Dilute Electron Gas near the Metal-Insulator Transition: Role of Valleys in Silicon Inversion Layers

Alexander Punnoose and Alexander M. Finkel'stein

Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot, 76100, Israel

(Received 25 May 2001; published 18 December 2001)

We emphasize the role of valleys in the transport properties of the dilute electron gas in Si-MOSFETs. Close to the critical region of the metal-insulator transition the decrease in the resistivity up to 5 times has been captured in the correct temperature interval by a renormalization group analysis of the interplay of interaction and disorder. No adjustable parameters are involved in the analysis if the electron band is assumed to have two distinct valleys. The considerable variance in the data obtained from Si-MOSFET samples of different quality is attributed to the sample-dependent scattering rate across the two valleys, while universal behavior is expected to hold when the intervalley scattering is negligible.

DOI: 10.1103/PhysRevLett.88.016802 PACS numbers: 72.10.–d, 71.30.+h, 71.10.Ay

The resistivity in a variety of high mobility two-dimensional (2D) electron/hole systems is seen experimentally to exhibit a number of interesting anomalies that do not, as yet, have an adequate theoretical understanding. (For an extensive bibliography, see Ref. [1].) The high quality of the samples allows measurements to be made at very low carrier densities corresponding to $r_s \gtrsim 10$, where $r_s = E_{e\text{-}e}/E_F$ is the ratio of the Coulomb energy to Fermi energy. When the resistivity at high temperatures is comparable to or less than the quantum resistance, h/e^2, the resistivity, $R_\square(T)$, drops noticeably as the temperature is reduced [2]. The drop appears to be completely quenched when a magnetic field is applied parallel to the plane [3]. This anomalously strong positive magnetoresistance, which is obviously related to the spin degrees of freedom, points to the importance of the electron-electron (e-e) interaction in this phenomenon.

These experimental observations have revived the fundamental question of localization or, alternatively, the existence of a metal-insulator transition in 2D systems in the presence of a strong e-e interaction.

Although the drop in the resistivity of dilute systems is generally considered to be universal, quantitative comparison indicates that the magnitude of the effect is very sensitive to the system used. The most pronounced anomaly has been reported in the cleanest (001) Si-MOSFET (metal-oxide-semiconductor field-effect transistor) samples, where a steep drop in $R_\square(T)$ of up to five to six times has been observed.

In Fig. 1 the temperature dependence of the resistivity for different densities in a Si-MOSFET sample with a high peak mobility has been reproduced [4]. The insulating region is labeled as **I** in Fig. 1. The range of densities where the resistivity depends nonmonotonically on temperature is labeled as $\mathbf{C}^*$. (For the specific sample used in Fig. 1, this region covers electron densities in the range $0.8 \times 10^{11} < n < 1 \times 10^{11}$ cm^{-2}.) A narrow range of densities between these two regions, in which the separatrix that separates the insulating phase from the metallic phase should lie (if a true metal-insulator transition exists), can be considered as the critical region **C**. The region with a drop in resistivity, but where no clear maximum is observed (unlike $\mathbf{C}^*$, is labeled as **M** in Fig. 1. The region $\mathbf{C}^*$ is the subject of this paper.

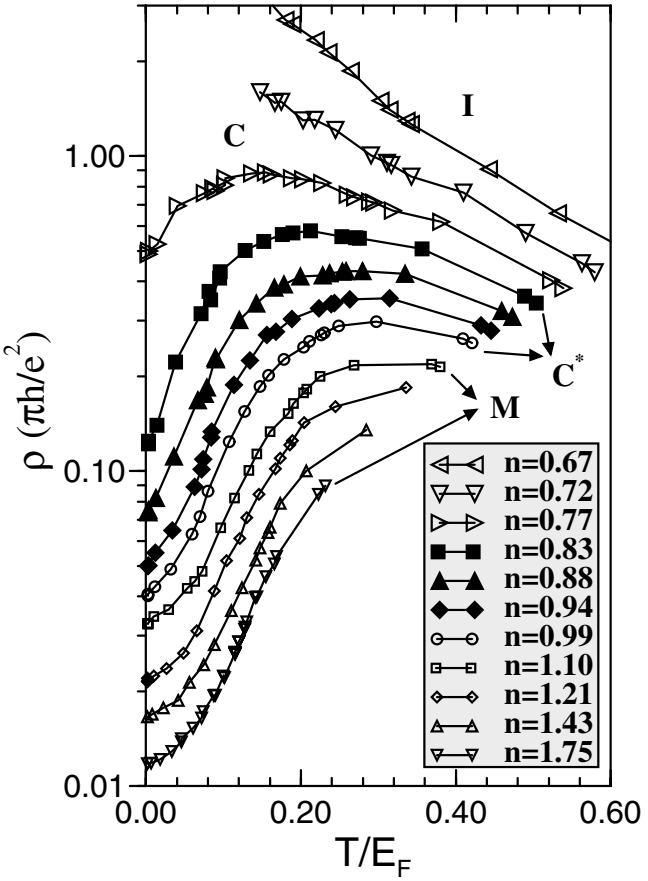

FIG. 1. Resistivity of a high mobility Si-MOSFET sample for various densities as a function of temperature. The electron densities, n, are defined in units of 10^{11} cm^{-2}. Data are reproduced from Fig. 1(a) of Ref. [4].

Since the maximum of the resistivity in the region $\mathbf{C}^*$ is comparable to $\sim h/e^2$, the transport mean-free path time, τ, of the electrons here is such that $\hbar/\tau \lesssim E_F$. The maximum together with the steep decrease in $R_\square(T)$ occurs at temperatures well below the Fermi energy (see Fig. 1). Therefore, the indications are that the nonmonotonic behavior of the resistivity in the region $\mathbf{C}^*$ is a manifestation of the physics of strongly interacting electrons that are in the diffusive regime: $T < \hbar/\tau \lesssim E_F$.

Away from the region $\mathbf{C}^*$ and deep in the region labeled **M** in Fig. 1 a naive estimate for τ can be extracted from the Drude expression, $\hbar/\tau = 4(e^2/h)E_F R_\square(T = 0)$, with $R_\square(T = 0)$ being the extrapolated value of the resistivity at $T = 0$. This estimate gives $\hbar/\tau$ that are well below the Fermi energy, while the steep drop in the resistivity develops at temperatures that are comparable to or larger than $\hbar/\tau$. This implies that the anomalies in the region **M** occur in the temperature range $\hbar/\tau \lesssim T < E_F$, and their origin may be attributed (at least partially) to a strong temperature dependence of the single particle mean-free path time $\tau(T)$ [5–7]. (For a recent discussion of the region $T > \hbar/\tau$, see Ref. [8].) On the contrary, since the anomalies in the region $\mathbf{C}^*$ are in the temperature range $T < \hbar/\tau \lesssim E_F$, the effects of thermal smearing of $\tau(T)$ are quenched by the disorder [7].

These considerations lead us to the conclusion that the anomalous decrease in the resistivity in the two regions, $\mathbf{C}^*$ and **M**, may have different origins and are hence best studied separately. In this paper we analyze the transport properties in the region $\mathbf{C}^*$, close to the critical region **C**, where the transport is controlled by the propagation of diffusive collective modes. We demonstrate that the phenomenon in this region can be understood within the framework of a theory describing the effect of the *e-e* interaction on the propagation of these modes. (For a review, see Ref. [9].) The peculiarity of dilute conductors is that at low temperatures the antilocalizing component of this effect becomes dominant.

Although universal behavior is generally expected to hold in the critical region, no universal scaling of the $R_\square(T)$ curves has been found. A considerable variance is seen even in the data obtained from Si-MOSFET samples of similar origin. Hence, for a quantitative understanding of the temperature dependence of the resistivity in the region not far from the transition some system-specific nonuniversal mechanism should be necessarily invoked. The conduction band of the electrons in a (001) Si-MOSFET surface has two almost degenerate valleys located at points $\pm Q_0$ [10]. In what follows, the sensitivity of the transport properties of the dilute electron gas to the scattering rate across the two valleys is presented as a possible explanation for the absence of this universality. At temperatures comparable to the rate of the intervalley scattering, $\hbar/\tau_\perp$, a crossover occurs between a band with two distinct valleys and a band where the two valleys are effectively unified due to the intervalley scattering. We believe that in a typical sample the value of $\hbar/\tau_\perp$ falls within the experimentally relevant temperature interval. Hence due to the crossover at $T \sim \hbar/\tau_\perp$ the resistivity $R_\square(T)$ will be nonuniversal. Only in an ultraclean sample (like the one presented in Fig. 1), where the intervalley scattering is weak and the two valleys are well separated, should universal behavior hold.

To understand the temperature dependence of $R_\square(T)$ in the case of two valleys, we study the interplay of the appropriate collective modes. These modes describe fluctuations of the local density of particles, spin, and the fluctuations involving electron states from different valleys. The evolution of the collective modes at large scales is described by a singular propagator with a diffusion pole $\propto 1/(Dq^2 - i\omega)$, where D is the diffusion constant. These singular propagators when combined with the *e-e* interaction are known to lead to the appearance of nonanalytical corrections to the resistivity. On the other hand, the amplitudes of the *e-e* interaction that affects the propagation of the collective modes are themselves known to have divergent corrections due to the disorder. The program to self-consistently take into account these corrections, which in fact corresponds to a derivation of a system of renormalization group (RG) equations, has been realized to lowest order in the resistivity (disorder), and fortunately to all orders in the *e-e* interaction amplitudes [11].

The diffusion propagators of the electron-hole pairs in the presence of two valleys in addition to spin quantum numbers are labeled by quantum numbers $|\tau\rangle$, where $|\tau\rangle = \pm$ are the two-valley indices similar to the up and down spin states $|\sigma\rangle = \uparrow, \downarrow$. Altogether there are $4 \times 4 = 16$ electron-hole states that break up into one singlet and fifteen multiplet states. In the case of strong intervalley scattering, however, the modes that are made of states from different valleys acquire a gap proportional to $\hbar/\tau_\perp$. This implies that for temperatures, or frequencies, less than $\hbar/\tau_\perp$ such modes do not yield diverging contributions and hence become ineffective [9,12,13]. (This is the origin of the crossover discussed above.) As a result, of the 16 modes only one spin singlet and three spin triplet combinations retain a diffusion pole, and the situation becomes equivalent to the case with no valleys.

In 2D the leading divergences are logarithmic. The RG equation describing the evolution of the resistivity [11,14] can be easily generalized to include two valleys:

$$\frac{d\rho}{d\xi} = \rho^2\left[n_v + 1 - (4n_v^2 - 1) \times \left(\frac{1 + \gamma_2}{\gamma_2}\ln(1 + \gamma_2) - 1\right)\right], \quad (1)$$

where $n_v = 1$ when $T < \hbar/\tau_\perp$, and $n_v = 2$ when $T > \hbar/\tau_\perp$ and the two valleys are distinct. Here $\xi = -\ln(T\tau)$ and the dimensionless parameter $\rho = (e^2/\pi h)R_\square$; note the additional factor $1/\pi$ that has been introduced in ρ. The first term in the square brackets corresponds to the quantum interference correction (Cooperon) in the presence of n_v valleys. The second term is the contribution of

the singlet density mode which, due to the long range nature of the Coulomb interaction, is universal [9]. The last term describes the contribution of the $(4n_v^2 - 1)$ multiplet modes. The parameter γ_2 is the Fermi liquid amplitude that controls the e-e interaction in all the multiplet channels normalized by the density of states for a single spin and valley species.

The contributions from the singlet and the multiplet modes differ in sign favoring localization and antilocalization, respectively. In conventional conductors the initial values of the amplitude γ_2 are small, and the net effect is in favor of localization. In dilute systems, however, this amplitude may be enlarged due to e-e correlations. The total effect, including the Cooperon contribution, will favor antilocalization if γ_2 is greater than an n_v-dependent value γ_2^*. As a result of the increase in the number of multiplet modes from 3 to 15, γ_2^* reduces considerably from $\gamma_2^* = 2.04$ for $n_v = 1$ to $\gamma_2^* = 0.45$ for $n_v = 2$. This strong reduction of γ_2^* makes it easier in the case of two distinct valleys to reach the stage where the resistivity starts to decrease.

In addition, in 2D the amplitude γ_2 also experiences logarithmic corrections due to the disorder [11,14]. The equation describing the RG evolution of γ_2 is the same for both one and two valleys:

$$\frac{d\gamma_2}{d\xi} = \rho \frac{(1 + \gamma_2)^2}{2}. \quad (2)$$

As the temperature is lowered γ_2 increases monotonically and when it increases beyond the value γ_2^* the resistivity will pass through a maximum. Although the initial values of ρ and γ_2 are not universal and depend on the system, the flow of ρ according to the RG equations can be described by a universal function $R(\eta)$ [11]:

$$\rho = \rho_{\max} R(\eta) \quad \text{and} \quad \eta = \rho_{\max} \ln(T_{\max}/T), \quad (3)$$

where $T_{\max}$ is the temperature at which ρ reaches its maximum value $\rho_{\max}$, i.e., $\gamma_2(T_{\max}) = \gamma_2^*$. For the case of two valleys, the function $R(\eta)$ is found here by numerically integrating Eqs. (1) and (2) with $n_v = 2$ and the boundary conditions: $\rho(\xi = 0) = \rho_{\max} = 1$ and $\gamma_2(\xi = 0) = \gamma_2^* = 0.45$.

Thus, if the experimental data of the resistivity are scaled as in Eq. (3), then the data should collapse on the function $R(\eta)$. This analysis has certain limitations, however. The RG equations have been derived in the lowest order in ρ and therefore cannot be applied in the critical region **C** where $\rho \gtrsim 1$. On the other hand, for $\rho \ll 1$, exponentially small temperatures are needed for changes in the resistivity to become noticeable. In addition, some other (not yet completely identified) mechanism operating in the region **M** may mask the discussed logarithmic corrections that are very weak when $\rho \ll 1$.

For these reasons, only curves in the region $\mathbf{C}^*$ with maximum ρ ranging from $\rho_{\max} \approx 0.3$ to $\rho_{\max} \approx 0.6$ have been used to test the RG analysis. The result is presented in Fig. 2. The decrease in the resistivity up to five times, together with its saturation, has been captured in the correct temperature interval by this analysis. Note that no adjustable parameters were used in the procedure. We emphasize again that this universal behavior will be observed only in ultraclean samples, and will not be found in samples that are only moderately clean, because of the crossover at $T \sim \hbar/\tau_\perp$.

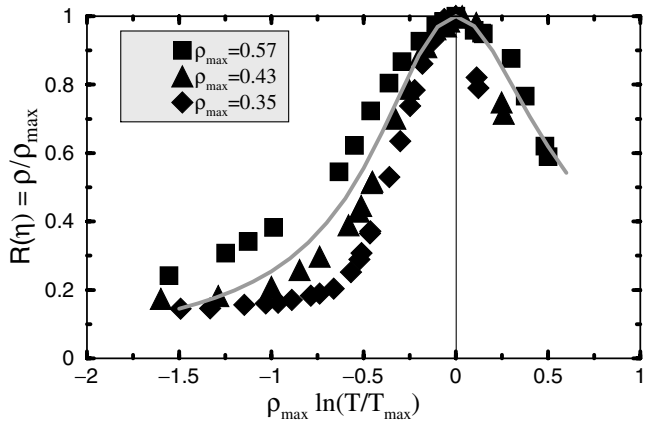

FIG. 2. The data corresponding to $n = (0.83, 0.88, 0.94) \times 10^{11}$ cm^{-2} in Fig. 1 are scaled according to Eq. (3). The solid line corresponds to the solution of the RG equations (1) and (2) with $n_v = 2$; no adjustable parameters have been used in this fit.

We believe that the reason that samples with the highest mobility remain metallic even for resistance that is a few times larger than the quantum resistance h/e^2 can also be attributed to the presence of two distinct valleys. Because of the large increase of the contribution acting against localization, the transition is shifted to higher resistance. However, in MOSFET samples with moderate mobility, as well as in p-SiGe and p-GaAs, it occurs at values close to h/e^2 or less.

Next, in samples with a low mobility, where a description in terms of an effective single valley is relevant, the large value for $\gamma_2^* = 2.04$ makes it difficult for the nonmonotonicity to be observed since the initial values of γ_2 are, most probably, far away from 2.04. Then, to scale the amplitude γ_2 till the value γ_2^* will, for $\rho \ll 1$, demand exponentially small temperatures as the corrections depend on the temperature only logarithmically. On the other hand, for ρ near the critical region, where changes in the resistivity develop rapidly, the resistivity flows to such large values that the system instead of passing through the maximum, becomes insulating.

To summarize, we have argued that in the region not far from the transition it is not the large value of r_s that makes the physics in high mobility MOSFET samples so different from that in lower mobility samples, but the difference in their number of effective valleys. Note that in some samples the discussed anomalies have not been observed even for $r_s \approx 10$.

The strong magnetoresistance in a parallel magnetic field [3] can also be understood by the reduction of the

number of diffusion modes that contribute to the antilocalizing corrections [15,16]. Here, the Zeeman splitting induces a gap in the propagators of the diffusion modes that are made of states with different spin projections. As a result these modes will no longer contribute to the antilocalization corrections. In a very strong magnetic field when the electrons are completely polarized, the system becomes identical to one with no valleys with the original valleys acting as fictitious spin projections. The difference in the resistivity of two- and one-valley systems, which is large at low enough temperatures, will be recovered as the magnetic field is applied resulting in a very strong positive magnetoresistance.

In conclusion, we have demonstrated that in an ultraclean (100) Si-MOSFET the temperature behavior of the resistivity in the region $\mathbf{C}^*$ is well described by the RG analysis of the interplay of the *e-e* interaction and disorder when the electron band has two distinct valleys. For ρ not too large, the system of RG equations in the case of two valleys is an internally consistent theory (for all practical purposes), unlike that for a single valley where γ_2 diverges at $\eta \approx 1$ after the maximum of ρ is passed. This divergence points to some instability of a magnetic nature in the electron gas. This instability also occurs in the case of two valleys but at such low temperatures ($\eta \approx 10^4$) that it has no practical significance.

Finally, a few remarks concerning the electron gas in Si-MOSFETs.

The mobility at low densities is determined by the charged centers at the Si/SiO_2 interface. This may not be the case for the intervalley scattering as it involves a transfer of a large momentum $2Q_0 \propto 1/a$, where a is the lattice constant of Si. Since the width of the inversion layer $z_0 > a$, the intervalley scattering amplitudes involving Coulomb interactions with the charged centers will be proportional to a high power of the parameter $1/Q_0 z_0$, which is small. The imperfections on the interface, on the other hand, can be of the atomic scale and their matrix element will contain Fourier components of high momenta. We assume, therefore, that the rate $\hbar/\tau_\perp$ is sample dependent. At low densities the width of the 2D plane increases making the distance from the interface larger. As a result the intervalley scattering is suppressed and may become negligible as the density is lowered in the ultraclean samples.

Some information about the rate $\hbar/\tau_\perp$ can be obtained from the magnetoresistance measurements in a weak magnetic field perpendicular to the conduction plane. The results of these measurements [17], which yield a negative magnetoresistance, have been fitted with a standard expression containing the Digamma function, Ψ [18]. Depending on the rate $\hbar/\tau_\perp$, the theory predicts different values for the prefactor α in this expression: $\alpha = 1$ in the absence of the intervalley scattering and $\alpha = 0.5$ in the case of strong intervalley scattering [12]. The experimental situation for the sample used in Fig. 1 (but after some age degradation, however) remains uncertain. The optimal fit gives values for α between 0.6 and 0.8, with a tendency to be larger when the density decreases [17]. We consider the fact that α is noticeably larger than 0.5 as an indication that the intervalley scattering is not too strong in the system at low density.

We have ignored the valley splitting, as it is known that the splitting is small at low densities [10]. We also ignore the chiral splitting of the electron bands due to spin-orbit interactions as there are no reasons to expect considerable splitting in n-channel inversion layers.

We have benefited greatly from numerous discussions with V. M. Pudalov and we appreciate his contribution to our understanding of the experimental data. This work was supported by the Israel Science Foundation–Centers of Excellence Program, and by the German Ministry of Science (DIP). A. P. acknowledges the support of the Feinberg Graduate School of the WIS.

[1] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).
[2] S. V. Kravchenko *et al.,* Phys. Rev. B **50**, 8039 (1994); S. V. Kravchenko *et al.,* Phys. Rev. B **51**, 7038 (1995).
[3] D. Simonian *et al.,* Phys. Rev. Lett. **79**, 2304 (1997); Phys. Rev. B **57**, R9420 (1998); V. M. Pudalov *et al.,* JETP Lett. **65**, 932 (1997); V. T. Dolgopolov *et al.,* JETP Lett. **55**, 733 (1992).
[4] V. M. Pudalov *et al.,* Physica (Amsterdam) **3E**, 79 (1998).
[5] F. Stern, Phys. Rev. Lett. **44**, 1469 (1980).
[6] A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986).
[7] S. Das Sarma, Phys. Rev. B **33**, 5401 (1986).
[8] G. Zala, B. N. Narozhny, and I. L. Aleiner, cond-mat/0105406.
[9] B. L. Altshuler and A. G. Aronov, in *Electron-Electron Interactions in Disordered Systems,* edited by A. L. Efros and M. Pollack (North-Holland, Amsterdam, 1985).
[10] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[11] A. M. Finkel'stein, Sov. Phys. JETP **57**, 97 (1983); Z. Phys. B **56**, 189 (1984); Sov. Sci. Rev. A, Phys. Rev. **14**, 1 (1990).
[12] H. Fukuyama, J. Phys. Soc. Jpn. **50**, 3562 (1981).
[13] R. N. Bhatt and P. A. Lee, Solid State Commun. **48**, 755 (1983).
[14] C. Castellani *et al.,* Phys. Rev. B **30**, 1596 (1984).
[15] P. A. Lee and T. V. Ramakrishnan, Phys. Rev. B **26**, 4009 (1982).
[16] C. Castellani, C. Di Castro, and P. A. Lee, Phys. Rev. B **57**, R9381 (1998).
[17] G. Brunthaler *et al.,* Phys. Rev. Lett. **87**, 096802 (2001).
[18] B. L. Altshuler *et al.,* Phys. Rev. B **22**, 5142 (1980).