus electron density. The dashed lines are guides to the eye. (from Shashkin *et al. [39]*)

9.1 Effective mass divergence at low densities

As discussed earlier (chapter 1), in Si-MOSFETs or GaAs/AlGaAs heterostructures, various experimental methods provide evidence for a sharp increase and possible divergence of the spin susceptibility at some finite sample-independent electron density, n_χ, which is at or very near the critical density for MIT in these high mobility samples.

In principle, the increase of the spin susceptibility could be due to an enhancement of either the electron g-factor, g^*, or the effective mass, m^*, (or both), through $\chi \propto m^* g^*$. Shashkin *et al.* [39] obtained g^* and m^* separately by analyzing their data for the temperature dependence of the conductivity in zero magnetic field using the recent theory of Zala *et al.* [46]. According to this theory, σ is a linear function of temperature

$$\frac{\sigma(T)}{\sigma_0} = 1 - A^* k_B T, \tag{9.1}$$

where the slope, A^*, is determined by the interaction-related parameters: the Fermi liquid constants F_0^a and F_1^s

$$A^* = \frac{(1+8F_0^a)g^*m^*}{\pi\hbar^2 n_s}, \qquad \frac{g^*}{g_0} = \frac{1}{1+F_0^a}, \qquad \frac{m^*}{m} = 1+F_1^s$$

where $g_0 = 2$ is the "bare" g-factor.) These relations allow a determination of the many-body enhanced g^* factor and effective mass m^* separately using the data for the slope A^* and the product g^*m^*. Values of g^*/g_0 and m^*/m_b determined from this analysis are shown as a function of the electron density in Fig. 9.1. In the high n region (relatively weak interactions), the enhancement of both g and m is relatively small, both values increasing slightly with decreasing electron density [155]. Also, the renormalization of the g-factor is dominant compared to that of the effective mass, consistent with theoretical studies [154, 152, 153]. In contrast, the renormalization at low n (near the critical region), where $r_s \gg 1$, is much more striking. As the electron density is decreased, the renormalization of the effective mass increases markedly with decreasing density while the g factor remains relatively constant. Hence, this analysis indicates that it is the effective mass, rather than the g-factor, that is responsible for the strongly enhanced spin susceptibility near the metal-insulator transition. To verify this conclusion, Shashkin *et al.* [40, 41] used an independent method to determine the effective mass through an analysis of the temperature dependence of the Shubnikov-de Haas (SdH) oscillations extended to much lower temperatures where the Dingle temperature is expected to be constant and the analysis reliable.

To probe a possible connection between the effective mass enhancement and spin and exchange effects, Shashkin *et al.* [40, 41] introduced a parallel magnetic field component to align the electrons' spins. They analyzed the effective mass as a function of the spin polarization, $P = (B_\perp^2 + B_\parallel^2)^{1/2}/B_c$ and showed that within the experimental accuracy, the effective mass does not depend on P. Therefore, the enhancement of m^* near the MIT is robust, and the origin of the mass enhancement has no relation to the electrons' spins and exchange effects.

9.2 Compressibility and effective mass

Now, as we know that the enhancement of the effective mass (through the increase in spin susceptibility) near MIT is quite evident, the next question is: whether this phenomenon somehow related to the divergence of the inverse compressibility in the same critical region. To address the this question, one has to recall that these effects are related to the metallic phase of the critical region. Therefore it is very natural to (in the first step) Fermi-liquid character of this phase! According to the Fermi-liquid theory one has the following relationships for the compressibility, spin susceptibility, and specific heat:

$$\frac{\kappa_0}{\kappa} = \frac{1+F_0^s}{1+F_1^s/3}, \tag{9.2}$$

$$\frac{\chi_0}{\chi} = \frac{1+F_0^a}{1+F_1^s/3}, \tag{9.3}$$

$$\frac{C_0}{C} = \frac{m}{m*} = \frac{1}{1+F_1^s/3}, \tag{9.4}$$

where F^s (F^a) are symmetric (asymmetric) Landau parameters in Fermi liquid theory. As mentioned earlier the divergence of spin susceptibility seen in the experiment near the MIT is believed to be due to the enhancement of effective mass near this critical point. This is indeed very clear from Eq. 9.4 and Eq. 9.3. On the other hands combining Eq. 9.4 and Eq. 9.2 we have:

$$\frac{\kappa_0}{\kappa} = \left(\frac{m}{m^*}\right)(1+F_0^s). \tag{9.5}$$

From this expression it is clear that although the inverse compressibility singularity and the effective mass divergence are related, they are not necessarily the same. The divergence in $1+F_0^s$ and F_0^s is clearly related to the screening properties of the fermion system. From Eq. 9.5 it is very clear that the divergence in effective mass does not give rise to the enhancement of the inverse compressibility, κ^{-1}, in a explicit way. In the following we will look at this in more detail. First we study the effect of the enhancement of the effective mass on the behavior of the compressibility at low densities near a critical point where mass diverges and try to compare it to the effect of the screening on the same quantity (including the fact that screening length also tends to divereg near the same critical point). After that we try to address the combined effect of the effectice mass and screening length divergence on the inverse compressibility.

Let us assume that we are in the regime of a very small amount of disorder and that the electron system in fairly homogeneous. In this case the LDA approximation applies and one can write the energy density (in Rydberg) of the charged particles as

$$\varepsilon_{LDA}(n) = \pi a_B^{*2} n - \frac{2}{3\pi} a_B^* \sqrt{\frac{2}{\pi}} n^{-1/2} \tag{9.6}$$

in which we also assumed that we are in the regime of small densities (large r_s) where the correlation energy is neglectable. $a_B^*$is the effective Bohr radius and is related to the free Bohr radius as $a_B^* = (m/m^*)a_B$. Now let us assume that we are near a critical density, n_c, where the effective mass shows strong enhancement as a function of the density of electrons/holes. Let us assume the following scaling relation for the effective mass near n_c

$$\frac{m^*}{m} \sim 1 + (n^*)^{-\alpha}, \tag{9.7}$$

where $n^* = \frac{n-n_c}{n_c}$ is the reduced density and α is a positive number. In this case near the critical point n_c the effective Bohr radius can be expressed as

$$a_B^* = \frac{a_B}{1+(n^*)^{-\alpha}}. \tag{9.8}$$

Substituting this relationship into the Eq. 9.6, we try to calculate the inverse compressibility, $\kappa^{-1} \propto \partial\mu/\partial n$, as

$$\begin{aligned}\frac{\partial\mu}{\partial n} &= \frac{\partial}{\partial n}(\varepsilon_{LDA}(n, n_c, \alpha)n) \\ &= \left(\frac{\partial\mu}{\partial n}\right)^{(0)} + \left(\frac{\partial\mu}{\partial n}\right)^{(1)} + \left(\frac{\partial\mu}{\partial n}\right)^{(2)} + \left(\frac{\partial\mu}{\partial n}\right)^{(3)}\end{aligned}, \tag{9.9}$$

where

$$\left(\frac{\partial\mu}{\partial n}\right)^{(0)} = 2\pi a_B^2 \left(\frac{m^*}{m}\right)^{-2} - \frac{a_B}{2\pi}\sqrt{\frac{2}{\pi}}\left(\frac{m^*}{m}\right)^{-1} n^{-1/2}, \tag{9.10}$$

$$\left(\frac{\partial\mu}{\partial n}\right)^{(1)} = \frac{\partial}{\partial n}\left(\frac{m^*}{m}\right)\left[-8\pi a_B^2\left(\frac{m^*}{m}\right)^{-3} n + \frac{2a_B}{\pi}\sqrt{\frac{2}{\pi}}\left(\frac{m^*}{m}\right)^{-2} n^{1/2}\right] \tag{9.11}$$

$$\left(\frac{\partial\mu}{\partial n}\right)^{(2)} = \left(\frac{\partial}{\partial n}\left(\frac{m^*}{m}\right)\right)^2\left[6\pi a_B^2\left(\frac{m^*}{m}\right)^{-4} n^2 - \frac{4a_B}{3\pi}\sqrt{\frac{2}{\pi}}\left(\frac{m^*}{m}\right)^{-3} n^{3/2}\right] \tag{9.12}$$

$$\left(\frac{\partial\mu}{\partial n}\right)^{(3)} = \frac{\partial^2}{\partial n^2}\left(\frac{m^*}{m}\right)\left[-2\pi a_B^2\left(\frac{m^*}{m}\right)^{-3} n^2 + \frac{2a_B}{3\pi}\sqrt{\frac{2}{\pi}}\left(\frac{m^*}{m}\right)^{-2} n^{3/2}\right] \tag{9.13}$$

In the limit of $n \to n_c$, effective mass ratio and its first and second derivatives reads as

$$\left(\frac{m^*}{m}\right) \sim n^{*-\alpha}, \quad \frac{\partial}{\partial n}\left(\frac{m^*}{m}\right) \sim \frac{-\alpha}{n_c} n^{*-(\alpha+1)}, \quad \frac{\partial^2}{\partial n^2}\left(\frac{m^*}{m}\right) \sim \frac{\alpha(\alpha+1)}{n_c^2}\, n^{*-(\alpha+2)}. \tag{9.14}$$

In this case the inverse compressibility (incompressibility) can be written as

$$\begin{aligned}\left(\frac{\partial\mu}{\partial n}\right)_{n\to n_c} &= 2\pi a_B^2\left[4\alpha n^{*2\alpha-1} + \alpha(2\alpha-1)n^{*2\alpha-2}\right] \\ &\quad - \frac{2a_B}{\pi}\sqrt{\frac{2}{\pi}} n_c^{-1/2}\left[\alpha n^{*\alpha-1} + \frac{\alpha(\alpha-1)}{3} n^{*\alpha-2}\right]\end{aligned} \tag{9.15}$$

In the effective mass measurement experiment by Shashkin *et al* [39] the exponent of effective mass scaling is believed to be between -0.4 to -0.6. In figure 9.2 we calculate $\partial\mu/\partial n$ numerically using the above method, trying to fit it with experimental data of Dultz *et al* [13] for the best choice of n_c.

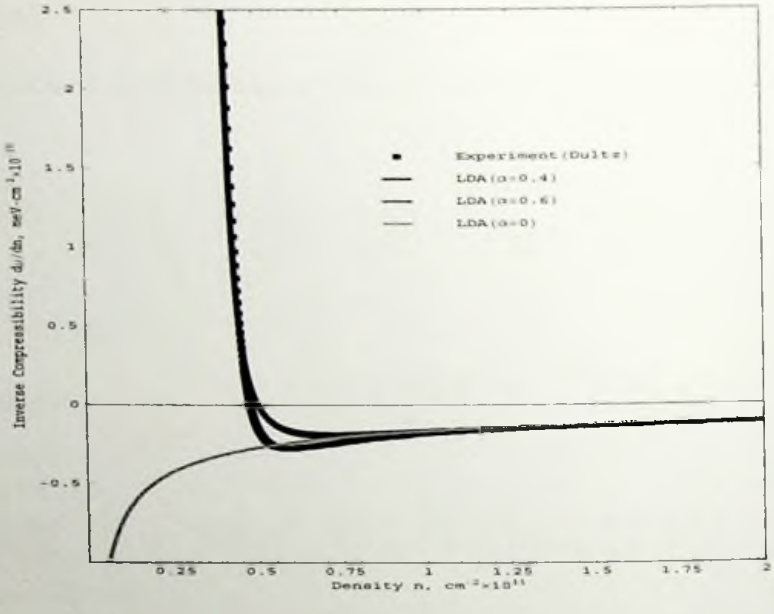


Figure 9.2: Incompressibility of LDA theory based on effective mass divergence.

More interesting is when we take $\alpha = 0.5$ in the above expression which gives the asymptotic behavior of κ^{-1} as

$$\left(\frac{\partial\mu}{\partial n}\right)_{n\to n_c} = 2\pi a_B^2 - \frac{a_B}{\pi}\sqrt{\frac{2}{\pi}}(n_c n^*)^{-1/2} + \frac{a_B n_c}{6\pi}\sqrt{\frac{2}{\pi}}(n_c n^*)^{-3/2}. \tag{9.16}$$

This exactly is the same expression of incompressibility of uniform electron gas including Si-Varma correction, not as a function of n alone but of reduced density, n^*. Using this comparison and by looking at the last term we see that the corresponding disorder strength, $1/\omega_0\tau_0$, is

$$(\omega_0\tau_0)^{-1} = \frac{2}{3}a_B^2 n_c. \tag{9.17}$$

For the above best fit $n_c = 3.7 \times 10^{10}\, cm^{-2}$ which gives $(\omega_0\tau_0)^{-1} = 0.001$. Although $(\omega_0\tau_0)^{-1}$ is very small, it is clear from the Fig 9.2 it is impossible to get the minimum right on the experimental curve.

By looking at the Eq.9.16 it is very interesting to ask what would we get if we introduce the criticality near n_c within Si-Varma formulation. This means substituting n by $n_{eff} = n_c n^*$ in the formulation of the chapter 5. In other words, we assume that Fermi-liquid collapses at the critical density n_c, which is a physically realistic assumption as experimentally at densities lower than n_c the electrons are

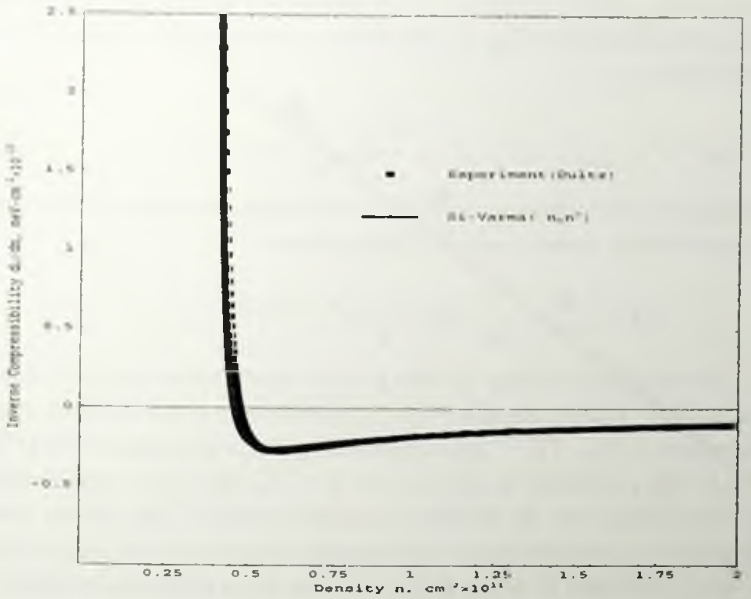


Figure 9.3: Incompressibility of Si-Varma theory based on $n_{eff} = n_c n^*$.

pined into the disorder sites and form a Wigner glass. Using this assumption we have the following equation for the electronic inverse compressibility

$$\left(\frac{\partial \mu}{\partial n}\right) = 2\pi a_B^2 - \frac{a_B}{2\pi}\sqrt{\frac{2}{\pi}}(n_c n^*)^{-1/2} + \frac{(\omega_0 \tau_0)^{-1}}{\pi a_B}\sqrt{\frac{2}{\pi}}(n_c n^*)^{-3/2}. \qquad (9.18)$$

Using this equation and calculate κ^{-1} numerically, we can fit the experimental data with a very good accuracy. The results has been shown in the Fig. 9.3.

So far we calculated the effect of the effective mass enhancement as well as the screening length divergence on the separate ground. Let us assume that the physics of large screening length is the correct description of the 2D electron liquid near MIT and at the same time the real transition occur at the finite density n_c. Due to the fact that effective mass is entering into the Si-Varma energy through a_B^*, one has to include the mass enhancement into the Si-Varma Physics as well. Therefore we can write the Si-Varma energy in its critical form as

$$\begin{aligned} \varepsilon_{SV}[n^*, m^*] &= -\frac{1}{\pi a_B^*}\sqrt{\frac{2}{\pi}}\frac{1}{\omega_0 \tau_0}(n_c n^*)^{-1/2} \\ &= -C(\frac{m^*}{m})(n^*)^{-1/2}, \end{aligned} \qquad (9.19)$$

where $C = \frac{1}{\pi a_B}\sqrt{\frac{2}{\pi}}\frac{1}{\omega_0\tau_0}$ and $n^* = (n/n_c - 1)$. Using the asymptotic expression for effective mass near critical point, Eq. 9.7, the inverse compressibility contribution coming from Si-Varma is

$$(\kappa^{-1}[n^*, m^*])_{SV} = (\kappa^{-1}[n^*])_{SV} - Cn_c\frac{\partial^2}{\partial n^2}((n^*)^{-\alpha+1/2}). \tag{9.20}$$

Therefore one can read the following correction to the inverse compressibility due to the mass enhancement, coming from Si-Varma physics

$$(\kappa^{-1})_{corr} = -\frac{C}{n_c}[(\alpha^2 - \frac{1}{4})(n^*)^{-(\alpha+3/2)}.$$

This is very interesting as in shows that the positive contribution from the mass enhancement to the Si-Varma compressibility is possible only if the effective mass diverges slower than $(n/n_c - 1)^{-1/2}$ upon approaching the critical point. It is clear that for $\alpha = 1/2$ this correction is zero and for $\alpha > 1/2$ the mass enhancement has negative contribution into the Si-Varma incompressibility. This means that, according to the above considerations, the exponent related to the divergence of incompressibility is less than 2! In the nest section we try to extract this exponent from the experimental data and try to address the question whether the MIT has anything to do with quantum criticality.

9.3 Compressibility and Quantum Criticality

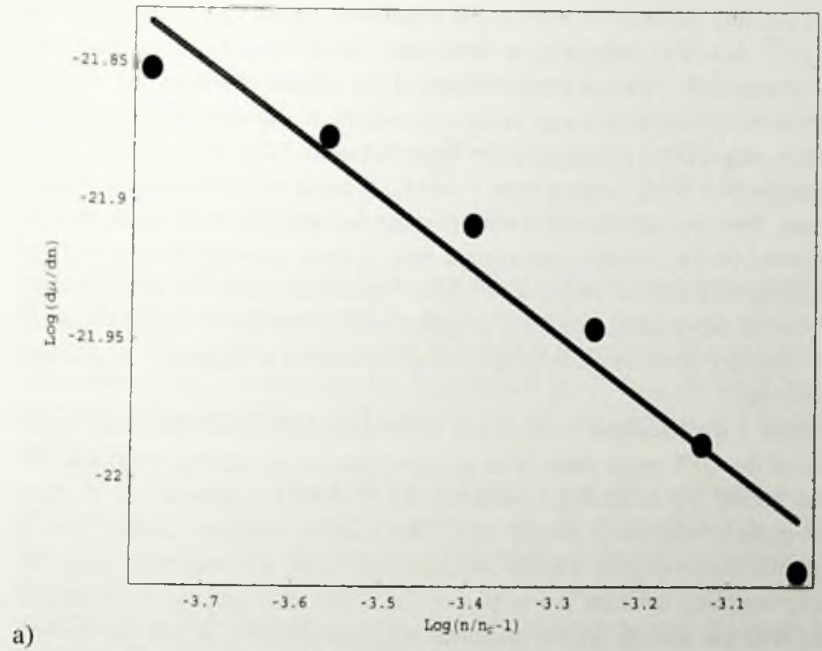
We start this section by looking more closely into the compressibility data coming from the experiment by Dutlz *et al.* on the 2D hole system [13]. Let us assume that there is a quantum critical point, n_c, at which some thermodynamical quantities, like specific heat or inverse compressibility, diverge. In the compressibility data this critical point, if there at all, is the point where the quantity, $\kappa^{-1} \propto \partial\mu/\partial n$, diverges. The critical behavior of inverse compressibility must reveal itself as a power-law type function near n_c. Let us assume that the inverse compressibility has the following power-law behavior near n_c

$$\frac{\partial\mu}{\partial n} \propto \frac{1}{\kappa} \sim (n^*)^{-\alpha_\kappa}, \tag{9.21}$$

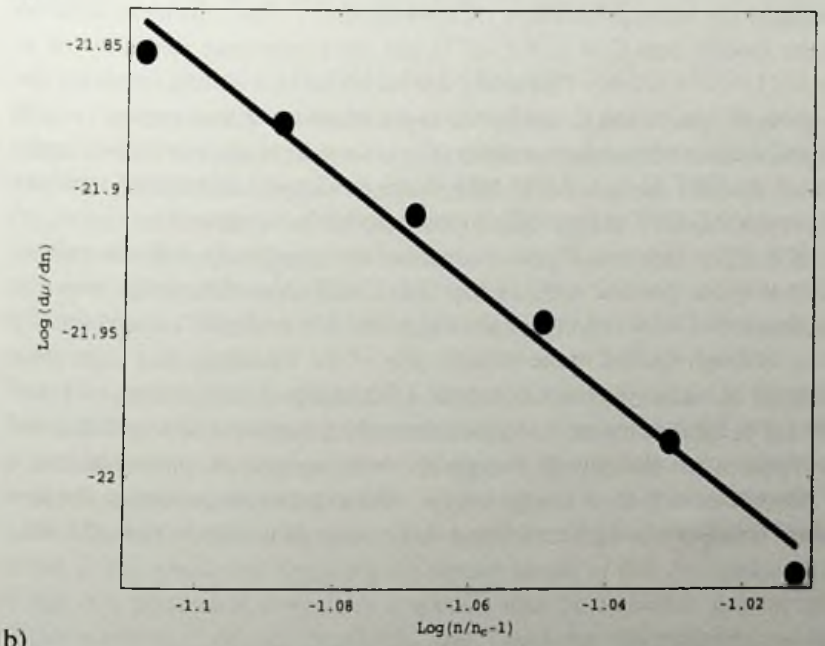
where $n^* = (n - n_c)/n_c$ and α_κ is the critical exponent associated with inverse compressibility. Therefore the log-log plot of the data on compressibility may help to get the critical exponent from the above relation

$$\alpha_\kappa = \frac{\log(\kappa^{-1})}{\log(n^*)}. \tag{9.22}$$

Figure 9.4 shows this plot from the compressibility data very close to the critical point n_c. Actually we do not know exactly what is n_c in compressibility measurement experiment, but it must be smaller and close to the minimum density at



a)



b)

Figure 9.4: Log-Log plot of the incompressibility data ($\partial\mu/\partial n$) as a function of reduced density $n^* = n/n_c - 1$, a) for the critical density $n_c = 3.9 \times 10^{10}$, b) for the critical density $n_c = 3.0 \times 10^{10}$. The solid curve is the fit to the $(n/n_c - 1)^{-\alpha_\kappa}$: a) $\alpha_\kappa = 0.24$, b) $\alpha_\kappa = 1.9$.

which this quantity has been measured. The minimum density of the measurement is 3.99×10^{10}. It is also interesting to mentioned that without knowing the value of n_c it is impossible to have a good estimate of the critical exponent α_κ. In fact due to the limited dynamical range in the experiment at low densities estimating α_κ is quite n_c-dependent. Using n_c in the range between 3.0×10^{10} to 3.9×10^{10}, gives the estimation of α_κ ranging from 1.9 to 0.24, using the best fit to the experimental data. Now the question is: considering the fact that the exact value of n_c is not clear based on the current experimental data, is there anyway that we can get more insight into the critical properties of the 2D electron system in a more general ground? On the other hand, is MIT a quantum phase transition? In the follwing we are going to use some of the scaling relation we derived in chapter 3, to address these questions.

In chapter 3 we discussed that one can construct a fool-proof thermodynamic diagnostic of the QPT in the form of an overcomplete set of scaling relations, derived from almost trivial scaling considerations (Si-Rosch scaling). Let us now apply this general wisdom to the 2D MIT. The coupling constant clearly corresponds with the carrier density n while the transition occurs at a critical n_c such that the reduced coupling constant is $r = (n - n_c)/n_c$. The thermodynamic quantities associated with the density are the chemical potential $\mu = \partial F/\partial n$ and the inverse compressibility (or incompressibility) $(1/\kappa) = (1/n)(\partial^2 F/\partial n^2)$, while in addition one has the specific heat $C = -(\partial^2 F/\partial T^2)$ and the temperature derivative of μ, $\eta_n = (\partial \mu/\partial T) = (\partial^2 F/\partial T \partial n)$. One easily derives a total of 9 scaling forms for the singular parts of $1/\kappa$, η and C, applicable to the quantum critical regime ($r = 0$, finite T) and the low temperature regimes ($T \to 0$, $r \neq 0$) of the two phases in the proximity of the QPT. In fact, the 2d MIT shares the density as coupling constant with the (suspected) QPT in optimally doped cuprates (see chapter 3).

The QPT of the cuprates, if present, represents a completely different universality class than the possible QPT of the 2DEG MIT. Also differently from the cuprates, the crucial information contained in η and $1/\kappa$ is already experimentally accessible, although limited to the metallic side of the transition. Let us assume that the metal at higher densities is indeed a Fermi liquid such that $y_0 = 1$ and (of course) $d = 2$ It is convenient to parametrize the expressions in terms of z and the zero-temperature analogue of the specific heat exponent α_r characterizing a thermal phase transition $\alpha_r = 2 - (d+z)/y_r$. The expressions governing the low temperature behaviors in the Fermi-liquid in the close proximity to the QPT simplify to

$$C_{cr}(T \to 0, r) = \frac{2\rho_0 c}{T_0^2} r^{(2-\alpha_r)(2-z)/(2+z)} T \tag{9.23}$$

$$\eta_{r,cr}(T \to 0, r) = -\frac{2\rho_0 c}{T_0^2} \frac{(2-z)}{(2+z)} \times (2-\alpha_r) r^{(2-\alpha_r)(2-z)/(2+z)-1} T \tag{9.24}$$

$$\frac{1}{\kappa_{r,cr}})(T \to 0, r) = -\rho_0 \tilde{f}(0)(1+z)(2-\alpha_r) r^{-\alpha_r}$$
$$-\frac{c\rho_0}{T_0^2}(2-z)((2-\alpha_r)\left(\frac{2-z}{2+z}\right)-1)$$
$$\times r^{(2-\alpha_r)(2-z)/(2+z)-2} T^2 \qquad (9.25)$$

One infers that the divergence of the prefactors of the temperature dependent parts of these quantities are governed by the fundamental 'mass' exponent $\alpha_m = (2-\alpha_r)(2-z)/(2+z)$ while independent information is obtained from the zero-temperature incompressibility characterized by the exponent α_r. In fact, the latter quantity has precisely the same fundamental status as the specific heat has at a thermal phase transition.

This is our first scaling relation, associating the mass-exponent α_m with the zero temperature incompressibility exponent α_r, on the metallic side of the transition. This is not yet decisive because of the unknown dynamical exponent z. However, Si-Rosch noticed the subtlety that in fact the *amplitude* of the following ratio (equivalent to the Grüneisen ratio when r corresponds with pressure) becomes universal: $\Gamma_r = \eta_r / C = (d - y_0 z)/(y_0 y_r) r^{-1}$, in our case

$$\Gamma_r = (2-z)(2-\alpha_r)/(2+z) r^{-1} \qquad (9.26)$$

Hence, by measuring the various quantities over a sufficiently large dynamical range in the Fermi-liquid regime a critical test of quantum critical scaling is feasible. An even more stringent test is possible when one can measure the truly independent scaling forms of the quantum critical regime. To reach this regime one has to get sufficiently close to the transition so that temperature exceeds the Fermi-energy avoiding the ultraviolet cut-off. In the regime where $T > E_F$ one finds that $C \sim T^{2/z}$, $\eta \sim T^{(2(1-\alpha_r)-z)/(z(2-\alpha_r))}$ and $(1/\kappa) \sim T^{-((2+z)\alpha_r/(z(2-\alpha_r))}$, amounting to three extra scaling laws!

How does the above relate to the thermodynamic experiments, feasible in the MOSFET's? First, several groups have reported measurements of the chemical potential and the incompressibility. It appears that indeed the incompressibility is diverging at some critical density which appears to be lower than the density where the MIT occurs in transport. The dynamical range of the data is too small to arrive at any conclusion regarding the precise nature of this divergence. However, if this is a power law divergence it appears that the exponent is quite large; we obtain a best fit $(1/\kappa) \sim r^{-\alpha_\kappa}$ with $\alpha_\kappa \simeq 2$. Given that this exponent has the same status as a specific heat exponent at a classical phase transition this might appear as troublesome given that specific heat exponents are usually small compared to unity.

The electronic specific heat cannot be directly measured in the MOSFET's. However, our interest in the coefficient of the linear specific heat γ which is directly proportional to the effective mass m^*/m, which is accessible by transport

measurements. Podalov *et al.* used [78] Schubnikov-de Haas oscillations, to find that $m*/m$ is diverging as r^{α_m} with $\alpha_m = -0.5 \pm 0.1$, at a critical density similar as the one governing the incompressibility divergence. Given the quality of the data this does not suffice to arrive at any definite conclusion. However, we can deduce a rough consistency check. We noticed that the incompressibility singularity has to be much stronger than a typical classical specific heat singularity. The scaling law for α_m can be written as $\alpha_r = 2 - \alpha_m(2+z)/(2-z)$. Since the mass diverges at the transition $\alpha_m < 0$ and since a negative z does not make sense it has to be that $\alpha_r > 2$! In other words, a QPT characterized by a diverging mass in the neighboring Fermi-liquid is characterized by an 'energy singularity' which is gigantic as compared to what is found at normal (classical-, bosonic) phase transitions, giving away an interesting clue regarding the rather mysterious nature of QPT's involving fermions.

Summarizing this part, the experimentalists should attempt to measure the temperature and density dependences of both the effective mass and the chemical potential in the same samples, trying to acquire a large dynamical range in the close proximity of the phase transition. In the above we have presented a strategy to deduce from such data if the MIT is a genuine quantum phase transition or not.

The short coming is of course is that it is a purely phenomenological scaling analysis and a-priori not much is learned regarding the physics at work at the phase transition. Although we do not claim it to be a genuine critical theory

9.4 Conclusions

Leaning on recent insights obtained in heavy fermion physics regarding the thermodynamic singularity structure associated with quantum phase transitions (Rosch-Si scaling analysis) we presented here an experimental strategy to establish if the zero-temperature transition in the disordered two dimensional gas is a real quantum phase transition. We derived scaling laws relating the density and temperature dependence of the chemical potential and the effective mass, and we addressed these theoretically starting from the idea of a diverging screening length by Si and Varma.

Appendix A

Long-range Potentials and Ewald Image Technique

There is a problem in using long-range potentials in periodic boundries; how does one deal with a potential that leaks across the boundaries? Suppose the bare potential in infinite space is $v(r)$. Let us define the Fourier transform by:

$$\tilde{v}_{\mathbf{k}} = \int_{-\infty}^{\infty} d^D\mathbf{r} e^{-i\mathbf{k}\mathbf{r}} v(r). \tag{A.1}$$

Here D is the dimensionality of space. Then its inverse is:

$$v(r) = \frac{1}{(2\pi)^D} \int_{-\infty}^{\infty} d^D\mathbf{k} e^{i\mathbf{k}\mathbf{r}} \tilde{v}_{\mathbf{k}}. \tag{A.2}$$

Now let us find the energy of a single particle interacting with an infinite rectangular lattice of another particle a distance $\mathbf{r}$ away. To make the sum converge for a long-range potential, we also add a uniform background of the same density ($1/volume = 1/V$) of opposite charge. Thus, the *image potential* is equal to

$$v_I(\mathbf{r}) = \sum_{\mathbf{L}} v(\mathbf{r}+\mathbf{L}) - \tilde{v}_0/V. \tag{A.3}$$

The $\mathbf{L}$ sum is over the Bravias lattice of the simulation cell $\mathbf{L} = (m_x L_x, ...)$ where m_x, m_y, ... ranage over all positive and negative integers. The second term is the background. Converting this to k-space using the Poisson sum formula:

$$v_I(\mathbf{r}) = \frac{1}{V} {\sum_{\mathbf{k}}}' \tilde{v}_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}}, \tag{A.4}$$

where the prime indicates that we omit the $\mathbf{k} = 0$ term; it cancels out with the background. The k-sum is over reciprocal lattice vectors of the simulation box: $\mathbf{k} = (2\pi n_x/L_x, ...)$.

Because both sums are so poorly convergent, we use both k-space and r-space; taking the long-range part into k-space and the short-range part in r-space. We write:

$$v(r) = v_s(r) + v_l(r). \tag{A.5}$$

(For the moment this division is arbitrary.) Since Fourier transforms are linear:

$$\tilde{v}_{\mathbf{k}} = \tilde{v}_{s\mathbf{k}} + \tilde{v}_{l\mathbf{k}}. \tag{A.6}$$

The image potential can be written as

$$v_I(\mathbf{r}) = \sum_{\mathbf{L}} v(|\mathbf{r}+\mathbf{L}|) + \frac{1}{V}\sum_{\mathbf{k}} \tilde{v}_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}}, \tag{A.7}$$

where the background requires $\tilde{v}_{l0} = -\tilde{v}_{s0}$.

Now let us work with N particles of charge q_i in a periodic box and let us compute the total potential energy of the unit cell. Particles i and j are assumed to interact with potential of $q_i q_j v(r_{ij})$. The image potential for the N-particle system is:

$$V_I = \sum_{i<j} q_i q_j v_I(\mathbf{r}_{ij}) + \sum_{i=1}^{N} q_i^2 v_M, \tag{A.8}$$

where v_M is the interaction of a particle with its own images; it is Madelung constant for particle i interacting with the perfect lattice of the simulation cell. If this term were not present, particle i would only see $N-1$ particles in the surrounding cells instead of N. We can find its value by considering the limit as two particle get close together with the image potential.

$$v_M = \frac{1}{2}\lim_{\mathbf{r}\to 0}[v_I(\mathbf{r}) - v(\mathbf{r})]. \tag{A.9}$$

Now we substitute the split up image potential and collect all the terms together:

$$V_I = \sum_{i<j} q_i q_j v_s(r_{ij}) + \frac{1}{2V}\sum_{\mathbf{k}}{}' v_{l\mathbf{k}} |\rho_{\mathbf{k}}|^2 + const. + dipole, \tag{A.10}$$

where:

$$const. = \frac{\sum_i q_i^2 v_c}{2} - \frac{\tilde{v}_{s0}[\sum_i q_i]^2}{2V}, \tag{A.11}$$

$$v_c = \lim_{r\to 0}(v_s(r) - v(r)), \tag{A.12}$$

and

$$\rho_{\mathbf{k}} = \sum_i q_i e^{i\mathbf{k}\mathbf{r}}. \tag{A.13}$$

The second term of the constant only appears if the system has an imbalances in the net charge of all the particles. (The background really balances out the charge, example are one-component plasma or jellium.)

The dipole term comes from how the infinite sums are finally dealt with for a macroscopic sample. It is only appropriate to use this term if the charge is localized in the sense that you define a dipole moment (for example if the system consists of positive and negative charges bound together.)

$$dipole = \frac{2\pi}{(2\varepsilon+1)V}|\sum_i q_i \mathbf{r}_i|^2. \tag{A.14}$$

where ε is the dielectric constant of the surrounding media; it would be infinite if the similation system is in contact with a metal so that surface charges connot accumulate and unity for a vacuum.

Now we give the standard forms for the beakup wich is done with a Gaussian charge distribution. It gives nice analytic results but not neccessarily optimal. (See the paper by Natoli and Ceperley, J. Comp. Phys. **117**, 171 (1994).)

For an interaction that goes as $v(r) = r^{-n}$ the needed functions are

$$v_s(r) = \frac{\Gamma(\nu, (\xi r)^2)}{\Gamma(\nu) r^n}, \tag{A.15}$$

$$v_{lk} = \frac{\pi^{D/2}(2/k)^{2\eta}\Gamma(\eta, (k/(2\xi))^2)}{\Gamma(\nu)}, \tag{A.16}$$

$$\tilde{v}_{0s} = \frac{\eta\pi^{D/2}}{\Gamma(\nu)\xi^{2\eta}}, \tag{A.17}$$

$$v_c = -\frac{\xi^n}{\nu\Gamma(\nu)}, \tag{A.18}$$

where $\Gamma(x,a)$ is the incomplete gamma function and $\nu = n/2$ and $\eta = (D-n)/2$. A free parameter ξ appears that is related to the width of the distribution. For large arguments the short-ranged potential drops off exponentially fast:

$$v_s(r) \propto e^{-(\xi r)^2}/r^2. \tag{A.19}$$

Specializing for the usual case of Coulomb interaction ($n = 1$) in three dimensions $D = 3$, we get:

$$v_s(r) = erfc(\xi r)/r, \tag{A.20}$$

$$v_{lk} = \frac{4\pi e^{-(k/(2\xi))^2}}{k^2}, \tag{A.21}$$

$$\tilde{v}_{0s} = \frac{\pi}{\xi^2}, \tag{A.22}$$

$$v_c = -\frac{2\xi}{\sqrt{\pi}}. \tag{A.23}$$

One usually chooses ξ so that the short-ranged potential is zero at the edge of the box (at $L/2$) and we can use the minimum image convension and then increase

the number of **k** points until convergence is achieved. The optimal value of ξ to get high accuracy in the Ewald sum is:

$$\xi^* = \sqrt{\pi}/L. \tag{A.24}$$

Bibliography

[1] P.A. Lee, T.V. Ramakrishnan, Rev. Mod. Phys. **57**, 287(1985).

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

[3] B.L. Altshuler, A.G. Aronov, and P.A. Lee, Phys. Rev. Lett. **44**, 1288(1980).

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

[5] G.J. Doan, and D.D. Osheroff, Phys. Rev. Lett. **43**, 721(1979); D.J. Bishop, D.C. Tsui, and R.C. Dynes, Phys. Rev. Lett. **44**, 1153(1980); M.J. Uren, R.A. Davis, and M. Pepper, J. Phys. C **13**, L985(1980).

[6] A. Gold, Phys. Rev. B **44**, 8818(1991).

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

[8] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251(2001).

[9] S. V. Kravchenko, and M. P. Sarachik, Rep. Prog. Phys. 67, 1(2004).

[10] M. P. Sarachik, and S. V. Kravchenko, Proc. Natl. Adac. Sci. USA **96**, 5900(1999).

[11] Y. Hanein, U. Meirav, D. Shahar, C. C. Li, D. C. Tsui, and H. Shtrikman, Phys. Rev. Lett. **80**, 1288(1998).

[12] D. Simonian, S. V. Kravchenko, M. P. Sarachik, and V. M. Pudalov, Phys. Rev. Lett. **79**, 2304(1997).

[13] S. C. Dultz, H. W. Jiang, Phys. Rev. Lett. **84**, 4689(2000).

[14] S. Ilani, A. Yacoby, D. Mahalu, and H. Shtrikman, Phys. Rev. Lett. **84**, 3133(2000).

[15] S. Ilani, A. Yacoby, D. Mahalu, and H. Shtrikman, Science **292**, 1354(2001).

[16] R. Fletcher , V. M. Pudalov, A. D. B. Radcliffe, C. Possanzini, Semicond. Sci. Tech. **16**, 386 (2001).

[17] C. Castellani, C. Di Castro, M. Grilli, and G. Strinati, Phys. Rev. B **37**, 6663(1988).

[18] S. V. Kravchenko, A. A. Shashkin, D. A. Bloore, and T. M. Klapwijk, Solid State Commun. **116**, 495(2000).

[19] A. A. Shashkin, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, Phys. Rev. Lett. **87**, 086801(2001).

[20] S. A. Vitkalov, H. Zheng, K. M. Mertes, M. P. Sarachik, and T. M. Klapwijk, Phys. Rev. Lett. **87**, 086401(2001).

[21] J. Zhu, H. L. Stormer, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, Phys. Rev. Lett. **90**, 056805(2003).

[22] A. M. Finkelstein, Zh. Espk. Teor. Fiz. **84**. 168(1983).

[23] A. M. Finkelstein, Z. Physik **56**, 189(1984).

[24] C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527(1984).

[25] C. Castellani, C. Di Castro, and P. A. Lee, Phys. Rev. B **57**, R9381(1998).

[26] T. R. Kirkpatrik, and D. Belitz, Phys. Rev. B **53**, 41364(1996).

[27] T. R. Kirkpatrik, and D. Belitz, Phys. Rev. B **62**, 966(2000).

[28] D. M. Ceperley, and B. J. Alder, Phys. Rev. Lett. **45**, 566(1980).

[29] S. Chakravarty, L. Yin, and E. Abrahams, Phys. Rev. B **58**, 559(1998).

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

[31] S. He, and X. C. Xie, Phys. Rev. Lett. **80**, 3324(1998).

[32] Y. Meir, Phys. Rev. Lett. **83**, 3506(1999).

[33] S. V. Kravchenko, D. Simonian, K. Mertes, M. P. Sarachik, and T. M. Klapwijk, Phys. Rev. B **59**, R12 740(1999).

[34] Q. Si, and C. M. Varma, Phys. Rev. Lett. **81**, 4951(1998).

[35] C. M. Varma, Z. Nussinov, and W. van Saarloos, Phys. Rep. **361**, 267(2002).

[36] S. Chakravarty, S. Kivelson, C. Nayak, and K. Voelker, Phill. Mag. B **79**, 859(1999).

[37] B. Spivak, Phys. Rev. B **64**, 085317(2001).

[38] B. Spivak, Phys. Rev. B **67**, 125205(2002).

[39] A. A. Shashkin, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, Phys. Rev. B **66**, 073303(2002).

[40] A. A. Shashkin, M. Rahimi, S. Anissimova, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, Phys. Rev. Lett. **91**, 046403(2003).

[41] A. A. Shashkin, S. V. Kravchenko, V. T. Dolgopolov, and T. M. Klapwijk, J. Phys. A: Math. Gen. **36**, 9237(2003).

[42] P. Phillips, Y. Wan, I. Martin, S. Knysh, and D. Dalidovich, Nature (London) **395**, 253(1998).

[43] B. L. Altshuler, D. L. Maslov, and V. M. Pudalov, Phys. Stat. Sol. (b) **218**, 193(2000).

[44] B. L. Altshuler, G. W. Martin, D. L. Maslov, V. M. Pudalov, A. Prinz, G. Brunthaler, and B. Bauer, cond-mat/0008005

[45] Y. Yaish, O. Prus, E. Buchstab, S. Shapira, G. Ben Yoseph, U. Sivan, and A. Stern, Phys. Rev. Lett. **84**, 4954(2000).

[46] G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **64**, 214204(2001).

[47] A. Punnoose, and A. M. Finkelstein, Phys. Rev. Lett. **88**, 016802(2002).

[48] S.T. Chui, and B. Tanatar, Phys. Rev. Lett. **74**, 458(1995).

[49] N. F. Mott, Rev. Mod. Phys. **50**, 203(1978).

[50] A. Mc Kinnon, *Localization, Interaction, and Transport Phenomena*, ed. by B. Kramer, G. Bergman, Y. Bruyseraede, Springer Ser. Solid-State Sci., Vol. 61 (Springer, Berlin, Heidelberg 1985).

[51] L. P. Gorkov, A. I. Larkin, and D. E. Khmelnitskii, Zh. Eksp. Teor. Phys. **30**, 248(1979).

[52] S. Hikami, Prog. Theor. Phys. **64**, 1466(1980).

[53] A. A. Abrikosov, L. P. Gorkov, and I. Dzialoshinsky, *Methods of Quantum Field Theory in Statistical Physics*, (Prentice-Hall, New Jersey, 1963).

[54] O. Betbeder-Matibet and P. Nozières, Ann. Phys. (N.Y.) **37**, 17(1966).

[55] B. L. Altshuler and A. G. Aronov, in *Eletron-Electron Interactions in Disordered Systems*, edited by A. L. Efros and M. Pollak, Elsevier, Science Publishers, New York (1985).

[56] C. Castellani, C. Di Castro, P. A. Lee, M. Ma, S. Sorella, and E. Tabet, Phys. Rev. B **30**, 1596(1984).

[57] B. L. Altshuler, D. L. Maslov, and V. M. Pudalov, Physica E **9** 209(2001).

[58] G. Brunthaler, A. Prinz, G. Bauer, and V. M. Pudalov, Phys. Rev. Lett. **87**, 096802(2001).

[59] V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, Physica E **3**, 79(1998).

[60] P.A. Lee, and T.V. Ramakrishnan,Phys. Rev. B **26**, 4009(1982).

[61] F. Stern, Phys. Rev. Lett. **44**, 1469(1980).

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

[63] S. Das Sarma, Phys. Rev. B **33**, 5401(1986).

[64] S. Das Sarma, and E. H. Hwang Phys. Rev. Lett. **83**, 164(1999).

[65] A. Gold, J. Phys.: Condens. Matter **13**, 11641(2001).

[66] A. Gold, and V. T. Dolgopolov, J. Phys.: Condens. Matter **14**, 7091(2002).

[67] I. F. Herbut, Phys. Rev. B **63**, 113102(2001).

[68] G. Zala , B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **65**, R020201(2002).

[69] A. Gold, J. Phys.: Condens. Matter **15**, 217(2003).

[70] C. Chamon, E. R. Mucciolo, and A. H. Castro Neto, Phys. Rev. B **64**, 245115(2001).

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

[72] E. C. Stoner Rep. Prog. Phys. **11**, 43(1946).

[73] L. D. Landau, Sov. Phys. — JETP **3**, 920(1957).

[74] C. Attaccalite, S. Moroni, P. Gori-Giorgi, and G. B. Bachelet Phys. Rev. Lett. **88**, 256601(2002).

[75] B. Castaing, and P. Nozieres, J. de Physique **40**, 257(1979).

[76] S. Bogdanovich, and D. Popović, Phy. Rev. Lett. **88**, 236401(2002).

[77] J. Jaroszyński, D. Popović, and T. M. Klapwijk, Phys. Rev. Lett. **89**, 276401(2002).

[78] V. M. Pudalov, M. E. Gershenson, H. Kojima, N. Butch, E. M. Dizhur, G. Brunthaler, A. Prinz, and G. Bauer, Phys. Rev. Lett. **88**, 196404(2002).

[79] V. Emery, and S. Kivelson, Physica (Amsterdam) **209**, 597(1993).

[80] J. Zaanen, and O. Gunnarson, Phys. Rev. B. **40**, 7391(1989).

[81] A. A. Koulakov, M. M. Fogler, and B. I. Shklovskii, Phys. Rev. Lett. **76**, 499(1996).

[82] M. M. Fogler, A. A. Koulakov, and B. I. Shklovskii, Phys. Rev. B **54**, 1853(1996).

[83] R. B. Laughlin, G. G. Lonzarich, P. Monthoux, and D. Pines; Adv. in Phys. **20**, 361(2001).

[84] S. Sachdev, Science, **288**, 475(2000).

[85] M. Vojta, Rep. Prog. Phys. **66**, 2069(2003).

[86] J. Ornstein, A. Millis, Science, **268**, 468(2000).

[87] S. L. Sondhi, S. M. Girvin, J. P. Carini, and D Shahar, Rev. Mod. Phys. **69**, 315(1997).

[88] D. Belitz, and T. R. Kirkpatrick, in *Dynamics: Models and kinetic methods for non-equilibrium many-body systems* ed. J Karkheck (Dordrecht: Kluwer 2000).

[89] S. Sachdev *Quantum Phase Transitions* (Cambridge: Cambridge University Press 1999).

[90] R. B. Laughlin, Adv. in Phys. **47**, 943(1998)

[91] D. Bitko, T. F. Rosenbaum, and G. Aeppli, Phys. Rev. Lett. **77**, 940(1996).

[92] P. Coleman, Physica B **259-261**, 353(1999).

[93] H. v. Löhneysen, J. Phys. Cond. Matter **8**, 9689(1996).

[94] E. Dagotto E, Rev. Mod. Phys. **66**, 763(1994); M. B. Maple, J. Magn. Magn. Mater. **177**, 18(1998).

[95] S-K. Ma, *Modern Theory of Critical Phenomena* (Reading: Benjamin 1976).

[96] N. Goldenfeld, *Lectures on phase transitions and the renormalization group* (Reading: Addison-Wesley 1992).

[97] N. F. Mott, *Metal–Insulator Transitions* (London: Taylor and Francis 1990).

[98] B. Widom,J. Chem. Phys. **43**, 3892(1965).

[99] K. G. Wilson, Phys. Rev. B **4**, 3174(1971); *ibid*, 3184(1971).

[100] S. Chakravarty, B. I. Halperin, and D. R. Nelson, Phys. Rev. Lett. **60**, 1057(1988); Phys. Rev. B **39**, 2344(1989).

[101] D. Belitz, and T. R. Kirkpatrick, J. Low Temp. Phys. **126** 1107(2003).

[102] L. Zhu , M. Garst, A. Rosch, and Q. Si, Phys. Rev. Lett. **91**, 066404(2003).

[103] E. Grüneisen, Ann. Phys. (Leipzig) **39**, 257(1912).

[104] L.D. Landau and E.M. Lifshitz, *Statistical Physics* (Butterworth-Heinemann, Oxford, 1959).

[105] R. Takke, M. Niksch, W. Assmus, B. Luthi, R. Pott, R. Schefzyk and D.K. Wohlleben, Z. Phys. **B** 44, 33(1981); P. Thalmeier and P. Fulde, Europhys. Lett. **1**, 367(1986).

[106] S. Kambe, J. Flouquet, P. Lejay, P. Haen, A. deVisser, J. Phys: Conden. Mat. **9**, 4917(1997).

[107] J. A. Hertz, Phys. Rev. B **14**, 1165(1976).

[108] A. J. Millis, Phys. Rev. B **48**, 7183(1993).

[109] Q. Si, S. Rabello, K. Ingersent, J. L. Smith, Nature **413**, 804(2001); D. R. Grempel and Q. Si, Phys. Rev. Lett. **91**, 026401(2003).

[110] A.V. Chubukov and S. Sachdev, Phys. Rev. Lett. **71**, 169 (1993).

[111] C. Castellani, C. di Castro and m. Grilli, Phys. Rev. Lett. **75**, 4650 (1995).

[112] R.B. Laughlin, Adv. Phys. **47**, 943 (1998); S. Chakravarty, R.B. Laughlin, D.K. Morr, C. Nayak, Phys. Rev. B **63**, 094503 (2001).

[113] C.M. Varma, Phys. Rev. Lett. **83**, 3538 (1999); C.M. Varma, Z. Nussinov and W. van Saarloos, Phys. Rep. **361**, 267 (2002).

[114] A. Kaminski, S. Rosenkranz, H.M. Fretwell, Z.Z. Li, H. Raffy, M. Randeria, M.R. Norman, J.C. Campuzano, Phys. Rev. Lett. **90**, 207003 (2003).

[115] J.L. Tallon, J.W. Loram, G.V.M. Williams, J.R. Cooper, I.R. Fisher, J.D. Johnson, M.P. Staines, C. Bernhard, Phys. Stat. Sol. B **215**, 531 (1999); S.H. Naqib, R.A. Chakalov, J.R. Cooper, Physica C **387**, 365 (2003); S. Nakamae, K. Behnia, N. Mangkorntong, M. Nohara, H. Takagi, S.J.C Yates, N.E. Hussey, Phys. Rev. B **68**, 100502 (2003).

[116] D. van der Marel, H. J. A. Molegraaf, J. Zaanen, Z. Nussinov, F. Carbone, A. Damascelli, H. Eisaki, M. Greven, P. H. Kes, M. Li, Nature **425**, 271 (2003).

[117] see, e.g., J. Custers, P. Gegenwart, H. Wilhelm, K. Neumaier, Y. Tokiwa, O. Trovarelli, C. Geibel, F. Steglich, C. Pépin, P. Coleman, Nature **424**, 524 (2003); T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, M.P.A. Fisher, Science **303**, 1490 (2004).

[118] N.E. Hussey, M. Abdel-Jawad, A. Carrington, A. P. Mackenzie, L. Balicas, Nature **425**, 814 (2003).

[119] L. Zhu, M. Garst, A. Rosch and Q. Si, Phys. Rev. Lett. **91**, 066404 (2003).

[120] R. Kuchler, N. Oeschler, P. Gegenwart, T. Cichorek, K. Neumaier, O. Tegus, C. Geibel, J.A. Mydosh, F. Steglich, L. Zhu, Q. Si, Phys. Rev. Lett. **91**, 066405 (2003).

[121] J.W. Loram, K.A. Mirza, J.R. Cooper and W.Y. Liang, Phys. Rev. Lett. **71**, 1740 (1993); Physica C **253**, 134 (1994); J.W. Loram, J. Luo, J.R. Cooper, W.Y. Liang, J.L. Tallon, J. Phys. Chem. Sol. **62**, 59 (2001); J.L. Tallon and J.W. Loram, Physica C **349**, 53 (2001).

[122] S.H. Naqib, J.R. Cooper, J.L. Tallon, R.S. Islam, R.A. Chakalov, Phys. Rev. B (submitted)(cond-mat/0312443).

[123] A.Ino, T. Mizokawa, A. Fujimori, K. Tamasaku, H. Eisaki, S. Uchida, T. Kimura, T. Sasagawa, K. Kishio, Phys. Rev. Lett. **79**, 2101 (1997); N. Harima, A. Fujimori, T. Sugaya, I. Terasaki, Phys. Rev. B **67**, 172501 (2003).

[124] M. Matlak and M. Pietruszka, Phys. Stat. Sol. B **231**, 299 (2002); A. Pimenov, A. Loidl, D. Dulic, D. van der Marel, I.M. Sutjahja, A.A. Menovsky, Phys. Rev. lett. **87**, 177003 (2001).

[125] G. Rietveld, N.Y. Chen and D. van der Marel, Phys. Rev. Lett. **69**, 2578 (1992); D. van der Marel and G. Rietveld, Phys. Rev. Lett. **69**, 2575 (1992).

[126] S. Ilani, A. Yacoby, D. Mahalu and H. Shtrikman, Phys. Rev. Lett. **84**, 3133 (2000); Science **292**, 1354 (2001).

[127] S.H. Dultz and H.W. Jiang, Phys. Rev. Lett. **84**, 4689 (2001).

[128] K. M. Mertes, D. Simonian, M. P. Sarachik, S. V. Kravchenko, and T.M. Klapwijk, Phys. Rev. B **60**, R5093(1999)

[129] J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, Phys. Rev B **50**, 1760(1994).

[130] S. Shapira, U. Sivan, P.M. Solomon, E. Buchstab, M. Tischler, G.B. Yoseph, Phys. Rev. Lett. **77**, 3181(1996).

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

[132] A. Yacoby, H.F. Hess, T.A. Fulton, L.N. Pfeiffer, K.W. West, Solid. State. Comm. **111**, 1 (1999).

[133] Y. Y. Wei, J. Weis, K. von Klitzing, K. Eberl, Phys. Rev. Lett. **81**, 1674 (1998).

[134] L. P. Kouwenhoven *et al*, in *Mesoscopic Electron Transport*, edited by L. L. Sohn, L. P. Kouwenhoven, and G. Schon, *NATO ASI series* (Kluwer Academic Publishers, 1996).

[135] B. L. Altshuler, and A. G. Aronov, Solid State Commun. **30**, 115(1979); Zh. Eksp. Teor. Fiz. **77**, 2028(1979) [Engl. transl.: Sov. Phys.-JETP **50**, 968(1979).

[136] H. Fukuyama, J. Phys. Soc. Jpn. **48**, 2169(1980).

[137] S. C. Dultz, B. Alavi, H. W. Jiang, cond-matt/0210584.

[138] J. Shi, and X. C. Xie, Phys. Rev. Lett. **88**, 86401 (2002).

[139] F. W. Van Keuls, H. Mathur, H. W. Jiang, and A. J. Dahm, Phys. Rev. B 56, 13263 (1997).

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

[141] L. H. Thomas, Proc. Cambridge Philos. Soc. **23**, 542 (1927).

[142] E. Fermi, Z. Phys. **48**, 73 (1928).

[143] P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964).

[144] W. Kohn, L. J. Sham, Phys. Rev. **140**, A 1133 (1965).

[145] Y. Imry, and M. Wortis, Phys. Rev. B **19**, 3580(1979).

[146] Y. Wang, J. Wang, H. Guo, and E. Zaremba, Phys. Rev. B **52**, 2738(1995).

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

[148] C. F. von Weizsäcker, Z. Phys. **96**, 431(1935).

[149] B. I. Shklovskii, A. L. Efros, JETP Lett. **44**, 123(1986).

[150] M. Rahimi, M. R. Sakr, S. V. Kravchenko, S. C. Dultz, H. W. Jiang, Phys. Rev. B **67**, 081302(R) (2003).

[151] D. Pines, and P. Nozieres, *The Theory of Quantum Liquids*, vol. I, Addison-Wesely, Reading, MA (1966).

[152] N. Iwamoto, Phys. Rev. B **43**, 2174 (1991).

[153] Y. Kwon, D. M. Ceperley, and R. M. Martin, Phys. Rev. B **50**, 1684 (1994).

[154] G. H. Chen, and M. E. Raikh, Phys. Rev. B **60**, 4826 (1999).

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

[156] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterlings, "*Numerical Recipes in Fortran*", Cambridge University Press; 2 edition (1992).

Summary

This thesis, called "*On The Thermodynamics at Quantum Phase Transitions in Two Dimensional Electron Systems*", is dedicated to the theory of the metal-insulator transition in two-dimensional electron systems. The distinction between metals and insulators appears at first sight to be simple - metals conduct electricity whereas insulators do not. Interestingly, for the past 25 years, arguments have raged over whether a two-dimensional (2D) electron system can become a real metal at all. In a 2D system - such as a very thin metal film, or the active region of many semiconductor transistors - the electrons are constrained to move in a plane of negligible thickness. Although these structures may conduct at room temperature, it was generally accepted that as the temperature (T) is reduced to absolute zero they have to become insulating. This is, in fact, based on the theory by Anderson which won the Nobel prize in 1977; namely Anderson (weak) localization. Therefore, for a long time it was believed that according to experiments the 2D electron system is indeed always insulating.

The notion that 2D metals cannot exist dates back 20 years, when powerful "scaling" theories indicated that any amount of disorder would trap electrons, so preventing conduction and the existence of a metallic state. These arguments are based on the quantum-wave nature of the electrons, whereby a traveling electron wave can be scattered from impurities back to its starting point. If these returning waves interfere constructively, the electrons become localized in one place and are less able to diffuse through the solid. At high temperatures this effect is weak and the sample appears metallic. As the temperature is reduced quantum interference becomes more important, so that at absolute zero all the electrons are localized and completely unable to move. Low-temperature experiments with both thin-metal films and 2D sheets of electrons in field-effect transistors confirmed these theoretical predictions, and for nearly two decades it was generally accepted that there can be no 2D metal.

In 1994, the community was shocked when it was discovered experimentally that these 2D electron systems can actually become metals and even exhibit a metal-insulator transition (MIT) in two-dimension. These transitions were found in extremely low-disorder silicon field-effect transistors, in apparent contradiction with the prevailing scaling theory. Similar behavior was subsequently observed in other material systems, independent of the sign of the charge carriers, indicating that the metallic state is a universal property of all low-disorder 2D systems, and

the metal-insulator transition does exist in 2D. At present, there is no theoretical consensus as to the nature of this unusual metallic phase, but experiments suggest that strong interactions between the charge carriers (not considered in the original scaling theory) play a significant role.

According to quantum mechanics electrons have a dual nature: particles and waves. The localization theory of Anderson deals with the wave aspect of electrons and we just discussed its prediction in 2D. The particle aspect is manifested through the effect that the electrons have mutual electromagnetic interactions. At the meta-insulator transition (seen in the above systems) the electron-electron interaction energy is 10 times as large as the kinetic energy. This condition can be parameterized by a dimensionless parameter, the ratio of the Coulomb energy to the kinetic energy of the electrons, r_s. Therefore one can say that in these systems r_s is larger than 10. In the absence of disorder and when interactions are weak the behavior of the 2D electron system is well understood based on the Fermi-liquid theory. The Fermi liquid is qualitatively analogous to the non-interacting Fermi-gas, in the following sense: the system's dynamics and thermodynamics at low energies (small r_s) and temperatures may be described by substituting for the non-interacting fermions so-called "quasi-particles", each of which carries the same spin, charge and momentum as the original particles. For large density of electrons this theory predicts a metallic phase for the electron system in the absence of disorder.

Remarkably, at low enough densities, due to the effect of strong Coulomb interaction, the electrons form a regular crystal structure known as Wigner crystal. Now generally accepted that, in the presence of a disorder potential, at low densities electrons are pinned not in a regular form of crystal but rather in a random position known as Wigner glass; while the experiments indicate Fermi-liquid at high densities. Although the scaling theory does not explicitly consider the effect of the Coulomb interaction between electrons, early theoretical works (in the 1980's) predicted that weak electron-electron interactions increase the localization even further. This was taken as a prove of Anderson localization which might happen at very low temperatures including the effect of interactions. But clearly the experiments in 1994 showed that this might not be true. To date, no analytical theory has been developed in the limit of strong interactions. The old problem of the interplay of disorder and electron-electron interaction is presented here in an extreme limit. Various explanations have been suggested based on the various physical processes which might be expected in these systems. Chapter 2 briefly reviews some of the most important of these theoretical models.

To date, most of the informations on the properties of the 2D metallic system near the metal-insulator transition comes from the transport measurements. Recently there have been very interesting and crucial experimental findings which probe the change in the thermodynamic behavior of the 2D electron system near the MIT. These experiments include the measurement of the electronic compressibility and the (thermodynamical) spin susceptibility. The importance of these measurements can be emphasized based on -at least- two reasons: first, they are quite free

of any complications related to the quantum coherency which one may encounter in transport measurements and second, if a true zero temperature phase transition exist, it may reveal itself also in the thermodynamics. These experiments indicate a very clear change in the thermodynamic properties across the MIT. For example, the inverse of electronic compressibility exhibits an unusual anomaly across the MIT point in the 2D electron (hole) systems. A sharp enhancement of the spin susceptibility is also reported in these systems upon approaching the metal-insulator transition. Based on the Fermi-liquid theory, this observation might also be related to the enhancement of the effective mass of the charge carriers at very low densities. Chapter 4 and the first section of chapter 9 look into these experiment in more detail.

This thesis deals with theoretical considerations of these (thermodynamical) quantities. This is in fact very important as to this date almost all of the theoretical attempts for understanding the metal-insulator transition are biased by focusing on the transport properties. With a focus on the compressibility and effective mass, this thesis deals with two theoretical issues:

1) *The screening properties of the 2D electron system*: Over the years, it has been suggested that the interacting 2D electrons in the presence of disorder can be understood by the so-called "droplet" model. According to this picture, due to the presence of disorder, the electrons are separated into "liquid" droplets with local density higher than the average density and "vapor" islands with lower density. For a given disorder, due to the potential fluctuation of the donors, as the Fermi energy is decreased, the "vapor" region increases. In this model, a crossover from metallic behavior to insulating behavior should be seen around the percolation threshold. Although this model might work well in relatively high densities, we will make the case in chapter 5 that for the highly mobile (weakly disordered) and strong interacting electrons (holes) this kind of physics will fail. In fact, all percolation like scenarios need to start with quite inhomogeneous metallic phase, by incorporating either screened Coulomb interaction or strong and -at the same time- smoothed disorder potential. Both assumptions prove to be wrong according to the experimental evidences.

A possible correct description of the screening properties of the 2D electron liquid near the MIT has been suggested recently by Si and Varma. According to this theory, at low enough densities and weak disorders, the screening length (the length beyond which the electrons do not interact) of a disordered interacting electron gas may become larger than its mean free path (the distance electrons can move freely without colliding with each other). The compressibility, being proportional to the inverse screening length, is expected to vanish as the transition to the insulating state is approached. This, indeed, is the most important physical concept behind the line of a research carried out in this thesis. Quite interestingly, the effect of this screening properties of the electron liquid at low densities and low disorders can be expressed in a simple way in terms of a local density of the charge carriers in the 2D system. This in turn allows us to construct a density functional theory (DFT) in which, the total energy of the electrons in 2D system in

the metallic regime can be expressed entirely in terms of the their local density.

By employing the numerical simulation, we will show that, although the "critical" density where the compressibility changes its behavior implied by these calculations is smaller than the experimentally observed critical density, but such a DFT formulation for 2D electrons indicates a substantial improvement for the description of the behavior of compressibility near the MIT critical point. At the same time, we will see that due to the fact that the screening length becomes large, it forces the electron liquid to become increasingly incompressible with the obvious effect that density fluctuations are suppressed.

2) *Metal-insulator quantum phase transitions:* there also have been suggestions that the observed (continuous) metal-insulator transition in the 2D electron system at low temperatures might be a reminiscence of the more robust zero temperature Quantum Phase Transition (QPT), driven by interaction and disorder in the system. QPTs are a type of transitions, which are accessed at zero temperature by variation of a non-thermal control parameter (in contrast to classical phase transitions), and these can influence the behavior of electronic systems over a wide range of temperatures. Generally speaking, quantum phase transitions occur as a result of competing ground state phases. In general, quantum critical points are more difficult to characterize compared to their classical counterparts. Quite remarkably, it has been shown recently by Rosch and Si that using a very simple thermodynamic ratio one can, in fact, probe the existence of quantum critical point in a critical system. They showed that at the generic quantum critical point, the ratio of the thermal expansion and the specific heat, known as Grüneisen ratio, diverges. Employing the scaling analysis of Rosch and Si and using elementary power counting arguments, we have discovered an empirical strategy which should make it possible to decide if the 'quantum criticality' of the cuprates has to do with universality (see chapter 3). Using the same analysis related to the scaling behavior of effective mass and compressibility at the critical region, we also address the same question for 2D electrons in MOSFETs and give the experimental strategy to determine if the MIT is a genuine quantum phase transition or not (see chapter 9).

Samenvatting

Dit proefschrift, genaamd "Over de Thermodynamica van Kwantum Fase Overgangen in Twee-Dimensionale Elektronen Systemen", gaat over de theorie van de metaal-isolator overgang in twee-dimensionale elektronen systemen. Het onderscheid tussen metalen en isolatoren lijkt op het eerste gezicht simpel te zijn: metalen geleiden elektriciteit terwijl isolatoren dat niet doen. Toch woedt de laatste 25 jaar al een discussie of een twee-dimensionaal (2D) elektronen systeem een echt metaal kan worden of niet. In een 2D systeem, zoals een zeer dunne metalen film, of het actieve gebied van veel halfgeleider transistoren, zijn elektronen gedwongen om te bewegen in een vlak van verwaarloosbare dikte. Hoewel deze structuren bij kamertemperatuur stroom geleiden, is het algemeen geaccepteerd dat zij isolerend worden als de temperatuur (T) afneemt tot het absolute nulpunt. De verklaring voor dit feit is gebaseerd op de Anderson (zwakke) lokalisatie theorie waarmee Mott en Anderson in 1977 de Nobelprijs voor Natuurkunde wonnen. Daarom werd gedurende lange tijd geloofd dat het 2D elektronen systeem altijd isolerend is.

De notie dat 2D metalen niet kunnen bestaan dateert zo'n 20 jaar terug, toen krachtige "schalings" theorieën aangaven dat elke willekeurige hoeveelheid "disorder" elektronen gevangen zou houden. Op die manier wordt geleiding en dus het bestaan van een metallische toestand voorkomen. Deze argumenten zijn gebaseerd op de kwantum-golf natuur van de elektronen. Hierdoor kan een bewegende elektronengolf door onzuiverheden terugverstrooid worden naar zijn startpunt. Als deze terugkomende golven constructief interfereren kunnen de elektronen gelokaliseerd worden in één plaats en zijn ze hierdoor niet in staat om door een vaste stof te bewegen. Bij hoge temperaturen is dit effect zwak en lijkt het preparaat metallisch. Bij afnemende temperatuur wordt de kwantum interferentie belangrijker, totdat bij het absolute nulpunt alle elektronen gelokaliseerd zijn en hierdoor niet in staat zijn om te bewegen. Lage temperatuur experimenten met zowel dunne metaalfilms als met 2D-sheets van elektronen in veldeffect transistoren bevestigden deze theoretische voorspellingen en gedurende bijna twee decennia was het algemeen geaccepteerd dat er geen 2D metaal kan bestaan.

In 1994 werd de gemeenschap echter wakker geschud toen experimenteel werd aangetoond dat 2D elektron systemen toch metallisch kunnen worden en zelfs een metaal-isolator overgang kunnen laten zien. Deze overgangen werden gevonden in extreem zuivere silicium veldeffect transistoren, in schijnbare tegenspraak met de beschikbare schalingstheorie. Vergelijkbaar gedrag werd hierna waargenomen

in systemen bestaande uit andere materialen, met zowel positieve als negatieve ladingsdragers. Dit gaf aan dat de metallische toestand een universele eigenschap is van alle 2D systemen met een lichte mate van disorder. Op dit moment is er geen theoretische consensus over de essentie van deze ongewone metallische fase, maar de experimenten suggereren dat sterke wisselwerkingen tussen de ladingsdragers (dat niet meegenomen wordt in de originele schalingstheorie) een belangrijke rol spelen.

Volgens de kwantummechanica hebben elektronen een duaal karakter: ze zijn zowel deeltjes als golven. De lokalisatie theorie van Anderson, zoals besproken in hoofdstuk 2, gaat uit van de golf aspecten van de elektronen. Het deeltjes aspect komt tot uitdrukking door het effect dat de elektronen een gemeenschappelijke elektromagnetische wisselwerking hebben. Bij de metaal-isolator overgang (zoals gezien in bovenstaande systemen) is de elektron-elektron interactie energie 10 keer zo groot als de kinetische energie. Dit gegeven kan worden weergegeven door een dimensieloze parameter, de verhouding van de Coulomb energie tot de kinetische energie van de elektronen, r_s. Daarom kan men zeggen dat in deze systemen r_s groter is dan 10. Bij afwezigheid van disorder en een zwakke wisselwerking kan het gedrag van 2D elektronen systeem goed begrepen worden met de Fermi-vloeistof theorie. Een Fermi-vloeistof is op de volgende manier kwalitatief analoog aan een Fermi-gas zonder wisselwerkingen: de dynamica en thermodynamica van het systeem bij lage energie (kleine r_s) en temperatuur kan worden beschreven door de werkelijke elektronen te vervangen door zogenaamde "quasi-deeltjes": niet-wisselwerkende fermionen welke dezelfde spin, lading en impuls dragen als de originele deeltjes. Voor grote dichtheden van elektronen voorspeldt deze theorie, in de afwezigheid van disorder, een vloeibare fase voor het elektronensysteem.

Verrassend genoeg vormen de elektronen bij lage dichtheden, ten gevolge van het effect van de sterke Coulomb wisselwerking, een zeer reguliere kristalstructuur dat bekend staat als het Wigner kristal. Het wordt nu algemeen aanvaard dat in de aanwezigheid van een disorder potentiaal, en bij lage dichtheden, de elektronen wel gepinned worden, maar dan op willekeurige posities. Dit wordt een Wigner glas genoemd. Bij hogere dichtheden laten de experimenten een Fermi-vloeistof zien. Hoewel de schalingstheorie niet expliciet rekening houdt met de invloed van Coulomb wisselwerking tussen de elektronen, laat eerder theoretisch werk (uit de jaren tachtig) zien dat zwakke elektron-elektron wisselwerking de lokalisatie verder kan versterken. Dit werd gezien als bewijs dat Anderson lokalisatie plaats kon vinden bij zeer lage temperaturen als de effecten van interactie werden meegenomen. Maar de experimenten uit 1994 lieten duidelijk zien dat dit het mogelijk niet waar is. Tot nu toe is er geen analytische theorie ontwikkeld voor de limiet van sterke wisselwerkingen. Het oude probleem van de wisselwerking tussen disorder en elektron-elektron wisselwerking is hier aanwezig in een extreme limiet. Verschillende verklaringen zijn gesuggereerd gebaseerd op de verschillende fysische processen die plaats kunnen vinden in deze systemen. Hoofdstuk 2 gaat kort in op enkele van de belangrijkste van deze theoretische modellen.

Tot op dit moment komt de meeste informatie over de eigenschappen van het 2D metallische systeem in de buurt van de metaal-isolator overgang uit transport metingen. Recentelijk zijn er een aantal belangrijke experimentele metingen gedaan welke meer informatie leveren over de verandering van het thermodynamisch gedrag van het 2D elektron systeem vlak bij de MIT. Dit zijn de kompressibiliteit en de (thermodynamisch) spin susceptibiliteit metingen. Deze experimenten zijn vanwege twee redenen belangrijk: ten eerste zijn ze vrij van complicaties gerelateerd aan de kwantum coherentie welke men tegenkomt in transport metingen. Ten tweede, als een zero temperatuur fase overgang bestaat, kan het zichzelf ook laten zien in de thermodynamica. Deze experimenten laten duidelijk zien dat bijvoorbeeld de elektronische compressibiliteit een ongewone verandering laat zien in de buurt van het MIT punt in 2D elektronen (of gaten) systemen. Tevens wordt een scherpe toename van de spin susceptibiliteit gerapporteerd in deze systemen bij nadering van de metaal-isolator overgang. Gebaseerd op de Fermi-vloeistof theorie kan deze observatie worden gerelateerd aan de toename van de effectieve massa van de ladingsdragers bij zeer lage dichtheden. In hoofdstuk 4 en de eerste sectie van hoofdstuk 9 wordt in meer detail gekeken naar deze experimenten.

Dit proefschrift kijkt naar de theoretische achtergronden van deze (thermodynamische) eigenschappen. Dit is zeer belangrijk omdat tot nu toe vrijwel alle theoretische pogingen voor een begrip van metaal-isolator overgangen gebaseerd zijn op de transport eigenschappen. Wij kijken naar twee aspecten van de thermodynamica van deze systemen (met een focus op de compressibiliteit en effectieve massa):

1) *De afschermingseigenschappen van 2D elektronen systemen*: gedurende de laatste jaren is er gesuggereerd dat wisselwerkende 2D elektronen in aanwezigheid van disorder kunnen worden begrepen met het zogenaamde "druppel" model. Volgens dit beeld worden door de aanwezigheid van disorder de elektronen opgesplitst in "vloeibare" druppels met een lokale dichtheid groter dan de gemiddelde dichtheid en "gasvormige" eilanden met een lagere dichtheid. Voor een gegeven sterkte van de disorder neemt het "gasvormige" gebied toe als de Fermi energie afneemt ten gevolge van de potentiaal fluctuaties van de donoren. In dit model zou een crossover van metallisch gedrag naar isolerend gedrag zichtbaar moeten zijn bij de percolatie drempel. Hoewel dit model goed zou moeten werken voor relatief hoge dichtheden, laat hoofdstuk 5 zien dat voor zeer mobiele (zwakke disorder) en sterke wisselwerkende elektronen en gaten dit type van fysica niet correct is. In feite moeten namelijk alle op percolatie gebaseerde scenarios starten met een sterk niet homogene metallische fase, door òf gebruik te maken van een afgeschermde Coulomb wisselwerking òf sterke en afgevlakte disorder potentiaal. Beide aannames blijken foutief te zijn volgens de experimentele metingen.

Een mogelijke correcte beschrijving van de afschermingseigenschappen van 2D elektronenvloeistoffen vlakbij de MIT werd recentelijk gesuggereerd door Si en Varma. Als de dichtheid laag genoeg is en er een zwakke disorder is, dan wordt de afschermingslengte - de lengte waarboven de elektronen geen interactie meer

hebben - van een disordered wisselwerkend elektronen gas groter dan zijn gemiddelde vrije weglengte - de afstand elektronen kunnen bewegen zonder met elkaar te botsen. Hierdoor kan verwacht worden dat de compressibiliteit, die evenredig is met de omgekeerde afschermingslengte, verdwijnt als de overgang naar de isolerende toestand wordt genaderd. Dit is het belangrijkste fysische idee achter de onderzoekslijn dat gevolgd wordt in dit proefschrift.

De effecten van deze afscherming van de elektronen vloeistof bij lage dichtheden en bij lage disorder kan op een verbazingwekkend eenvoudige manier worden verdisconteerd in een lokale dichtheidsfunctionaal. Dit kan aan de metallische kant van de MIT gebruikt worden in een dichtheids functionaal theorie (DFT) formalisme. Door gebruik te maken van numerieke simulaties wordt de "kritische" dichtheid waar de compressibiliteit van gedrag verandert, berekend. Hoewel het resultaat uit deze berekeningen kleiner is dan de experimenteel geobserveerde kritische dichtheid, laat de DFT formulering in dit proefschrift echter een duidelijke verbetering van de beschrijving van compressibiliteit vlak bij het MIT overgangspunt zien. Zoals we zullen zien wordt tegelijkertijd, tengevolge van het feit dat de afschermingslengte groot wordt, de elektronenvloeistof gedwongen om steeds meer niet comprimeerbaar te worden met als duidelijk effect dat de dichtheidsfluctuaties worden onderdrukt.

2) *Metaal-isolator kwantum fase overgang*: er zijn ook suggesties dat de waargenomen (continue) metaal-isolator overgang in 2D elektron systemen bij lage temperaturen een overblijfsel is van een meer robuust kwantum fase overgang bij het absolute nulpunt. Deze overgang wordt gedreven door wisselwerking of disorder (of beide) in het systeem. Kwantum fase overgangen zijn een type van faseovergangen die plaatsvinden bij het absolute nulpunt, door verandering van een niet-thermische controle parameter (zoals wel het geval is bij klassieke fase overgangen). Zij kunnen het gedrag van het elektronische systeem over een groot temperatuurbereik beïnvloeden. In het algemeen vinden kwantum fase overgangen plaats ten gevolge van concurrerende grondtoestanden.

In het algemeen zijn kwantum kritische punten moeilijker te karakteriseren dan hun klassieke tegenhangers. Recentelijk hebben Rosch en Si echter laten zien dat een zeer eenvoudig thermodynamische ratio inzicht geeft over een generiek kwantum kritisch punt. Zij laten zien dat de verhouding van de thermische uitzetting en de soortelijke warmte, bekend als de Gruneisen verhouding, bij een kwantum kritisch punt divergeerd. Gebruik makend van elementaire dimensionale analyse argumenten ("powercounting"), suggereert dit proefschrift een nieuwe empirische strategie welke het mogelijk maakt om te beslissen of de "kwantum kritikaliteit" van de cuprates gerelateerd is aan de universaliteit. Door gebruik te maken van dezelfde schalings analyse, gerelateerd aan het schalingsgedrag van de effective massa en de reciproke compressibiliteit, gaan we ook in op dezelfde vraag voor 2D elektronen systemen en geven een experimentele strategie om te bepalen of de MIT een echte kwantum fase overgang is of niet.

Publications

1. *Effects of quantum fluctuations in the slave-bosons approach to the Hubbard model*, B. Hosseinkhani, Condensed Matter Theories, Vol. 14, by David J. Ernst (Editor), *et al*, Nova Science Publishers Inc; ISBN: 1560728116; (2000).

2. *Thermodynamics and quantum criticality in cuprate superconductors*, J. Zaanen and B. Hosseinkhani, Phys. Rev. B, Rapid Communications, in press.

3. *Quantum criticality and the metal-insulator transition in 2D: a critical test*, B. Hosseinkhani and J. Zaanen, submitted to Physical Review Letters.

4. *A density functional description for the disorder driven screening length divergence in the 2D electron gas*, B. Hosseinkhani and J. Zaanen, submitted to Physical Review B.

Curriculum Vitae

I was born in 27th of August 1973 in the old city of Kermanshah in the west part of Iran. From 1979 to 1992, I did my primary and high school study in the port city of Bandar-e-Abbas, in the south of the country next to the Persian Golf. In 1992 after taking the national entrance examination I started my study in Tehran University and graduated in 1997 with a B.Sc. degree in Applied Physics. My B.Sc. project was *"Quantum Rotor as a Model for Electron Spin: Path Integral Approach"* , under the supervision of Prof. S. Bayegan. In 1993 I awarded one year studentship from Iranian Physical Society as one of the first rank students in Physics departments in universities of Tehran. From 1997 to 1999 I did my Master study in Theoretical Physics at the Institute for Advanced Studies in Basic Sciences (IASBS) in Zanjan (Iran). My Supervisor was Prof. S. Azakov and the title of my thesis was *"Quantum Fluctuations in Slave Boson Approach to the Single-band Hubbard Model"*. The work of my Master thesis also accepted for a short oral presentation on the XXII International Workshop on Condensed Matter Theories, Vanderbilt, TN, USA (June 1998). In 1999 I started my PhD work in the Instituut-Lorentz in Leiden as a Junior Research Assistance (OIO), employed by The Foundation for Fundamental Research in Matter (FOM). My PhD research work was carried out in the field of Theoretical Condensed Matter Physics and under the supervision of Prof. Jan Zaanen. During this period I attended several national and international conferences. In April 2000 I took part in the major NATO research school on the "New Theoretical Approaches to Strongly Correlated System" at the Isaac Newton Institute for Mathematical Sciences in Cambridge (UK). I also was teaching assistance for the course of "Electrodynamics" and "Advanced Condensed Matter Theory" in the Physics department of Leiden University.

Acknowledgments

I spent such a great time in the Insituut-Lorentz during my PhD study. In this respect I would like to acknowledge all the people who created such a pleasant and stimulating environment in which I was able to conduct my study and research. First of all I want to mention Hanni Andela who I am indebted to her as she was the first person I met from the institute and helped me a lot to get used to my new life in Leiden. I am particularly very grateful to Chandra Varma who suggested me to pursue this line of research during his visit as a Lorentz Professor in Leiden. I particularly enjoyed lots of useful discussions on disordered systems with Philippe, Peter, Zohar and Inanc. I learned a great deal of Linux programming from Michael and Marco who also helped me for the dutch summary of my thesis. I also would like to acknowledge Qimiao Si (Rice Univ) for his critical and helpful comments on my work, S. Ilani (Wiesmann Institute) and H.W. Jiang (UCLA) for providing me with their experimental data. It was great pleasure to work with all the member of the condensed matter theory group both as colleagues and as friends.

Finally I would like to thank my cat Kitty who was the only one staying awake with me during the nights I was writing my thesis. I am also very grateful to Leticia whom, being very patient, without her support finalizing this thesis was impossible.

Stellingen

belonging to the thesis

On The Thermodynamics at Quantum Phase Transitions in Two Dimensional Electron Systems

1. It is remarkable that it has taken approximately a decade for the realization to emerge that simple thermodynamic quantities[1] can be used as critical tests for quantum criticality.
   Chapter 3 of this thesis
   [1] *L. Zhu, M. Garst, A. Rosch and Q. Si, Phys. Rev. Lett. 91, 066404 (2003).*

2. The local compressibility measurements of the 2D electron system near the metal-insulator transition[1] has been taken by the community at large as evidence for strong density inhomogeneity. Unfortunately, on general ground one expects that the relation between density and the compressibility has to be highly non-linear in this regime.
   Chapter 4 of this thesis
   [1] *S. Ilani, A. Yacoby, D. Mahalu, and H. Shtrikman, Science 292, 1354 (2001).*

3. Spivak has argued, based on classical electrostatics, that the 2D electron system in MOSFETs should have an intrinsic tendency towards the formation of density inhomogeneities[1]. However, the metallic state becomes increasingly incompressible upon approaching the metal-insulator transition and this diminishes these effects completely.
   Chapter 8 of this thesis
   [1] *B. Spivak, Phys. Rev. B 67, 125205 (2002).*

4. The popularity of the idea that the metal-insulator transition in two dimensions is associated with percolation physics seems more given in by the appealing simplicity of the idea than by empirical fact.
   Chapter 9 of this thesis

5. In the interpretation of the STM information on charge ordered structures it should be realized that one is dealing with the microscopic electronic structure. For instance, the "fork" effect of Martin *et al.*[1] leads to a blurring of these structures in the actual STM images[2].
   [1] *I. Martin, A.V. Balatsky and J. Zaanen, , Phys. Rev. Lett. 88, 097003 (2002).*
   [2] *T. Hanagari, C. Lupien, Y. Koshaka, D. H. Lee, M. Azuma, M. Takano, H. Tagachi and J. C. Davis, Nature, in press.*

6. When Hawking would take his most recent idea about the black holes[1] completely serious, he should abandon the notion that the event horizon and the related holographic principle are sharply defined.
   [1] *S. Hawking,* "The information paradox for black holes", *abstract, The 17th International Conference on General Relativity and Gravitation, 18th - 23rd July 2004 in Dublin, Ireland.*

7. It is questionable if theoretical biophysics, defined as being a continuation of statistical physics, will ever contribute substantially to the understanding of the important questions in biology.

8. All organisms with complex nervous systems are faced with the moment-by-moment question that is posed by life: What shall I do next?

9. Is it a coincidence that both Adam and Newton changed the nature of human existance because of an apple, or is it so that God's blessing always comes with a forbidden fruit?!

10. According to the latest survey 36% of dutch people believe that in fact Palestinians have occupied Israel. This might not be an incident, given the fact that the Dutch Prime Minister Balkenende during his visit of Dutch troops in Iraq on January 7th 2004[1] called this country Iran!
[1] *Volkskrant, 7-1-2004*

11. -¿Donde vas?
-¡Manzanas traigo!
Margarita

Babak Hosseinkhani
8 September 2004