Scaling Theory of Two-Dimensional Metal-Insulator Transitions

V. Dobrosavljević,[1] Elihu Abrahams,[1,2] E. Miranda,[1] and Sudip Chakravarty[3]
[1]*National High Magnetic Field Laboratory, Florida State University 1800 E. Paul Dirac Drive, Tallahassee, Florida 32306*
[2]*Serin Physics Laboratory, Rutgers University, Piscataway, New Jersey 08855-0849*
[3]*Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, California 90095-1547*
(Received 10 April 1997)

We discuss the recently discovered two-dimensional metal-insulator transition in zero magnetic field in the light of the scaling theory of localization. We demonstrate that the observed symmetry relating conductivity and resistivity follows directly from the quantum critical behavior associated with such a transition. In addition, we show that very general scaling considerations imply that any disordered two-dimensional metal is a *perfect metal*, but most likely *not* a Fermi liquid. [S0031-9007(97)03673-9]

PACS numbers: 71.30.+h, 72.15.Rn

In an influential 1979 paper [1], simple scaling arguments were put forward which have shaped much of our thinking about metal-insulator transitions (MITs) in disordered systems. Focusing on noninteracting electrons, the authors demonstrated that in two dimensions (2D) even weak disorder is sufficient to localize the electrons, and no true metallic behavior is possible at $T = 0$. Ever since, it has been widely believed that two is the lower critical dimension for MITs in disordered systems, and this prompted much theoretical activity based on expansions around dimension $d = 2$. The scaling ideas were quickly extended to interacting systems, but despite many years of painstaking effort no satisfying theoretical picture has emerged for the interacting case. Nevertheless, the belief that all the states are localized in $d = 2$ has remained largely unquestioned.

Recently, a remarkable experiment [2] performed on a 2D electron gas in zero magnetic field suggested that the conventional picture might be incomplete. In this work, fairly convincing evidence was presented which supports the existence of a true MIT in 2D. Despite the beauty and elegance of the data, this work has met considerable resistance, largely due to its apparent conflict with the scaling theory of localization as well as its uniqueness.

The major assumption of [1] was based on an earlier idea of Thouless [3] about the length (scale) dependence of the conductance. In [1] it is asserted that the "beta function" $\beta(g) = d[\ln(g)]/d[\ln(L)]$ is a function of the conductance g itself but not an explicit function of the length scale L. Now $\beta(g)$ is known in the two limits of very large and very small disorder, and it is reasonable to assume that it is continuous (smooth) in between. From Ohm's law, $\beta(g) = (d - 2)$ for $g \to +\infty$, while for g small, one expects exponentially localized states, so that $\beta(g) \to -\infty$ in this limit. Since the metallic behavior is possible only for $\beta(g) > 0$, the form of $\beta(g)$ at large g is sufficient to determine the stability of the metallic phase. In particular, for noninteracting electrons, this has a form $\beta(g) = (d - 2) - 1/g + \cdots$ [4], indicating that the metallic phase is unstable in $d \leq 2$.

However, it is important to emphasize that the last step in this analysis is valid *only* for noninteracting electrons; the scaling theory is only complete in the absence of electron-electron interaction. In the following, we show that the existence of a 2D transition does not contradict any general scaling principles. However, we argue that within the scaling theory such a transition has a number of unusual features, many of which are already apparent in the existing data.

No fundamental principle requires that $\beta(g)$ be monotonic or, in fact, negative in 2D when g is large. Instead, if we assume that the leading behavior [5] at large g is

$$\beta(g) = (d - 2) + A/g^{\alpha} + \cdots, \qquad A > 0, \qquad (1)$$

then it is clear that the beta function for all $d \geq 2$ has to change sign at some finite $g = g_c$, leading, in particular, to a metal-insulator transition in 2D. Alternatively, the existence of a MIT in 2D *requires* that $\beta(g) > 0$ at g large, where the exponent α simply parametrizes how the system approaches the metallic limit as $g \to \infty$.

This idea is not new. Such a scenario is realized for noninteracting electrons in the presence of spin-orbit scattering [6]. In this case, $\alpha = 1$ and a MIT is expected in 2D. However, this behavior has never been taken too seriously in the context of real 2D systems because it is known that adding interactions to the spin-orbit universality class results in *reversing* the sign of the first quantum correction [4] and all the states again end up localized in 2D. A more complicated situation was found in the absence of spin-orbit and time-reversal breaking perturbations. There, an interaction driven enhancement of the conductivity *was* found at weak disorder, which was expected to overwhelm the localizing effects of coherent backscattering (weak localization) [7]. However, the analysis revealed that the effective interaction strength diverges upon scaling, making it difficult to determine what will actually happen at long scales or low temperatures [7]. Nevertheless, in that case, the relevant quantum corrections can, in fact, reduce the resistivity as the temperature is lowered, so

 © 1997 The American Physical Society

that the existence of a metallic state at low temperature is a possibility in the theory of Ref. [7].

At present, the above mentioned problems in the interaction case make it difficult to determine the form of the leading quantum corrections even at weak disorder. Rather than elaborating on this, we shall *assume* that interactions lead to a positive β at g large, so that the 2D metal exists. The first question that one can address is what is the low temperature form of the conductivity in the metallic phase. Within the scaling theory this is determined *only* by the form of $\beta(g)$ at g large. Assuming a form of Eq. (1), and integrating the beta function from ℓ (mean free path) to L (sample size), we find

$$g(L) = [\, g_0^{\alpha} + A\alpha \ln(L/\ell)]^{1/\alpha}. \tag{2}$$

If we consider an infinite sample at finite temperature, then $L_{\text{eff}} \sim T^{-1/z}$. Here z is the dynamical exponent associated with the *metallic* ($g = \infty$) fixed point. One expects $z = 2$ in a conventional diffusive regime, in the absence of dangerously irrelevant variables [8]. We thus find that at low temperatures in the metallic phase

$$g(T) \sim \ln^{1/\alpha}(T_0/T)\,. \tag{3}$$

In other words, the conductance will *diverge*, i.e., the resistance will *vanish* at $T = 0$ in the metallic phase, albeit in a weak, logarithmic fashion. This should be true at sufficiently low temperatures throughout the metallic phase, i.e., the temperature dependence should have a *universal* form, as in Fermi liquid theory. However, we emphasize that this is *not* a Fermi liquid, not only because of its unusual temperature dependence, but, more fundamentally, since we expect the nature of these electronic states to change completely if the interactions are turned off: they would localize.

Another experimentally relevant question is what is the temperature dependence of the conductance in the *quantum critical region* associated with the metal-insulator transition. Within scaling theory, this question can be answered very precisely as follows. In our formulation, the transition occurs at $g = g_c$, where the beta function changes sign, i.e., $\beta(g_c) = 0$. We recall that deep in the insulating phase the beta function is *logarithmic* in g [1]. It is thus plausible to introduce the quantity $t = \ln(g/g_c)$ as a natural scaling variable. To determine the critical behavior it is sufficient to consider the linear approximation to the beta function near its 0 at $t = 0$. The slope of $\beta(t)$ at $t = 0$ determines the correlation length exponent ν:

$$\beta(t) = \frac{dt}{d(\ln L)} \approx \frac{1}{\nu}\, t + O(t^2)\,. \tag{4}$$

By integrating this equation from ℓ to L, we find

$$t(L) = t_0 \left(\frac{L}{\ell}\right)^{1/\nu}, \tag{5}$$

where $t_0 = \ln(g_0/g_c)$ is determined by the starting value g_0 of the conductance at scale ℓ. In the critical region, where we start with a g_0 very close to g_c, $t_0 \approx (g_0 - g_c)/g_c \propto \delta n$, where n is the carrier concentration, and $\delta n = (n - n_c)/n_c$ measures the distance to the critical point. Then from Eq. (5), we have for the conductance

$$g(L) = g_c \exp[A\delta n (L/\ell)^{1/\nu}]\,. \tag{6}$$

At nonzero temperature, the length scale L is determined by the temperature through the dynamical exponent as $T \sim L^{-z}$ and we obtain the temperature dependence of the conductance

$$g(\delta n, T) = g_c \exp(\mathrm{sgn}(\delta n) A [T_0(\delta n)/T]^{1/\nu z})\,, \tag{7}$$

where we have defined a crossover temperature T_0 corresponding to the inverse correlation time as

$$T_0(\delta n) \sim |\delta n|^{\nu z}. \tag{8}$$

Here A is an unknown dimensionless constant of order 1. Let us define the scaled conductance as $g^*(\delta n, T) = g(\delta n, T)/g_c$. From Eq. (7), we then immediately find a striking symmetry relating the conductance on the metallic side ($\delta n > 0$) to the resistance on the insulating side ($\delta n < 0$) of the transition at $g^* = 1$ (i.e., at $\delta n = 0$):

$$g^*(\delta n, T) = 1/g^*(-\delta n, T)\,. \tag{9}$$

Remarkably, precisely this behavior is clearly seen in the experiments [2], as emphasized in Ref. [9].

In related work [10], a similar symmetry for the dependence on electric field was found for the same 2D system. If one uses the conventional electric field scaling hypothesis [11], the preceding arguments carry over with the electric field E replacing T and $1 + z$ replacing z, so that this symmetry too appears as a consequence of the scaling argument we have presented. In fact, the behavior holds whatever the exponent relating the electric field to a characteristic length.

In order to fully appreciate the significance of these findings, we should carefully qualify them and comment on the range of their validity.

(1) In [9], the authors point out a similarity of this symmetry with the one observed in the context of the quantum Hall liquid-to-insulator (QHI) transitions [12]. However, we emphasize that the theory described in [12] is fundamentally different from that of the present paper. The symmetry found in a 2D MIT is restricted to the *quantum critical region*, which is defined by the crossover temperature scale, i.e., it is expected to hold only for $T > T_0(\delta n)$. In fact, the experiments completely confirm this expectation, as the authors themselves point out [9]. A careful examination of the data in [2] reveals that the temperature at which a departure from symmetry is observed is $T \approx T_0(\delta n)$, with T_0 shown in the inset of Fig. 3 of [9]. The same feature is clearly seen in also the earlier data of [2], where in Fig. 4 the resistance is plotted as a function of the reduced temperature $\tau = T/T_0(\delta n)$. Here the obvious symmetry of the resistance and the conductance on the respective metallic and insulating sides of the MIT is seen and it is violated below

$T \approx T_0(\delta n)$, as expected. The duality observed in QHI is thought to be a charge-flux duality that holds for the full Chern-Simons Lagrangian, ignoring the roles of irrelevant operators, cutoffs, or disorder. If this is so, it has little resemblance to the symmetry in MIT. On the other hand, it may be that the data in QHI transition can be understood from the present perspective, where $(B - B_c) \propto (1/\nu - 1/\nu_c)$ is the control parameter that tunes the system through a quantum critical point; here B is the magnetic field and ν is the filling fraction.

(2) We emphasize that the above results were derived by *linearizing* the beta function around the transition point $g = g_c$. The symmetry is then simply a consequence of the smoothness of the beta function in the critical region. However, without further knowledge about the form of the beta function, the result is strictly speaking valid only for $\delta n \ll 1$, i.e. it only defines the form of the leading high-temperature correction in the quantum critical regime. In other words, to be consistent, one should keep only the leading term of the expansion of the exponential of Eq. (7):

$$g(\delta n, T) \approx g_c[1 + A\delta n/T^{(1/\nu z)}]. \qquad (10)$$

This expression is generally expected to be valid only for $|g(T) - g_c|/g_c \ll 1$. On the other hand, if we *assume* that the linearized expression for the beta function [Eq. (4)] is a good approximation over an appreciable conductance range $g_{\min} < g_c < g_{\max}$, then the full exponential behavior is valid in that range. The experimental data reveal that the symmetry is found in a much broader range than allowed by the form of Eq. (10). In fact, the data clearly display an exponential temperature dependence as in Eq. (7), which is particularly striking in the metallic phase, where the resistance is found to drop almost by an order of magnitude within the "symmetry" regime. The deviations from leading behavior are more clearly seen in Fig. (2b) of [9], where the conductance and the resistance are plotted at a fixed temperature, for a range of carrier concentrations. The two curves show perfect mirror symmetry which holds over an extended range of concentrations, where the conductance shows a dramatic exponential dependence on concentration, in agreement with Eq. (7). This should be contrasted with the leading correction given by Eq. (10), where the density dependence is *linear*, as expected only for $\delta n \ll 1$.

(3) We conclude that the experimental results of [2,9,10] go beyond just confirming the expectations based on general scaling arguments in the quantum critical region. They also provide striking evidence about the form of the beta function in the critical region. In particular, they indicate that in a wide range of conductances $1/4 < g/g_c < 4$ the beta function is well-approximated by the linear expression in $t = \ln(g/g_c)$ as in Eq. (4). How can we rationalize this finding? Deep in the insulating regime, ($g \ll g_c$) the beta function is *exactly* given by $\beta(g) \sim \ln(g)$. The above experimental result can thus be interpreted as evidence that the same slow logarithmic form of the beta function persists beyond the insulating limit well into the critical regime. Further support for this idea can be found in the experimental findings of Hsu and Valles [13] on ultrathin films. While no MIT was found, their data yield a beta function which is close to linear in $\ln(g)$ ("strong localization") all the way to the standard crossover conductance $g \simeq e^2/2h$. If it is generally true, this feature could be used as a basis of approximate calculations of the critical exponents. In contrast to the well-known $2 + \varepsilon$ expansion, here one would try to obtain the form of the beta function in the critical region by an expansion around the *strong disorder* limit.

(4) There are physical reasons which support our conjecture that the beta function must be a logarithmic function close to g_c, if g_c happens to fall in the regime of strong disorder, which the experiments discussed here indicate. The assumption of the scaling theory is that the conductance g is the only important parameter at large distances and low energies. That is, the distribution of conductances of a hypercube of linear dimension L is so sharply peaked that the most probable value is the only value, hence the mean value. This assumption can be justified in the limit of weak disorder, where the sample-to-sample fluctuations of the physical properties are weak, but not when the disorder is strong and the distribution of the conductances is broad. At strong disorder, it is more plausible to assume, if we still want to describe the problem with a single scaling variable, that the distribution is sharply peaked when described in terms of $\log g$ [14], or more generally in terms of a function $\ln[\phi(g)]$, where $\phi(g)$ is a smooth function of g that tends to $b_{-1}/g + b_0 + b_1 g + \cdots$, as $g \longrightarrow 0$, but tends to $a_0 + a_1/g + a_2/g^2 \cdots$, as $g \longrightarrow \infty$. At strong disorder, it is natural to believe that Green's function between two points can be computed in a hopping parameter expansion [15]. It is plausible that the impurity averaged square of Green's function for p hops on an effective lattice of spacing a larger than the mean free path is

$$\overline{|G^2(p)|} = \left[\frac{\phi(g_c)}{\phi(g)}\right]^p. \qquad (11)$$

If $g > g_c$, the perturbation theory does not converge as $p \longrightarrow \infty$; this signals a failure of the localization assumption. When $g < g_c$, this is not so, and we can define a localization length. However, for a finite system of linear dimension L, one can meaningfully use this equation for all g because p is restricted to be less than (L/a). This allows us to derive the beta function as follows. From Eq. (11), the physical localization length is

$$\xi = \frac{2a}{\ln[\phi(g)/\phi(g_c)]}. \qquad (12)$$

If we keep the physical ξ fixed as we vary the lattice spacing, that is, set $d\xi/da = 0$, we get

$$\beta = \frac{\partial \ln \phi(g)}{\partial \ln a} = \ln\left[\frac{\phi(g)}{\phi(g_c)}\right]. \qquad (13)$$

To check that this is correct, consider the limit $g, g_c \longrightarrow 0$. We find that

$$\frac{\partial \ln g}{\partial \ln a} = \left(1 + \frac{b_0}{b_{-1}} g + \cdots\right) \ln\left(\frac{g}{g_c}\right), \qquad (14)$$

which is precisely the locator expansion [115]. The difficult task is to be able to compute $\phi(g)$ from a microscopic model. Note, however, that local electron-electron interactions should not pose any special problems in the locator expansion.

(5) The symmetry associated with the quantum critical region is specific to 2D systems. In particular, standard arguments suggest that in general dimension, the conductivity as a function of reduced concentration and temperature should assume the scaling form

$$\sigma(\delta n, T) \sim T^{(d-2)/z} f(\delta n / T^{1/\nu z}). \qquad (15)$$

Here $f(x)$ is a universal scaling function such that $f(0) = 1$, in the absence of dangerously irrelevant variables. As a result of the extra temperature prefactor present when $d > 2$, the conductivity will vanish at the transition ($\delta n = 0$, $T = 0$), ruining the reflection symmetry of the quantum critical region. This fact would make it very difficult, if not impossible, to extract the information about the details of the beta function from the temperature dependence in the quantum critical region when $d > 2$.

(6) The samples showing the 2D MIT are distinguished from those in previous work, as pointed out by the authors of [2], by the fact that the density n is so low that the Coulomb interaction $U \propto \sqrt{n}$ is immense, almost an order of magnitude greater than the Fermi energy $E_F \propto n$. Furthermore, for small n, high mobility μ is required to reach the critical conductance $g_c = n_c e \mu \simeq e^2/3h$. This also leads to the conclusion that $k_F l \lesssim 1$ for the samples showing MIT. Thus it is misleading to assume that because the mobilities are high they are weakly disordered. These conditions both lead to the conclusion that the one-electron picture of [1] cannot be expected to be valid and may explain why the MIT is not observed in some samples.

In summary, we have discussed the recently discovered 2D metal-insulator transition in zero magnetic field in the light of the scaling scenario for localization. We have shown that very general scaling considerations imply that any disordered 2D metal is a *perfect metal*, but most likely *not* a Fermi liquid. In addition, we have demonstrated that the observed symmetry relating conductivity and resistivity follows directly from the quantum critical behavior associated with such a transition. Furthermore, the fact that this symmetry is found over an extended range of conductances provides important information about the form of the beta function in the critical region.

The support of NSF Grants No. DMR 96-32294 (E. A.) and No. DMR 95-31575 (S. C.) is acknowledged. E. A. thanks the NHMFL for its hospitality. V. D. was supported by the NHMFL/FSU and the Alfred P. Sloan Foundation. E. M. was supported by the NHMFL/FSU. We thank P. W. Anderson, R. Bhatt, N. Bonesteel, L. Engel, S. Kravchenko, C. Nayak, D. Popovic, M. Sarachik, D. Simonian, and X-G. Wen for useful discussions.

[1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
[2] S. V. Kravchenko, W. E. Mason, G. E. Bowker, J. E. Furneaux, V. M. Pudalov, and M. D'Iorio, Phys. Rev. B **51**, 7038 (1995).
[3] D. J. Thouless, Phys. Rep. **13C**, 93 (1974).
[4] For a review of localization, see P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[5] If the beta function is to be analytic, then α is restricted to integer values.
[6] S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980); T. Ando, Phys. Rev. B **40**, 5325 (1989); S. N. Evangelou and T. Ziman, J. Phys. C **20**, L235 (1987).
[7] A. M. Finkel'stein, Zh. Eksp. Teor. Fiz. **84**, 168 (1983) [Sov. Phys. JETP **57**, 97 (1983)]; [Zh. Eksp. Teor. Fiz. Pis'ma Red. **40**, 63 (1984)] [Sov. Phys. JETP Lett. **40**, 796 (1984)]; [Z. Phys. B **56**, 189 (1984)].
[8] Note that there can be another z associated with the critical region of the MIT.
[9] D. Simonian, S. V. Kravchenko, and M. P. Sarachik, Phys. Rev. B **55**, R13 421 (1997).
[10] S. V. Kravchenko, D. Simonian, M. P. Sarachik, W. Mason, and J. E. Furneaux, Phys. Rev. Lett. **77**, 4938 (1996).
[11] A. Yazdany and A. Kapitulnik, Phys. Rev. Lett. **74**, 3037 (1995).
[12] D. Shahar, D. C. Tsui, M. Shayegan, E. Shimshoni, and S. Sondhi, Science **274**, 589 (1996).
[13] Shih-Ying Hsu and J. M. Valles, Jr., Phys. Rev. Lett. **74**, 2331 (1995).
[14] For a discussion of this point, see P. W. Anderson, D. J. Thouless, E. Abrahams, and D. S. Fisher, Phys. Rev. B **22**, 3519 (1980).
[15] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).