

Localized – delocalized electron quantum phase transitions

V F Gantmakher, V T Dolgoplov

DOI: 10.1070/PU2008v051n01ABEH006471

Contents

1. Introduction	3
1.1 Thermal and quantum fluctuations	
2. Quantum phase transitions	5
2.1. Definition of a quantum phase transition; 2.2 Parallels and differences between classical and quantum phase transitions; 2.3 Critical region of a quantum phase transition	
3. Flow diagrams for metal – insulator transitions	9
4. Three-dimensional electron gas	10
5. Two-dimensional electron gas	13
5.1 Noninteracting electron gas; 5.2 Spin – orbit interaction; 5.3 Interacting electron gas	
6. Quantum transitions between the different states of a Hall liquid	18
7. Conclusion	21
References	22

Abstract. Metal – insulator transitions and transitions between different quantum Hall liquids are used to describe the physical ideas forming the basis of quantum phase transitions and the methods of application of theoretical results in processing experimental data. The following two theoretical schemes are discussed and compared: the general theory of quantum phase transitions, which has been developed according to the theory of thermodynamic phase transitions and relies on the concept of a partition function, and a theory that is based on a scaling hypothesis and the renormalization-group concept borrowed from quantum electrodynamics, with the results formulated in terms of flow diagrams.

1. Introduction

Quantum phase transitions that occur in an electron gas and result in localization represent the mainstream in the study of electrons in solids; it is directed toward low temperatures and interactions. We think that a dangerous gap between theory and experiment has appeared in this field of science. The related theories are so complex that experimentalists, as a rule, use ready recipes and formulas of these theories without knowing (and, hence, without controlling) all the initial assumptions and limitations.

In this review, we try to close this gap. Our review does not contain a sequential description of the mathematical techniques involved in formulating hypotheses and theories. Instead, we outline the physical ideas, concepts, and assumptions that are often omitted in theoretical works and reviews, especially at the stage of a developed theory. This specific feature is thought to make our review interesting and useful for experimentalists.

On the other hand, we did not tend to accumulate, classify, and estimate the huge experimental data obtained in this field. We address experiment to demonstrate the use of a theory for its interpretation and the related difficulties and problems. This specific feature is thought to make our review interesting and useful for theorists.

We first qualitatively discuss the ideas that constitute the basis of the theory of quantum phase transitions. We then consider how the general theoretical scheme can describe the well-known data on metal – insulator transitions and transitions between different quantum Hall liquids.

1.1 Thermal and quantum fluctuations

Before discussing quantum phase transitions in essence, we briefly recall some facts from statistical physics. We consider a macroscopic system consisting of a huge number of particles. This system is almost isolated and behaves mainly as a closed system; however, it can exchange particles and energy with a larger system, which serves as a reservoir. In other words, our system is a subsystem of this reservoir.

We first analyze the classical subsystem at a finite temperature T . According to classical statistics formulas, the probability p_i of the residence of the subsystem in a state with energy ε_i is

$$p_i = \frac{\exp(-\varepsilon_i/T)}{Z}, \quad (1)$$

V F Gantmakher, V T Dolgoplov Institute of Solid State Physics, Russian Academy of Sciences, ul. Institutskaya 2, 142432 Chernogolovka, Moscow region, Russian Federation
Tel. (7-496) 522 29 42, (7-496) 522 29 46. Fax (7-496) 524 97 01
E-mail: gantm@issp.ac.ru, dolgop@issp.ac.ru

Received 8 October 2007

Uspekhi Fizicheskikh Nauk 178 (1) 3–24 (2008)

DOI: 10.3367/UFNr.0178.200801a.0003

Translated by K V Shakhlevich; edited by A M Semikhatov

where the coefficient $1/Z$ is determined by the normalization condition $\sum p_i = 1$,

$$Z = \sum_i \exp\left(-\frac{\varepsilon_i}{T}\right) \quad (2)$$

(hereafter, temperature is given in energy units). Z is called the partition function and plays a key role in the description of the thermodynamic properties of classical objects. That the subsystem energy obeys distribution (1) rather than being fixed means the presence of classical thermal fluctuations.

When quantum mechanics appeared, the basis of classical thermodynamics and its principal formulas were revised. The classical expressions were found to have a limited field of application. Quantum statistics requires an advanced apparatus of quantum mechanics. Before addressing this apparatus, we assume that the subsystem is completely closed and does not interact with the larger system. As a result, we can use the standard methods of quantum mechanics and describe the subsystem by the Schrödinger steady-state equation. The resulting set of steady-state energies ε_i and the corresponding wave functions $\varphi_i(q)$ of the subsystem can be considered subsystem attributes in the zeroth approximation. These energies ε_i enter Eqn (2). The set of $\varphi_i(q)$ functions is convenient because it is related to the subsystem under study and is complete; therefore, it can be conveniently used for series expansions.

The presence of a huge number of particles in the subsystem implies that it has a very dense energy distribution of quantum levels. Due to a weak interaction with the reservoir, the subsystem is in the so-called mixed state, in which no measurement or a set of measurements can lead to unambiguously predicted results. By the definition of a mixed state, the repeated measurements of any physical quantity give a near-average result that differs from the previous one. The scatter of the measured physical quantities is interpreted as a result of fluctuations.

Because the subsystem interacts with the environment, it cannot be described by a wave function; therefore, the wave function is substituted by a density matrix,

$$\rho(q', q) = \int \psi^*(q', X) \psi(q, X) dX, \quad (3)$$

where q corresponds to the set of subsystem coordinates, X are the remaining coordinates of the system, and $\psi(q, X)$ is the wave function of the closed system.

The mean values of any physical quantity $\langle s \rangle$ are now calculated using the density matrix

$$\langle s \rangle = \frac{1}{A} \int \left\{ \hat{s}[\rho(q', q)] \right\}_{q'=q} dq, \quad A = \int \rho(q, q) dq \quad (4)$$

rather than a wave function. In Eqn (4), we apply the operator \hat{s} , which acts on functions of the variable q , to the density matrix $\rho(q', q)$; we then set $q' = q$ and integrate. The operator that formally enters the expression for the normalization factor A is identically equal to unity; therefore, the condition $q' = q$ is naturally taken into account in the integrand.

According to general rule (4), the mean of coordinate $\langle q \rangle$, for example, is given by

$$\langle q \rangle = \frac{1}{A} \iint \psi^*(q, X) q \psi(q, X) dq dX = \frac{\int q \rho(q, q) dq}{\int \rho(q, q) dq}. \quad (5)$$

The physical meaning of the density matrix can be clarified if we write it in an explicit form for the subsystem in statistical equilibrium at a finite temperature T ,

$$\rho(q', q) = \sum_i \varphi_i^*(q') \varphi_i(q) \exp\left(-\frac{\varepsilon_i}{T}\right). \quad (6)$$

For the subsystem in statistical equilibrium, it follows from Eqns (5) and (6) that

$$\begin{aligned} \langle q \rangle &= \frac{1}{A} \sum_i \int \varphi_i^*(q) \varphi_i(q) \exp\left(-\frac{\varepsilon_i}{T}\right) q dq \\ &= \frac{1}{A} \sum_i \langle q \rangle_i \exp\left(-\frac{\varepsilon_i}{T}\right), \end{aligned} \quad (7)$$

$$\langle q \rangle_i = \int \varphi_i^*(q) \varphi_i(q) q dq.$$

The averaging with the density matrix performed in Eqn (7) accounts for both the probabilistic description in the form of $\langle q \rangle_i$ in quantum mechanics and incomplete information on the system (statistical averaging).

We use the $\varphi_i(q)$ functions and write the function $\rho(q', q)$ in the matrix form

$$\rho(q', q) = \|\rho_{ij}\| = \left\| \int \varphi_j^*(q) \rho(q', q) \varphi_i(q) dq \right\|. \quad (8)$$

In Eqn (4), we then have

$$\langle s \rangle = \frac{\sum_{ij} s_{ij} \rho_{ji}}{\sum_i \rho_{ii}} = \sum_{ij} s_{ij} w_{ji}, \quad w_{ij} = \frac{\rho_{ij}}{\sum_i \rho_{ii}}. \quad (9)$$

The w_{ij} matrix constructed from the set of $\varphi_i(q)$ functions is called the statistical matrix and, in essence, is the normalized density matrix.

We replace \hat{s} with the energy operator \hat{H} in Eqn (4) and obtain

$$\langle \varepsilon \rangle = \sum_i w_{ii} \varepsilon_i. \quad (10)$$

This means that the probability p_i of detecting the energy ε_i in the subsystem is equal to the diagonal element of the matrix w_{ii} ,

$$p_i = w_{ii}. \quad (11)$$

Equation (11) is the quantum analog of Eqn (1).

The statistical matrix has a number of universal properties. By definition, it is normalized:

$$\sum_i w_{ii} = 1. \quad (12)$$

As follows from Eqn (6), the statistical matrix is diagonal in statistical equilibrium. Its diagonal elements are functions of only the energy ε_i of the corresponding subsystem state. As long as at least one w_{ii} with $i \neq 0$ has a nonzero value, the subsystem state remains mixed, and any measured quantity undergoes fluctuations.

The macroscopic system is in a mixed state mainly due to a finite temperature. A comparison of Eqns (1) and (11) demonstrates that in the high-temperature limit, the statistical matrix obeys the Gibbs distribution

$$w_{ii} = \frac{\exp(-\varepsilon_i/T)}{Z} \quad (13)$$

with a good accuracy.

Because the integral of the squared wave function $\varphi_i(q)$ over the entire coordinate space is

$$\int \varphi_i^*(q) \varphi_i(q) dq \equiv 1,$$

the partition function Z can be represented as the sum of the diagonal matrix elements of the operator $\exp(-\hat{H}/T)$ over the complete set of eigenfunctions φ_i :

$$\begin{aligned} Z &= \sum_i \int \varphi_i^*(q) \varphi_i(q) \exp\left(-\frac{\varepsilon_i}{T}\right) dq \\ &= \sum_i \left\langle \varphi_i \left| \exp\left(-\frac{\hat{H}}{T}\right) \right| \varphi_i \right\rangle. \end{aligned} \quad (14)$$

Here, we use that φ_i are eigenfunctions of the Hamiltonian \hat{H} to obtain

$$\exp\left(-\frac{\hat{H}}{T}\right) \varphi_i = \exp\left(-\frac{\varepsilon_i}{T}\right) \varphi_i.$$

A comparison of Eqns (4)–(7) with Eqn (13) gives another form of the partition function,

$$Z = \int \rho(q, q) dq \equiv \text{Sp} [\rho(q', q)]. \quad (15)$$

If Z is regarded as a normalization factor, Eqns (4) and (9) demonstrate that representation (15) is valid in both the classical limit and general case.

Because the subsystem is open, it interacts with the environment. Even if this interaction is very weak, the states of a macroscopic system become mixed in a certain small energy range due to the extremely high density of the system energy levels. Therefore, even at zero temperature, the unclosed subsystem is in a mixed state. This means that as the temperature decreases, each w_{ii} ($i \neq 0$) does not tend to zero exponentially but approaches a temperature-independent constant, which depends on ε_i and the sizes and properties of the subsystem. Thermal fluctuations are replaced by quantum fluctuations. This statement is illustrated by Fig. 1a, which qualitatively shows the temperature dependence of the probability p_i of detecting the energy ε_i in the subsystem. The dashed lines separate the regions of predominantly quantum and predominantly thermal fluctuations. The limiting value $p_i(0)$ is specified by the energy ε_i , sizes, and properties of the subsystem.

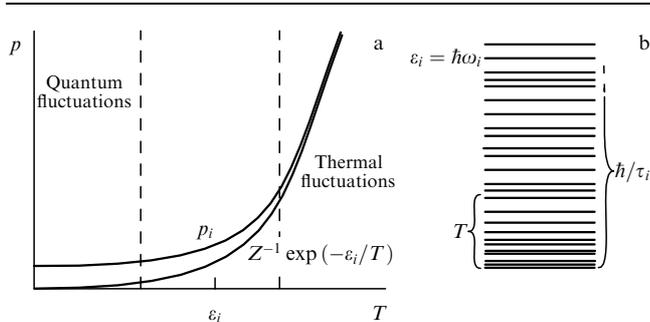


Figure 1. (a) Low-temperature behavior of the diagonal elements w_{ii} of the statistical matrix corresponding to low energies. (b) Simultaneous excitation of thermal and quantum fluctuations.

In this subsystem, $p_i(0)$ obviously decreases with an increase in ε_i . We assume that this dependence is a power-law function, $p_i(0) \propto 1/\varepsilon_i^\alpha$, and obtain the lower estimate of the exponent α . The energy of thermal fluctuations is taken from the reservoir. A finite, not exponentially small, probability of detecting the subsystem in a state $\varepsilon_i \gg T$ means a violation of the energy conservation law. This violation may occur, but must be constrained by the uncertainty relation

$$\varepsilon_i \tau_i \sim \hbar, \quad (16)$$

where τ_i is the lifetime of the state ε_i . Therefore, even if transitions to all ε_i states were equiprobable, $p_i(0)$ would decrease according to the law $p_i(0) \propto \varepsilon_i^{-1}$, as follows from Eqn (16). As a result, we have

$$\alpha \geq 1. \quad (17)$$

It is essential that the energy range in which quantum fluctuations occur is temperature independent, which follows from uncertainty relation (16).

The relative role of thermal and quantum fluctuations is illustrated in Fig. 1b, which shows the excitation spectrum of the subsystem in one of the phases, i.e., far from the phase transition point $x = x_c$. The brace in the bottom left part of Fig. 1b indicates the temperature-related scale. Only modes with frequencies $\omega_i \lesssim T/\hbar$ are classically excited modes. The range of thermal modes is bounded from above by temperature, but they have large occupation numbers. Modes with $\omega_i > T/\hbar$ are mainly excited due to quantum processes and have small occupation numbers. However, as the temperature decreases, the role of quantum excitations increases.

2. Quantum phase transitions

2.1 Definition of a quantum phase transition

Quantum phase transitions are phase transitions that can occur at the absolute zero $T = 0$ and consist of a change in the *ground state* of a system in the case where a certain control parameter x takes a critical value x_c . (By the ground state, we mean the lowest possible mixed state.) The control parameter can be a magnetic field or an electron concentration. Quantum phase transitions belong to the class of continuous transitions at which none of the physical functions of state has a discontinuity at the transition point.

A quantum fluctuation is the only cause that can change the ground state of the system at zero temperature; the phase transitions under study are therefore called quantum phase transitions. As in the case of thermodynamic transitions, the concept of a correlation length (ξ), which has the meaning of the average quantum-fluctuation size, is introduced into the theory of quantum phase transitions. At the absolute zero temperature, ξ is determined only by the deviation δx of the control parameter from the critical value. The dependence of ξ on δx is assumed to be a power-law function,

$$\xi \propto |x - x_c|^{-\nu}. \quad (18)$$

Real experiments are always performed at finite temperatures, where thermal fluctuations exist in addition to quantum fluctuations. The theoretical goal is to predict the manifestation of the phase transition that occurs only at zero

temperature in the properties of the system at a finite temperature.

The theory of quantum phase transitions (see, e.g., book [1] or reviews [2, 3]) is analogous to the theory of thermodynamic phase transitions. In the (x, T) plane (i.e., in the plane control parameter–temperature), the quantum transition point can be the endpoint of the line of thermodynamic transitions, $x_c(T) \rightarrow x_c(0)$ as $T \rightarrow 0$. For example, this behavior is characteristic of magnetic transitions with the magnetic field used as a control parameter. Quantum phase transitions of another type also exist; they are represented by an isolated point x_c on the abscissa axis of the (x, T) plane. The metal–insulator transition is an example of such a transition. In this review, we restrict ourselves to the case of isolated-point transitions.

2.2 Parallels and differences between classical and quantum phase transitions

As noted above, quantum phase transitions belong to the class of continuous transitions; there are no two coexisting competing phases and, hence, no stationary boundary between them. This class also includes thermodynamic second-order phase transitions. At a thermodynamic phase transition point, the system transforms into another phase as a whole as a result of thermal fluctuations. Fluctuations exist on either side of the transition, and their characteristic size ξ is called the correlation length. As the transition is approached from either side, ξ diverges [5, 6].

It is natural to expect a similar situation to occur in the vicinity of a quantum transition with the participation of quantum fluctuations. An analogy between classical and quantum phase transitions does exist, and it is rather unexpected. *The behavior of a quantum system in the vicinity of the transition point at a finite temperature in a d -dimensional space is analogous to the behavior of a classical system in a space with dimension $\mathcal{D} > d$.* This statement requires extensive explanations, which should include an algorithm for the introduction of such an imaginary system and the determination of its dimension.

First, we have to clearly distinguish between the dimension of the geometric space in which the system is located and the dimension of the space of generalized coordinates, which depends on the number of particles \mathcal{N} in the system. If ϱ is the density of particles in space, we have $\mathcal{N} = \iint \varrho d^d X$, where the dimension d determines the multiplicity of the integral. Usually, a d -dimensional space is assumed to be infinite in all directions, but the range of one or several X_i coordinates can be limited.

Second, we note a similarity between the operator $\exp(-\hat{H}/kT)$, which was used to write Eqn (14), and the operator \hat{S} that describes the time evolution of a closed quantum system in accordance with the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi, \quad \hat{S} = \exp\left(-\frac{i}{\hbar} \hat{H}t\right). \quad (19)$$

This similarity becomes obvious if we change the variables as

$$\frac{it}{\hbar} = \frac{1}{T}. \quad (20)$$

With substitution (20), we can interpret the matrix elements in Eqn (14) differently. The element $\langle i | \exp(-\hat{H}/T) | i \rangle$ can be regarded as the amplitude of the probability that, starting from a state $\langle i |$, the subsystem

evolves under the action of \hat{S} and returns to the initial state $|i\rangle$ within an imaginary time \tilde{t} , which is equal to $-i\hbar/T$. The imaginary time \tilde{t} is often called the Matsubara time. For definiteness, we assume that the number of steps is fixed and equal to $N + 1$, $N \gg 1$, and that, in time $i\hbar/T$, the subsystem has passed through N virtual states and occupied each state for a time $(\delta\tilde{t})_j$; as a result, we have

$$\sum_{j=0}^N (\delta\tilde{t})_j = \frac{i\hbar}{T}. \quad (21)$$

The amplitude of ‘the probability of returning to the initial state’ is the sum of the amplitudes of the probabilities of returning for all possible trajectories in the space of states. We consider a set of trajectories consisting of N steps. The operator entering each matrix element in Eqn (14) is represented as

$$\exp\left(-\frac{\hat{H}}{T}\right) = \exp\left(\frac{i}{\hbar} \hat{H} \sum_{j=0}^N (\delta\tilde{t})_j\right). \quad (22)$$

We take a complete set of wave functions $|m_j\rangle$ of an arbitrary operator that does not commute with the Hamiltonian and assume that the trajectories are realized in these states. Then, instead of Eqn (14), we obtain

$$\begin{aligned} Z = & \sum_i \sum_{m_1, m_2, \dots, m_N} \left\langle i \left| \exp\left(-\frac{i\hat{H}(\delta\tilde{t})_1}{\hbar}\right) \right| m_1 \right\rangle \\ & \times \left\langle m_1 \left| \exp\left(-\frac{i\hat{H}(\delta\tilde{t})_2}{\hbar}\right) \right| m_2 \right\rangle \\ & \dots \left\langle m_N \left| \exp\left(-\frac{i\hat{H}(\delta\tilde{t})_N}{\hbar}\right) \right| i \right\rangle. \end{aligned} \quad (23)$$

The product of the matrix elements in the summand corresponds to a chain of consecutive virtual transitions. The summation over all the $\{m_j\}$ combinations means that all possible closed chains of N links are taken into account. The modification of the expression for Z —the replacement of Eqn (14) with Eqn (23)—means that we take the quantum properties of the system into account by adding virtual transitions to real ones.

The introduction of virtual transitions is illustrated in Fig. 2, where the abscissa axis stands for the initial d -dimensional space $\{^d X\}$ and all possible classical states of the system are located along this axis in the order of increasing energy. All other horizontal lines are copies of this space. These copies form a set of N elements in the segment $[0, \hbar/T]$.

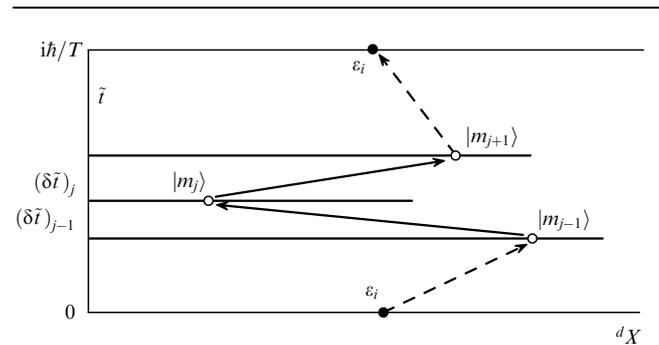


Figure 2. A set of quantum statistical d -dimensional systems located in a one-dimensional strip of width \hbar/T and virtual transitions between them.

The black dot on the abscissa axis represents the initial state of the system with an energy ε_i . The second black dot on the upper horizontal line corresponds to the ‘final’ state of the system with the same energy ε_i at the maximum value of the imaginary time $|\tilde{t}| = \hbar/T$. The arrows indicate the chain of virtual transitions through the set of virtual states $|m_j\rangle$ designated by white dots. These are mixed states without a definite energy; therefore, they are not eigenstates of the operator \hat{H} . The virtual transitions reflecting the quantum properties of the statistical system under study are shown by arrows. The chain of virtual transitions shown in Fig. 2 corresponds to one term in the sum in (23). The summation over all $\{m_j\}$ combinations means that we account for all possible chains between the fixed initial and final ε_i points.

Now, the problem is to construct a classical system whose partition function is represented by Eqn (23). In the original Eqn (14), the summation over i meant the summation over all states realized in a d -dimensional system. The number of terms in the sum is now increased, and our real system does not have such a large number of different states. Nevertheless, we can construct an imaginary higher-dimensional classical system using the scheme shown in Fig. 2.

We add an extra imaginary-time axis to the d axes of the original space. In the graphical terms of Fig. 2, this means that an ordinate axis is added to the abscissa axis. It is seen from Eqn (21) that the coordinate along this new axis changes in the range

$$0 \leq |\tilde{t}| \leq \frac{\hbar}{T}. \quad (24)$$

The width of strip (24) in Fig. 2, in which the image point representing a virtual state of the quantum system is located, increases with decreasing the temperature. At $T = 0$, the strip transforms into a half-plane; an increase in the temperature, in contrast, narrows the strip and decreases its contribution to the statistical properties of the quantum system.

We copy the d -dimensional classical system N times by placing replicas along the imaginary time \tilde{t} axis at a distance $(\delta\tilde{t})_j$ from each other. The states of the replicas in this ensemble can be different. Let these states be m_j . We fix a certain ‘initial’ state of the ‘lower’ subsystem ε_i . The energy of the formed classical \mathcal{D} -dimensional system is equal to the sum of the energies of all layers, with allowance for the interaction between them. Each $\varepsilon_{i,\{m\}}$ corresponds to a certain energy ε_i of the original d -dimensional classical system and a certain chain of states entering Eqn (23). Therefore, the number of terms in the sum $\sum_{\{m\}} s_{i,m}$ in Eqn (23) is equal to the number of different $\varepsilon_{i,\{m\}}$ energies. We note that each $s_{i,m}$ term in sum (23) has a form typical of partition functions, since the product of the matrix elements in this term eventually reduces to the product of $\exp(-i\varepsilon_j(\delta\tilde{t})_j/\hbar)$. The only difference is that for the classical system, each term in Z is a real positive number, but in Eqn (23), a real positive number is represented by the sum $s_i = \sum_{\{m\}} s_{i,m}$ over $\{m\}$ of all $s_{i,m}$ products. With this difference neglected, we can consider Eqn (23) the partition function Z of the imaginary \mathcal{D} -dimensional classical system.

Of course, in the general form, the resulting mathematical construction is absolutely impractical. However, as is customary in deriving scaling relations, it is important to reveal some significant properties of sum (23) rather than to calculate it. As is shown in Section 2.3, one of these properties is anisotropy, i.e., nonequivalence of the axes of the \mathcal{D} -dimensional space.

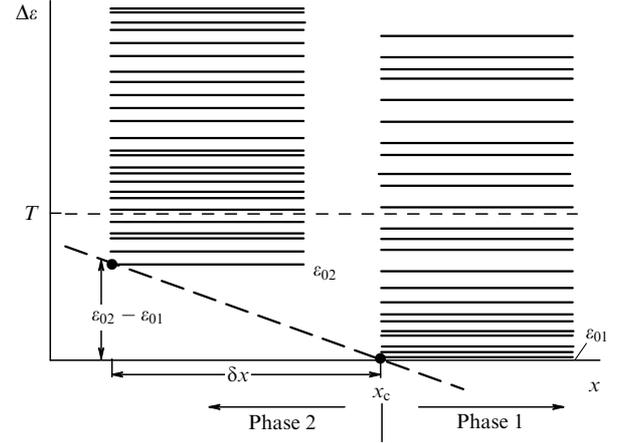


Figure 3. The difference in the lowest energies of two phases $\Delta\varepsilon = \varepsilon_{02} - \varepsilon_{01}$ versus a control parameter (dashed line). At a control parameter $x = x_c - \delta x$, phase 1 is subjected to fluctuation excitation inside equilibrium phase 2. Because $|\Delta\varepsilon| < T$, thermal fluctuations play a key role in this excitation.

At a thermodynamic phase transition point, the partition function Z has a special feature. The sensitivity of function (14) to the presence of a phase transition is caused by the fact that in approaching the transition, the energy range T determining significant terms in sum (14) contains not only φ_i levels from the set corresponding to the equilibrium phase but also φ'_i levels from the nonequilibrium phase (Fig. 3). This allows fluctuation transitions between $|i\rangle$ and $|i'\rangle$ levels from different sets. Additional possibilities appear after Eqn (14) is replaced with Eqn (23). If the temperature is low ($|\Delta\varepsilon| > T$), the fluctuation-induced appearance of another phase is also possible, but due to quantum fluctuations.

2.3 Critical region of a quantum phase transition

The region in which all physical quantities depend only on the correlation length ξ always exists in the phase plane near a classical phase transition. This region is called the critical or scaling region. Near a quantum phase transition at $T = 0$, there also exists a control parameter δx range in which physical quantities are expressed through the length ξ determined by Eqn (18). In Section 4, we show this behavior by the example of a metal–insulator transition in a three-dimensional system of noninteracting electrons and determine the boundaries of this region. At a finite temperature $T \neq 0$, however, the scaling region of a quantum phase transition is more complex.

The space $\{^d X, \tilde{t}\}$ has dimension $d + 1$ because of the extra imaginary time axis. We introduce correlation lengths in the space $\{^d X, \tilde{t}\}$. We retain the traditional notation for the correlation length (ξ) in an ordinary d -dimensional subspace and let ξ_φ denote the correlation length along the additional axis. The subscript φ is a reminder that ξ_φ is related to the quantum aspect of the problem and to the specific features of wave functions. The dimension of ξ_φ is \hbar/T rather than length; that is, it is measured in seconds rather than centimeters.

As $x \rightarrow x_c$ and $T \rightarrow 0$, both correlation lengths diverge at the transition. According to the theory of continuous thermodynamic transitions, the divergence is described by power-law functions; but the exponents of the two correlation

lengths can be different in general. This is usually written as

$$\xi \propto \delta x^{-\nu}, \quad \delta x = |x - x_c|, \quad (25)$$

$$\xi_\varphi \propto \xi^z. \quad (26)$$

The exponents ν and z are called critical indices; z is called the dynamic critical index. Both names and designations originate from the theory of thermodynamic processes. In particular, the dynamic critical index in the theory of thermodynamic processes enters the relation between the lifetime and size of thermal fluctuations, $\tau \propto \xi^z$. In the quantum problem at $T = 0$, the situation is similar: the length ξ characterizes spatial correlations, i.e., the characteristic size of quantum fluctuations, and the time ξ_φ characterizes time correlations.

Formula (26) immediately fixes the dimension of the space in which the imaginary classical system should be placed. A real length must be the coordinate along all axes of this space; that is, length axes should be made from the \tilde{r} axis. As follows from dimensional considerations and Eqn (26), the length equivalent to the ‘correlation pseudolength’ ξ_φ is proportional to $\xi_\varphi^{1/z}$,

$$\zeta \propto \xi_\varphi^{1/z}, \quad (27)$$

and therefore the volume element in the space is

$$(d\xi)^d d\xi_\varphi \propto (d\xi)^d (d\xi)^z, \quad (28)$$

and a z -dimensional subspace with the conventional ‘spatial’ coordinates appears instead of the one-dimensional imaginary time axis. Hence, we have

$$\mathcal{D} = d + z. \quad (29)$$

It is convenient to discuss the consequences of Eqns (25) and (26) using the (x, T) plane (Fig. 4). At $T = 0$, the control parameter x on the abscissa axis of the (x, T) plane only affects one independent correlation length (ξ), and ξ_φ (and the length ζ) can be formally obtained from ξ using Eqn (26). All the physical quantities can be expressed only in terms of x_c in a certain segment $|x_1, x_2|$ that contains the point x_c .

We assume that $T \neq 0$ and move in the (x, T) plane along a horizontal line $T \neq 0$ (Fig. 4, line aa) in the $x \rightarrow x_c$ direction. At a certain value of δx , the length ξ_φ , which changes in accordance with Eqns (25) and (26), reaches its

maximum value \hbar/T [which is determined by inequality (24)]. At lower values of δx , Eqn (26) does not hold on line aa , because ξ increases according to Eqn (25) and ξ_φ remains equal to its maximum value \hbar/T . The parameters ξ and ξ_φ become *mutually independent*. The region of this independence is called the critical region. If we now move toward the transition inside the critical region (e.g., along line bx_c), both parameters still diverge at the transition. However, the divergence of one of them is controlled by δx , $\xi \propto \delta x^{-\nu}$, and the divergence of the other is controlled by temperature, $\xi_\varphi = \hbar/T$.

The name of the region derives from the fact that for a quantum transition, the critical region is considered to be the region where quantum fluctuations play an essential role in mixing the two-phase states. As can be seen from Fig. 3, this occurs only under the condition

$$|\varepsilon_{02} - \varepsilon_{01}| > T. \quad (30)$$

As the temperature decreases, the range of the control parameter δx where condition (30) is satisfied narrows. As a result, the shape of the critical region is unusual: it widens in going from the transition.

Inside the critical region, the quantity $\xi_\varphi = \hbar/T$ can be associated with a real length. The quantity

$$L_\varphi \propto \left(\frac{\hbar}{T}\right)^{1/z} \quad (31)$$

has the required properties: it is equivalent to the parameter $\xi_\varphi = \hbar/T$ and, according to Eqn (26), has the dimension of length. The physical meaning of L_φ can be understood from the following considerations. Because of a finite temperature, the quantum problem acquires a characteristic energy T that separates classical and quantum fluctuations (see Fig. 1b). Quantum fluctuations are fluctuations with energies $\hbar\omega_\varphi > T$. The spatial size of these fluctuations is $l_\varphi \propto 1/\omega_\varphi$ and is bounded from above because the frequency $\omega_\varphi > T/\hbar$ is bounded from below. The length L_φ is the upper boundary of the size of quantum fluctuations. Therefore, L_φ is often called the dephasing length, i.e., the length beginning from which coherence in a system of electrons is broken.

The appearance of two independent parameters in the critical neighborhood of a quantum phase transition is caused by inequality (24). A scaling description in the critical neighborhood of a quantum phase transition is therefore called a finite-size scaling. An imaginary thermodynamic system is thought to exist in a hyperstrip in a \mathcal{D} -dimensional space with d variables ranging from 0 to ∞ and z variables ranging from 0 to $(\hbar/T)^{1/z}$.

Inside the critical region, $L_\varphi < \xi$, and, along its boundary,

$$L_\varphi = \xi. \quad (32)$$

The lengths L_φ and ξ are determined by Eqns (25) and (26) up to a multiplicative constant; however, Eqns (25) and (26) rigidly fix a power-law relation between these variables. Therefore, the equation for the critical region boundaries has the form

$$T = C(\delta x)^{\nu z}, \quad (33)$$

where the constant C can be different on either side of the transition in general.

The constructed critical region is based on Eqns (25) and (26). However, the use of Eqn (25) at all temperatures actually

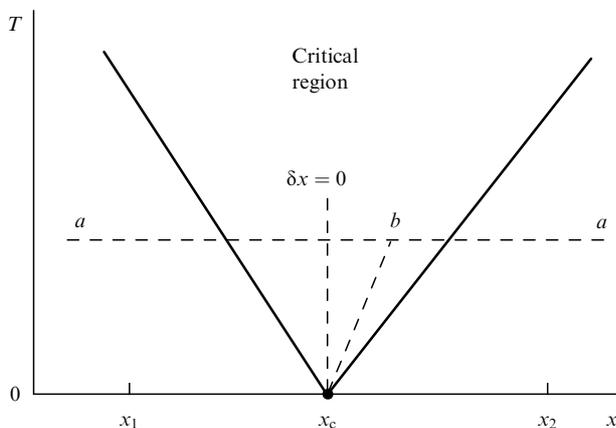


Figure 4. Shape of the critical region of a quantum phase transition depicted by an isolated point in the (x, T) plane.

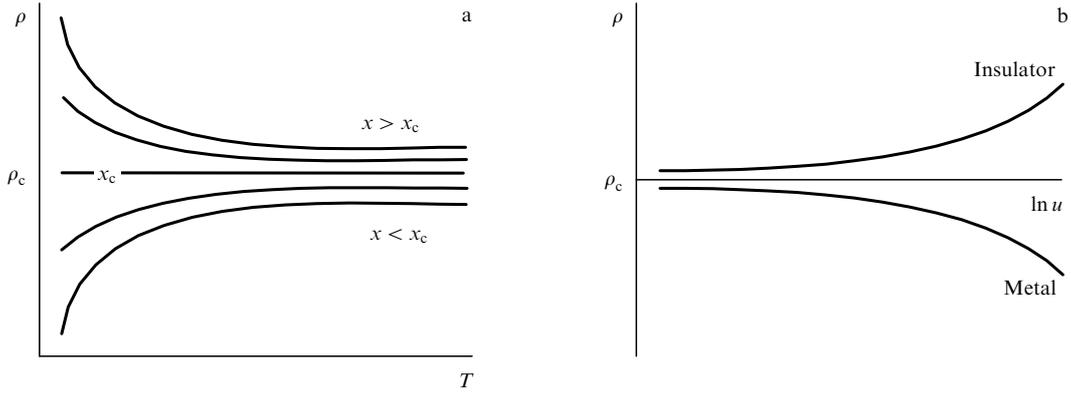


Figure 5. (a) Temperature dependences of the resistivity $\rho(T)$ at various values of x with the horizontal separatrix $x = x_c$. (b) Reduction of all the $\rho(x, T)$ curves to two curves (34) of the dependence of the resistivity ρ on scaling variable (35).

assumes that the length ξ depends only on δx and is temperature independent. This simplified scheme was usually applied to the transitions considered in this review. As is shown below, this simplification is not always valid.

Because the critical region has two independent parameters, the scaling expressions for physical quantities become more complex. We only present and discuss expressions for the electric conductivity and resistivity,

$$\sigma = \xi^{2-d} F\left(\frac{L_\varphi}{\xi}\right), \quad \rho = \xi^{d-2} F_1\left(\frac{L_\varphi}{\xi}\right), \quad F_1(u) = \frac{1}{F(u)}, \quad (34)$$

where $F(u)$ is an arbitrary function. The exponent in the first factor in Eqn (34) is specified by how the length enters the expressions for conductivity at various dimensions d . As can be seen from Fig. 4, this factor is determined by the x component of the distance from the transition. The L_φ/ξ ratio is dimensionless; therefore, scaling rules can be used to regard L_φ/ξ as an argument of an arbitrary function. This relation depends on both T (i.e., the y component of the distance from an image point to the transition in the phase plane) and the distance from this point to the critical-region boundary along the x axis.

As an argument, we can also take any power of the L_φ/ξ ratio; as a result, we write the argument in various forms, such as

$$\frac{L_\varphi}{\xi}, \quad \frac{\hbar/T}{\xi_\varphi}, \quad \frac{\hbar/T}{\xi^z}, \quad \frac{\hbar(\delta x)^{zv}}{T}, \quad \text{or} \quad \frac{\delta x}{(T/\hbar)^{1/zv}}. \quad (35)$$

These various forms elucidate various aspects of the physical meaning of this ratio. In the last two forms, the argument is dimensional and the dimensionality of the control parameter is not specified a priori. These forms show how temperature enters the argument of the scaling function.

To conclude this section, as an example, we present an implication of Eqn (34) for the resistivity of a two-dimensional system. If a quantum phase transition induced by electron localization occurs in a two-dimensional system and manifests itself in the transport properties of this system, we can write

$$\rho(x = x_c) \equiv \frac{1}{F(0)}, \quad \text{i.e.} \quad \rho(x_c, T) = \text{const} = \rho_c, \quad \text{at } d = 2. \quad (36)$$

This implication (a horizontal separatrix in the set of temperature dependences at various values of x) is schematically shown in Fig. 5a. We note that it is valid only under assumption (25), which indicates that the correlation length ξ is temperature independent. At an arbitrary value of x , we obtain the curve $\rho(u) = \rho(\delta x/T^{1/zv})$ with a finite slope at the point $x = x_c$ (cf. Fig. 13).

Usually, with δx replaced by the modulus $|\delta x|$, the $\rho(u)$ curve is represented in the form of two branches as a function of $\ln u$ (Fig. 5b). The values of zv are then chosen such that these curves fit all the experimental points obtained at various temperatures.

3. Flow diagrams for metal–insulator transitions

So far, we have not specified the type of a quantum phase transition. Hereafter, we speak about transitions related to a change in the electron localization, which, in turn, is caused by the degree of disorder in the system. The general theory of quantum phase transitions initially assumes that a control parameter affects the interparticle interaction and that a disorder is not more than a perturbative factor in the initial scheme of quantum phase transitions. Therefore, the applicability of this scheme to metal–insulator transitions, in which the degree of disorder is the main factor and the interaction is a secondary factor, is not obvious a priori.

The fundamental difference between an insulator and a metal is that the electron states at the Fermi level are localized in an insulator and delocalized in a metal. If an insulator is transformed into a metal due to a change in a certain parameter, the properties of the wave functions at the Fermi level change. The main physical property that is radically different in materials of these two types is the conductivity, i.e., the possibility of carrying an electric current at an arbitrarily weak electric field. This gives a ‘yes–no’ type signature: the conductivity is either zero ($\sigma = 0$) or arbitrarily small but nonzero. However, at a finite temperature ($T \neq 0$), an insulator also carries a current owing to hopping conductivity. Therefore, the definition of an insulator given above is only related to the temperature $T = 0$, and the concept of a metal–insulator transition makes sense only at $T = 0$.

That the conductivity σ is not a function of the state (because it is realized only under nonequilibrium conditions) is an additional reason to doubt the applicability of the general theory. However, Thouless [4] noted that the

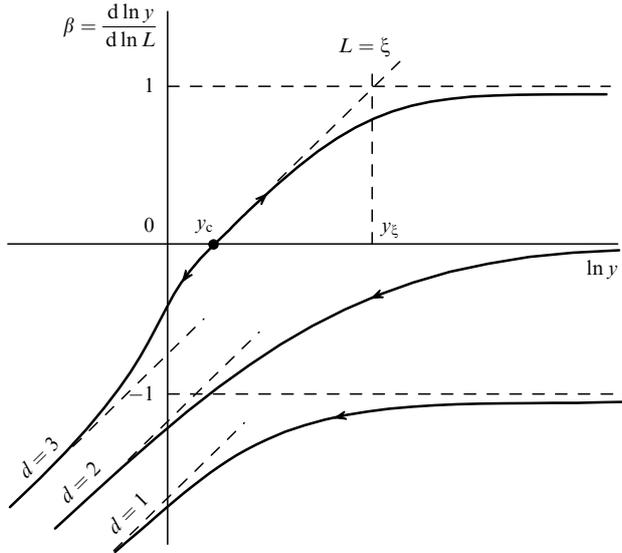


Figure 6. Universal $\beta(\ln y)$ functions for various dimensionalities [7].

transport properties can be used to characterize the decrease in the wave function of an electron placed at the center of an L^d cube in a d -dimensional space toward the edges of the cube. Therefore, in analyzing the behavior of conductance or conductivity near the transition, we actually follow the evolution of wave functions.

Thus, we have every reason to apply the general theory of quantum phase transitions to metal–insulator transitions. However, the first successful version of a theoretical description of metal–insulator quantum transitions [7] was based on the renormalization group theory [8] borrowed from quantum field theory. The essence of its conclusions, related to systems of noninteracting electrons in a random potential, is illustrated in Fig. 6, which shows the logarithmic derivative of the conductance of a sample with respect to the size L ,

$$\beta = \frac{d \ln y}{d \ln L} = \frac{L}{y} \frac{dy}{dL},$$

at the temperature $T = 0$ as a function of the conductance y ,

$$\frac{d \ln y}{d \ln L} = \beta_d(\ln y). \quad (37)$$

The $\beta_d(\ln y)$ curves describe the universal laws of a change in the conductance of the system when its size changes. The behavior of the system of noninteracting electrons in any material is described by a part of the curve for the corresponding dimension. The interpretation and method of using these curves are described in detail in review [9] and book [10].

For metal–insulator quantum phase transitions, the theory in [7] plays the role of an existence theorem. The shape of the $\beta_d(\ln y)$ curves and their position in the (y, β) plane indicate that these transitions are absent in one- and two-dimensional systems of noninteracting electrons and that such a transition can occur in three-dimensional systems, with the conductivity of the material changing continuously during this transition. This means that sufficiently large one- and two-dimensional samples are always insulators at absolute zero, whereas a three-dimensional material can be

either an insulator or a conductor. Below, we discuss this feature in detail. Here, we note that the $\beta = \beta_d(\ln y)$ curves are called differential flow diagrams or Gell-Mann–Low curves.

On the one hand, the theory in [1–3] is more specialized than that in [7] in some respects, because it only describes the vicinity of a quantum transition. On the other hand, the theory in [1–3] is more universal, because its applicability is not limited by either interaction or a strong magnetic field. Therefore, it is of interest to compare the conclusions of these theories for the systems to which they can be applied.

4. Three-dimensional electron gas

The possibility of a phase transition in a three-dimensional material follows from the fact that the $\beta_3(\ln y)$ curve intersects the abscissa axis $\beta = 0$. A certain point on the $\beta_3(\ln y)$ curve corresponds to the transport properties of a sample of size L made from a certain material (Fig. 7a). If this point is located in the lower half-plane ($\beta < 0$), the material is an insulator, and the image point shifts to the left as L increases; as a result, the conductance of the sample becomes exponentially small. If this point is located in the upper half-plane ($\beta > 0$), the image point shifts to the right; as a result, the material is a metal, and its conductance increases with the size L .

The curves in Figs 6 and 7a are not suitable for a direct comparison with experimental data because of a special choice of their coordinate axes. This disadvantage can be partly corrected by integrating the equation

$$\frac{d \ln y}{\beta_3(\ln y)} = d \ln L \quad (38)$$

for three-dimensional (3D) systems. Because the left-hand side of differential equation (38) becomes infinite at the point $y = y_c$, the curves corresponding to different integration constants decompose into two families: one of them corresponds to an insulator region and the other to a metallic region. To show this in the simplest way, we restrict ourselves to the immediate vicinity of point y_c , in which the $\beta_3(\ln y)$ curve can be approximated by a straight line,

$$\frac{d \ln y}{d \ln L} = s \ln \frac{y}{y_c}, \quad (39)$$

where s is the slope of the line with respect to the abscissa axis. Correspondingly, Eqn (38) can be replaced by linear equation (39). The general solution of linear differential equation (39) is given by

$$\ln \frac{y}{y_c} = \left(\frac{L}{\lambda} \right)^s \ln \frac{y_\lambda}{y_c}, \quad (40)$$

where λ plays the role of the initial condition that fixes the initial point in the $\beta_3(\ln y)$ curve; for example, it can be the length k_F^{-1} determined by the electron concentration. On the metallic side, k_F^{-1} is equal to the minimum free path length l_{\min} . The conductance y_λ corresponds to the point λ .

Figure 7b shows particular solutions of Eqn (39). If $y_\lambda > y_c$, we have $\ln y_\lambda/y_c > 0$, the solution is located in the right quadrant, and the conductance y increases with L (metal). If $y_\lambda < y_c$, we have $\ln y_\lambda/y_c < 0$, the solution is located in the left quadrant, and the conductance y decreases as L increases (insulator). The set of curves in Fig. 7b is called

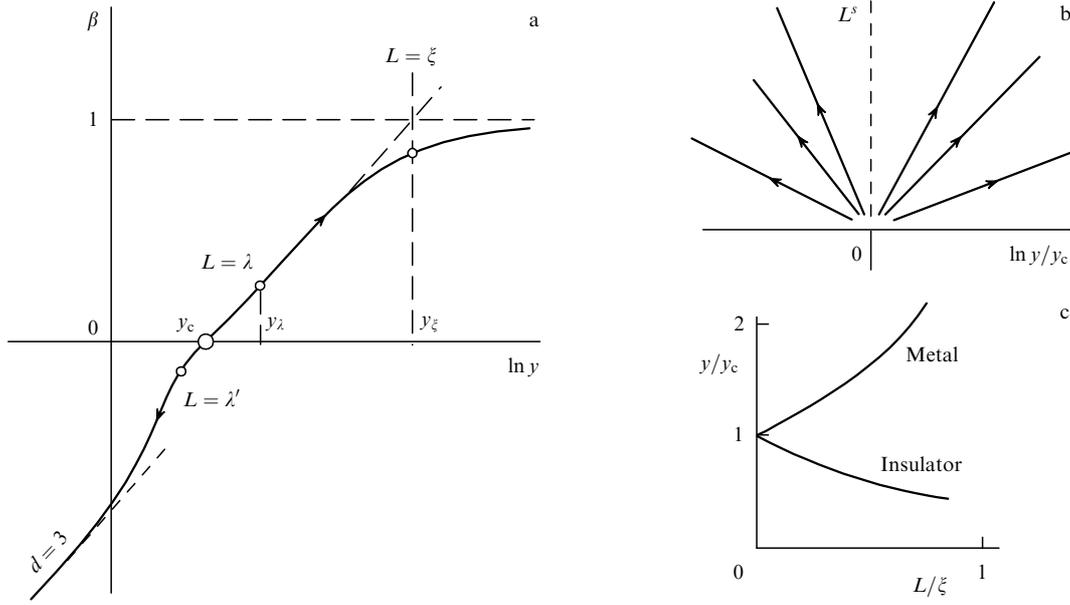


Figure 7. Noninteracting 3D electron gas: (a) differential scaling diagram from [7] (see also Fig. 6); the intersection of the $\beta_3(\ln y)$ curve with the abscissa axis $\beta = 0$ signifies the existence of a quantum phase transition; (b) flow diagram for a noninteracting 3D electron gas in the vicinity of the transition point; and (c) two universal curves of the dependence of the conductance of a 3D system on its size L obtained by scaling.

a flow diagram. The arrows in this diagram show the direction of the image point motion when the system size increases. The flow lines corresponding to the particular solutions of Eqn (39) fill both upper quadrants of the $(\ln y_i/y_c, L^s)$ plane. The boundary between them, the $\ln y_i/y_c = 0$ axis, is called the separatrix; in Fig. 7b, it is indicated by a dashed line.

We extend straight line (39) in Fig. 7a to its intersection with the asymptote $\beta = 1$; as a result, we approximate the $\beta_3(y)$ curve in the upper half-plane ($\beta > 0$) by a broken line consisting of segments of two straight lines. The size L corresponding to the intersection point is called the correlation length ξ . From Eqn (40), we have

$$\xi = \lambda \left(s \ln \frac{y_i}{y_c} \right)^{-1/s}. \quad (41)$$

With Eqn (41), we can rewrite general solution (40) as

$$\ln \frac{y}{y_c} = \left(\frac{L}{\xi} \right)^s, \quad (42)$$

normalize the size L by the length ξ (which is specific for each material), and reduce each of the two families in Fig. 7b to one scaling curve (Fig. 7c).

We make several remarks.

First, Eqns (39)–(42) involve a parameter s , which is unknown a priori. However, in the currently relevant case of a noninteracting 3D gas, this parameter is either equal to unity or very close to it [9]. Because s enters all formulas as a factor or exponent, this parameter can be dropped.

Second, the flow lines in Fig. 7b are straight only because we integrated the Gell-Mann – Low equation only in a small vicinity of the transition point. If we integrated the $\beta_3(\ln y)$ function over a wide range of its argument $\ln y$ using some model representation, we would obviously obtain curved flow lines.

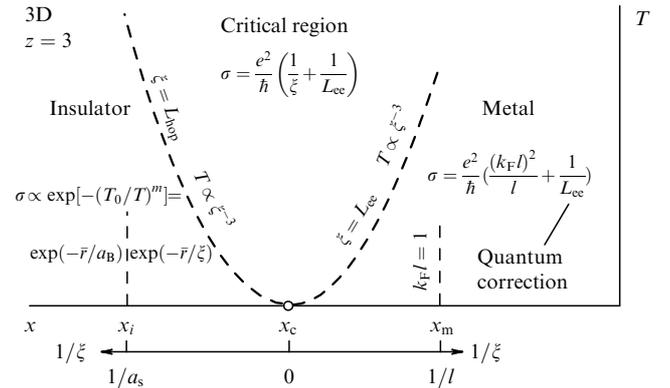


Figure 8. Vicinity of the metal – insulator transition in a noninteracting 3D electron gas in the (x, T) phase plane.

The third remark concerns the role of temperature. The parameter that specifies system motion along the flow lines indicated by arrows in Fig. 7b is the *size*. If the temperature is taken to be zero (as was assumed until now), this parameter is the sample size L . However, we can initially suppose that the system is large, as this is done in the theory of quantum phase transitions [1 – 3]. In this case, the limiting size is considered to be the dephasing length L_ϕ , i.e., the size at which quantum coherence in an electron system is realized. The image point then moves along flow lines in the direction indicated by arrows *as the temperature decreases*. Because L_ϕ depends on the temperature and becomes infinite as $T \rightarrow 0$, this procedure introduces the temperature into the set of physical quantities related to the flow diagram. However, we do not study this relation; based on the ‘existence theorem’ of a quantum transition in a 3D space, we turn to the (x, T) phase plane (Fig. 8) to construct the critical region (cf. Fig. 4).

Let small values of the control parameter x correspond to metallic states. Then, at small x on the abscissa axis, the

Drude formula $\sigma = \sigma_3 = ne^2 l / \hbar k_F$ holds, and, at small x and finite T , the quantum correction

$$\sigma = \sigma_3 + \frac{e^2}{\hbar} L_{ee}^{-1}, \quad L_{ee} = \sqrt{\frac{\hbar D}{T}}, \quad (43)$$

where D is the diffusion coefficient, is added to σ_3 . The diffusion length in Eqn (43) is taken to be the length L_{ee} determined by the Aronov–Altshuler effect [11], i.e., by the electron–electron interaction. In other words, we assume that dephasing is caused by internal processes that occur in the electron gas in the time

$$\tau_{ee} = \frac{\hbar}{T} \quad (44)$$

without external impacts, such as the electron–phonon interaction (see analogous discussion of the physical meaning of L_φ in review [2, p. 324]).

At any control parameter x , another special point x_m (the Mott limit) exists on the right of the transition point x_c . At this special point, we have

$$\sigma(x_m, T=0) = \sigma_{\text{Mott}} = \frac{e^2}{\hbar} k_F = \frac{e^2}{\hbar} \frac{1}{l}. \quad (45)$$

In the range between x_c and x_m , the Drude formula is invalid: describing conductivity with this formula implies introducing a free path length l that is shorter than the de Broglie wavelength. In this range at $T=0$, σ is expressed through the correlation length ξ ; as we know, this agrees with the general scheme of the theory of quantum phase transitions. As a result, the conductivity along the abscissa axis is expressed as

$$\sigma(T=0) = \left(\frac{e^2}{\hbar}\right) \times \begin{cases} 0, & x \geq x_c, \\ \frac{1}{\xi}, & x_c \geq x \geq x_m, \\ \frac{(k_F l)^2}{l}, & x \leq x_m. \end{cases} \quad (46)$$

To match the last two expressions at $x = x_m$, it suffices to set $\xi(x_m) = l$.

Thus, to describe conductivity in the metallic region, we introduced two parameters, ξ and L_{ee} , which have the dimension of length and the properties required for the critical region (they are mutually independent and diverge at the point x_c). Following [12], we construct the following interpolation function in the critical region:

$$\sigma = \frac{e^2}{\hbar} \left(\frac{1}{\xi} + \frac{1}{L_{ee}} \right) = \frac{e^2}{\hbar} \frac{1}{\xi} F\left(\frac{L_{ee}}{\xi}\right), \quad F(u) = 1 + \frac{1}{u}. \quad (47)$$

Expression (47) matches Eqn (43) on the straight line $x = x_m$ and gives the correct values of conductivity in the $x_c < x \leq x_m$ segment at $T=0$. The second form of the function demonstrates that function (47) satisfies general relation (34).

We now move from right to left along the line $T = \text{const}$ (Fig. 4, line *aa*). As long as the quantum correction is relatively small, electron diffusion occurs via scattering by impurities; i.e., it is controlled by the first term in Eqn (43). Therefore, the diffusion coefficient D in L_{ee} is temperature independent. But as we enter the critical region, σ_3 transforms into $(e^2/\hbar)\xi^{-1}$ and begins to decrease rapidly. Under these

conditions, D ceases to be a constant: diffusion is likely to be caused by the electromagnetic-field fluctuations that determine L_{ee} ; as a result, this diffusion becomes temperature dependent. We can then write a set of equations for the $\sigma(T)$ and $D(T)$ functions with the Einstein relation

$$\begin{cases} \sigma = \frac{e^2}{\hbar} \left(\frac{1}{\xi} + \sqrt{\frac{T}{\hbar D}} \right), \\ \sigma = e^2 g_F D, \end{cases} \quad (48)$$

where g_F is the density of states at the Fermi level, used as a second equation.

We eliminate D from Eqn (48), use that $1/\xi \ll 1/L_{ee}$ near the transition, and obtain the temperature dependence of the conductivity on the right-hand side of the critical region [12, 13],

$$\sigma(T) = \frac{e^2}{\hbar} \left(\frac{1}{\xi} + (T g_F)^{1/3} \right) \equiv \alpha + \beta T^{1/3}. \quad (49)$$

This means that inside the critical region, L_{ee} is not determined by Eqn (43) but is given by

$$L_{ee} = (T g_F)^{-1/3}. \quad (50)$$

Exactly at the transition, we have $\xi = \infty$ and $\alpha = 0$.

Temperature dependence (49) was repeatedly detected in experiments performed on a variety of materials [14–17]. Figure 9 shows two examples in which the control parameter is represented by the electron concentration and magnetic field.

Using condition (32) and Eqn (49), we can write the relation

$$T \propto (g_F \xi^3)^{-1} \quad (51)$$

for the right boundary of the critical region. If the relation $\xi \propto (\delta x)^{-1}$ holds (i.e., if $\nu = 1$, as is assumed in Fig. 8), the boundary of the critical region is represented by a cubic parabola $T \propto (\delta x)^3$. In the general case, we have $\nu \neq 1$ and

$$T \propto \frac{(\delta x)^{3\nu}}{g_F}. \quad (52)$$

A comparison with Eqn (33) demonstrates that in the metal–insulator transition in a 3D system of noninteracting electrons, the dynamic critical index is

$$z = 3. \quad (53)$$

Below curve (52) in the $x_c > x > x_m$ segment, the conductivity is also described by Eqn (47); however, L_{ee} enters Eqn (47) in the initial form (43) with the diffusion coefficient $D = \text{const}$. Therefore, the temperature dependence of the conductivity in this region must have the form $\sigma(T) = \alpha + \beta T^{1/2}$.

Up to this point, we have discussed the right-hand side of the phase diagram. The left-hand side of the phase diagram, i.e., the insulator region (where hopping conductivity is realized), also has two characteristic lengths. First, there is the decay length ξ of the localized-state wave functions, $\psi \propto \exp(-r/\xi)$. Far from the transition, ξ decreases to the Bohr radius $a_B = \kappa \hbar^2 / (m^* e^2)$ (where κ is the dielectric

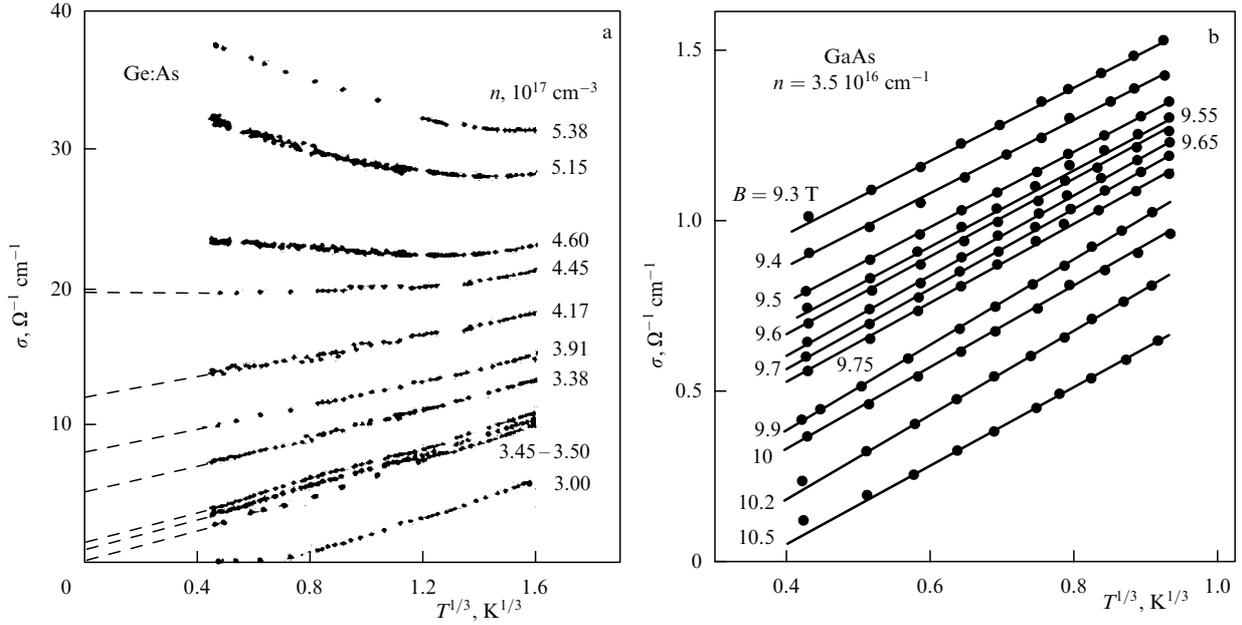


Figure 9. (a) Temperature dependences of the conductivity of Ge:As samples with various levels of doping in the region of a metal–insulator transition (borrowed from [18]); the critical concentration is determined by the extrapolation of the experimental data to $T = 0$. (b) Temperature dependences of the conductivity of a GaAs sample in various magnetic fields in the region of a metal–insulator transition (borrowed from [16]); the critical field determined by extrapolation is 9.78 T.

constant and m^* is the effective electron mass); at the transition, it diverges because the electrons become delocalized. Second, there is the average hopping distance \bar{r} . If the hopping conductivity is described by the Mott law, the average hopping distance is [10]

$$\bar{r} = \left(\frac{\xi}{g_F T} \right)^{1/4}. \quad (54)$$

The lengths \bar{r} and ξ cannot be used as two independent lengths in the critical region because they are connected by relation (54). However, using L_{ee} determined by Eqn (50), we can rewrite Eqn (54) as

$$\bar{r} = (\xi L_{ee}^3)^{1/4}. \quad (55)$$

We take into account that expression (50) for L_{ee} does not contain the kinetic characteristics of the electron gas and consider L_{ee} the dephasing length over the entire critical region, including its left-hand side, above the $x_i > x > x_c$ segment on the abscissa axis. Equation (52) then determines both the right and left boundaries of the critical region. Equations (54) and (55) demonstrate that $\bar{r} = \xi = L_{ee}$ at the left boundary.

That L_{ee} can be considered the dephasing length over the entire critical region is supported experimentally: as can be seen from Fig. 9, the $\sigma(T)$ temperature dependences straighten in terms of the $(T^{1/3}, \sigma)$ axes, not only in the right part of the critical region but also for those values of the control parameter x at which the system is in the left part of the critical region (e.g., at the electron concentration $n = 3 \times 10^{17} \text{ cm}^{-3}$ in Fig. 9a). However, the free term α in Eqn (49) becomes negative. This means that we should either assume that the correlation length ξ is negative in the insulator region or (which is formally preferred but is essentially the same) should replace interpolation formula (49) on the left-hand side of the critical region by the

formula

$$\sigma(T) = \frac{e^2}{\hbar} \left(-\frac{1}{\xi} + \frac{1}{L_{ee}} \right). \quad (56)$$

As follows from Eqns (55) and (56), we have $\sigma = 0$ along the left boundary of the critical region, which means that the conductivity along this boundary is determined up to an exponentially small hopping conductivity.

The hopping conduction mechanism is still operative below the lower boundary of the critical region above the $x_i > x > x_c$ segment, and the wave function decay length is $\xi \gg a_B$ rather than a_B [19]. Therefore, a_B does not enter the expression for the hopping conductivity, and σ is expressed in terms of the correlation length ξ (see Fig. 8):

$$\sigma \propto \exp \left(-\frac{\bar{r}}{\xi} \right).$$

The (x, T) phase diagram in Fig. 8 accumulates the results of the long-term experimental studies of the low-temperature transport properties of conducting systems, namely, the quantum corrections to metallic conductivity, the evolution of electron spectra during metal–insulator transitions, the temperature dependences of conductivity in the vicinity of the transitions, and hopping conductivity. In essence, this phase diagram was plotted irrespective of the theory of quantum phase transitions [1–3] in order to reveal the compatibility of all experimental data. Nevertheless, the considerations given above demonstrate that this diagram is absolutely adequate for the concepts following from this theory.

5. Two-dimensional electron gas

5.1 Noninteracting electron gas

We now pass to 2D systems. According to the theory in [7], a 2D ($d = 2$) system of noninteracting electrons is always an

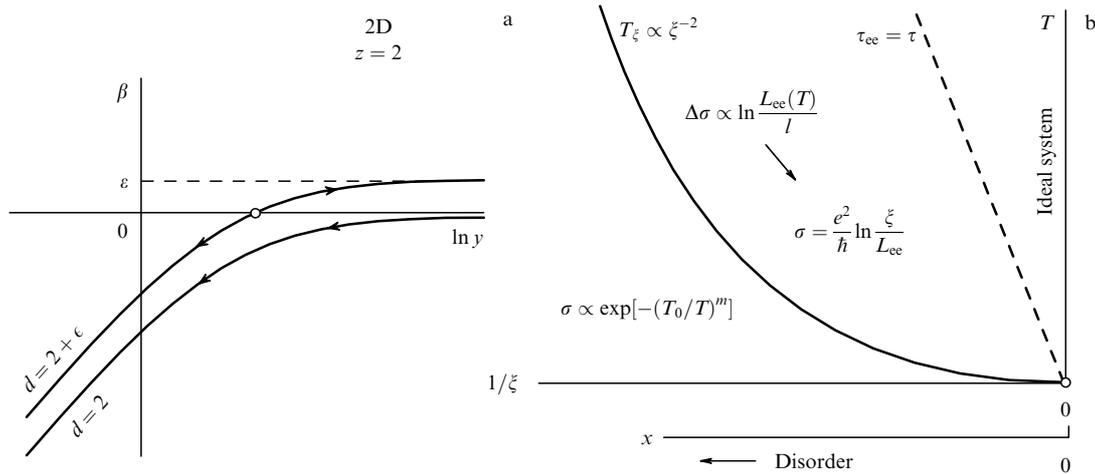


Figure 10. Noninteracting 2D electron gas: (a) differential flow lines for systems with dimension $d = 2$ taken from the scaling diagram in [7] (see also Fig. 6) and $d = 2 + \epsilon$ (see text), and (b) crossover from the logarithmic temperature dependence of conductivity to its exponential dependence, which can be treated as the boundary of the critical region of a virtual quantum transition (see text).

insulator in the sense that localization should inevitably occur in a sufficiently large sample ($L > \xi$) at a sufficiently low temperature ($T < T_\xi$) and an arbitrarily small disorder. This statement follows from the fact that the $d = 2$ flow line in Fig. 6 asymptotically approaches the $\beta = 0$ axis but does not intersect it (this line is shown in Fig. 10a). But in weakly disordered films, the correlation length ξ , which bounds the size L from below, can be unrealistically large, and the temperature T_ξ of the crossover from the region that contributes to the logarithmic quantum correction to the conductivity to the region that is characterized by an exponential temperature dependence is, in contrast, too low. Such films are called metallic films.

Estimates of ξ and T_ξ can be obtained from the assumption [20] that the logarithmically divergent quantum correction $\Delta\sigma$ to the conductivity

$$\sigma = \sigma_2 + \Delta\sigma = \frac{e^2}{h} k_F l - \frac{e^2}{h} \ln \frac{L_{ee}}{l} \quad (57)$$

is of the order of the classical conductivity σ_2 , and hence $\sigma \approx 0$ and

$$\ln \frac{L_{ee}}{l} \approx k_F l. \quad (58)$$

The diffusion length L_{ee} over which electron dephasing occurs is determined by Eqn (43). The value of L_{ee} determined from Eqn (58) is called the correlation length ξ ,

$$\xi = l \exp(k_F l), \quad (59)$$

and the temperature

$$T_\xi = \frac{D\hbar}{\xi^2} \quad (60)$$

determined from the relation $L_{ee} = \xi$ using Eqn (43) is called the crossover temperature. Of course, the conductivity does not vanish at $T = T_\xi$; however, the theory of weak localization is obviously not valid at this temperature, and the resistivity should begin to increase exponentially with

decreasing the temperature at $T < T_\xi$ in samples of size $L > \xi$.

We now plot function (60) and, for convenience, direct the $1/\xi$ axis to the left (see Fig. 10). As a result, this diagram can be conveniently compared with the diagram for a 3D system shown in Fig. 8. We add an axis of the control parameter x as abscissa; this parameter characterizes the degree of disorder, with $x = 0$ corresponding to an ideal system with no disorder.

It is easily seen that the $T(1/\xi)$ curve in Fig. 10 represents the left-hand side of the phase diagram in the vicinity of a metal–insulator transition in a 3D material (cf. Fig. 8) whose phase transition point is located at the edge of the diagram, at the origin ($T = x = 0$). Curve (60) is then the left boundary of the critical region, and the dynamic critical index is

$$z = 2. \quad (61)$$

Using Eqns (53) and (61), we can conclude that for metal–insulator transitions in systems of noninteracting electrons, the dynamic critical index is equal to the dimension, $z = d$.

Thus, systems with dimension $d = 2$ turn out to be boundary systems: a quantum transition is still present in the (x, T) phase plane but is shifted toward its corner, to the unreachable point $x = 0$. The boundary properties of 2D systems can also be found from the flow diagram in Fig. 6. We imagine that the dimension d is a continuous parameter, not restricted to integer values. Straight lines $\beta = d - 2$ represent the asymptotic positions of the flow lines at large values of y . Therefore, the flow lines of a system with dimension $d = 2 + \epsilon > 2$ inevitably cross the abscissa axis $\beta = 0$, and such systems have a metal–insulator transition (Fig. 10a). As $\epsilon \rightarrow 0$, the transition point shifts toward high conductance and goes to infinity.

This interpretation of the curve in Fig. 10 implies that the domain above the $T \propto (1/\xi)^2$ parabola is the critical region of the quantum transition. For 2D systems, one scaling variable (u) is retained in Eqn (34) written for the conductivity in the critical region; it is equal to the ratio of two characteristic lengths,

$$\sigma = F(u) \equiv F\left(\frac{L_\phi}{\xi}\right). \quad (62)$$

In this region, however, the conductivity is typically expressed as the difference between the classical conductivity and the quantum correction,

$$\sigma = \sigma_2 - \Delta\sigma = \frac{e^2}{\hbar} \left(k_{\text{F}} l - \ln \frac{L_{\text{ee}}}{l} \right). \quad (63)$$

We substitute the expression $l = \xi \exp(-k_{\text{F}} l)$ in Eqn (59) in the argument of the logarithm in Eqn (63), pull the exponent out of the logarithm, and obtain

$$\sigma = \frac{e^2}{\hbar} \ln \frac{\xi}{L_{\text{ee}}}. \quad (64)$$

The classical conductivity σ_2 cancels and the remaining part depends only on the scaling variable, as it should be in the critical region. Thus, in this regard, our (x, T) diagram also satisfies the requirements of the theory of quantum phase transitions.

Expression (64) does not hold in the entire ‘quasi-critical’ region $T > T_{\xi}$; moreover, it diverges on the $x = 0$ axis. But near this axis, the elastic mean free path $l \rightarrow \infty$ and the standard expression for L_{ee} loses its meaning because the necessary condition $\tau_{\text{ee}} \gg \tau$ is violated. The boundary of the part of the region where Eqn (64) is invalid ($\tau_{\text{ee}} = \tau$) is shown in Fig. 10 by a dashed straight line plotted under the assumption that $\tau \propto x^{-1}$.

As is shown in Section 5.3, the introduction of interaction can lead to the appearance of a metal–insulator transition in a 2D system. In terms of the phase diagram shown in Fig. 10, this means that the phase transition point shifts from the origin to a point $x_c \neq 0$ on the abscissa axis due to a certain cause, and a boundary additionally appears between the critical and metallic regions. Standard expression (63), which was written for the conductivity in the metallic region and was transformed into Eqn (64), was already used to describe the conductivity in the critical region. This means that the expression for the conductivity in the metallic region should be radically different and that we should expect the appearance of a marginal metal instead of a Fermi liquid. The nature of the electronic states in this hypothetical metal should be rather peculiar, because electrons are assumed to be localized when interaction is turned off [21].

5.2 Spin–orbit interaction

The boundary position of 2D systems makes them sensitive to various types of interaction, e.g., spin–orbit interaction or electron–electron interaction, which can cause a phase transition.

We first consider the spin–orbit interaction. This case is convenient because a flow diagram can be plotted using the (y, β) axes of the initial diagram shown in Fig. 6. At large y , the initial flow line $\beta_2(y)$ deviates down from its right asymptote $\beta = 0$ because of weak localization, which results in a decrease in the conductivity (Fig. 10a). But the spin–orbit interaction changes the sign of the quantum correction, i.e., changes weak localization into antilocalization. Therefore, the sign of the derivative on the right-hand side of the $\beta_2(\ln y)$ curve should change: the curve deviates upward from the asymptote $\beta = 0$, goes to the upper half-plane ($\beta > 0$), and, hence, inevitably crosses the abscissa axis $\beta = 0$ and reaches the left asymptote.

Antilocalization was studied in detail both experimentally [22] and theoretically [23]. In the 2D case ($d = 2$), the

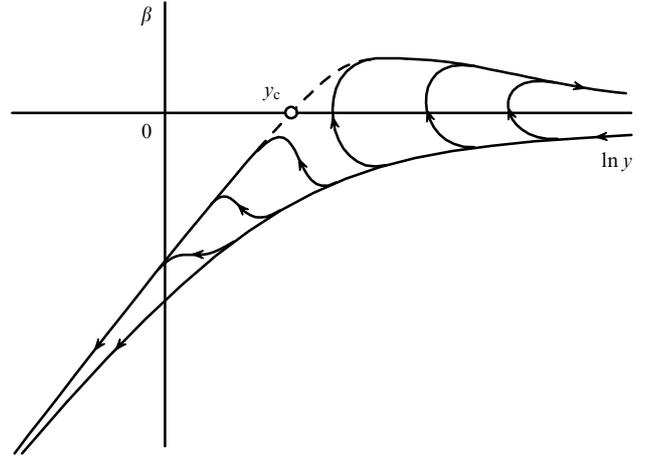


Figure 11. Flow diagram for a 2D electron gas with spin–orbit interaction.

quantum correction to the conductivity is given by

$$\Delta\sigma_2 \approx -\frac{e^2}{\hbar} \int_{\tau}^{\tau_{\text{ee}}} \frac{dt}{t} \left(\frac{3}{2} \exp\left(-\frac{t}{\tau_{\text{so}}}\right) - \frac{1}{2} \right), \quad (65)$$

where τ is the time between elastic collisions, $\tau_{\text{so}} \geq \tau$ is the time between spin-flop collisions, and $\tau_{\text{ee}} = \hbar/T$ is the dephasing time. The motion of the image point in the flow diagram now depends on two parameters, the conductance y and time τ_{so} . For various values of τ_{so} , Fig. 11 shows the family of flow lines located between two envelope curves. The lower curve is taken from the diagrams shown in Figs 6 and 10, and its right-hand side can be obtained from integral (65) if we set $\tau_{\text{so}} = \infty$ in the integrand, and hence the parenthesis in the integrand becomes equal to unity. The upper curve has the same asymptotes; however, its construction implies that $\tau_{\text{so}} = \tau$. Then, we have $t \gg \tau_{\text{so}}$ in the major portion of the integration range, and the parenthesis can be considered equal to $-1/2$.

We now take a sample of size L of a material with an intermediate value of τ_{so} ,

$$\tau \ll \tau_{\text{so}} \ll \infty.$$

The size L bounds the diffusion time t of an electron in a 2D sample until the collision with the boundary by the quantity τ_L , $t < \tau_L \sim \tau(L/l)^2$. Therefore, to describe weak localization in a sample of size L , we must replace the upper integration limit in Eqn (65) with τ_L ,

$$\tau_{\text{ee}} \rightarrow \tau_L \equiv \tau \left(\frac{L}{l} \right)^2.$$

Let L be first very small ($L \gtrsim 1$) and the temperature $T = 0$. Diffusion processes then have no time to develop; electron interference is virtually absent; the conductivity is equal to its classical value; and the image point is on the right on the axis $\beta = 0$. As the size L increases to

$$L_{\text{so}} = l \sqrt{\frac{\tau_{\text{so}}}{\tau}}, \quad (66)$$

spin-flip collisions are insignificant and correction (65) to the conductivity is negative. The image point moves to the left along the lower envelope curve, as in Fig. 10 in the absence of

the spin–orbit interaction. When the size L becomes larger than L_{so} , integral (65) and $\Delta\sigma_2$ change their sign and the image point moves to the upper envelope curve. If this passage occurs at $y < y_c$, the film becomes an insulator at $L \rightarrow \infty$ in any case. But if the spin–orbit interaction is strong, τ_{so} is small and the image point reaches the upper envelope curve at $y > y_c$. Then, the image point continues to move along the upper envelope curve toward large y , and the film retains its metallic properties as $L \rightarrow \infty$. Thus, a metal–insulator transition could be observed in an experiment with the spin–orbit interaction used as a control parameter. So far, only the transformation of weak localization into antilocalization has been demonstrated this way [22].

5.3 Interacting electron gas

We have to clarify what is meant by the absence or presence of the electron–electron interaction. The interaction manifests itself differently in the properties of electron systems; for example, it determines the probability of electron–electron scattering. In the classical theory of metals, electron–electron scattering is considered not to contribute to conductivity, since the total momentum of the electron system and the drift velocity are conserved. When quantum effects are taken into account, this statement becomes invalid because the conductance depends on the dephasing length and electron–electron scattering changes this length. Nevertheless, if the electron–electron interaction affects the conductance only through scattering, the electron system is considered non-interacting, because the interaction affects the conductance as an external action, e.g., scattering by phonons.

Electron–electron scattering is not the only channel of the effect of the interaction on conductance. By the interaction, we mean not the bare Coulomb interaction but the interaction between ‘dressed’ quasiparticles, i.e., the screened interaction that depends on the electron density, the diffusion coefficient of electrons, and (under certain conditions) the sample size or the quantum coherence length [24].

This interaction determines the structure of quantum levels and the ground-state energy, affecting the competition between phases in the vicinity of a phase transition. Electrons are mainly scattered by a screened impurity potential, whose properties depend on the effective electron–electron interaction. This interaction, in turn, depends on the diffusion coefficient and dephasing length, which results in a complex dependence of the effective interaction on all external parameters.

The case of spin–orbit interaction considered in Section 5.2 illustrates a renormalization of the interaction as the sample size L changes: the spin–orbit interaction is actually turned on only when L becomes larger than the value determined by Eqn (66).

As another example, we analyze the picture of possible states in the model with a multivalley electron spectrum developed in [25]. In this model, a change in L causes changes in both the conductance y and the interaction Θ . Therefore, an equation for Θ is added to Eqn (37). The flow diagram is a result of the solution of these two equations. Figure 12 shows part of this diagram calculated by the authors of [25] in slightly different coordinate axes. The variables were changed to facilitate a comparison of this diagram with those shown in Figs 6 and 10, although the quantitative information contained in the initial diagram in [25] is lost. The abscissa of the flow diagram shown in Fig. 12 is the conductance. Thus, the abscissa of all the diagrams in Figs 6–12 is the

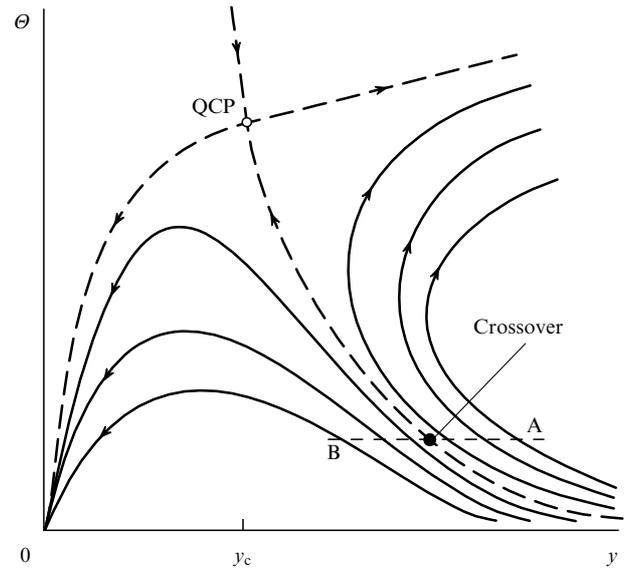


Figure 12. Part of the flow diagram plotted for a 2D electron gas with interaction [25].

same. The ordinate Θ of the flow diagram in Fig. 12 reflects the effective interaction. A noninteracting electron gas, which was discussed in Section 5.1, corresponds to the straight line $\Theta = 0$. As in all the previous diagrams (Figs 6–11), the size is the parameter that determines system motion along the flow lines indicated by arrows. This size is given by the sample size L if $T = 0$ or by the dephasing length L_φ (or L_{ee}), i.e., the maximum length over which quantum coherence is retained in an electron system.

The configurations of the flow lines in Fig. 12 clearly display the specific feature of the interaction Θ that was discussed at the beginning of this section: *as the size L or L_φ changes, the interaction in the system changes (is renormalized)*, and this change is different in different flow lines. The diagram in Fig. 12 is a two-parameter diagram: two parameters are required to specify the flow line of the image point. The flow lines (trajectories) in this diagram occupy the entire half-plane.

We assume that the position of a point in a flow line is determined by L_φ , i.e., the temperature. We assume that we are at point A in a flow line that is metallic because it recedes to the region of high conductivity as $T \rightarrow 0$. We fix the interaction Θ and vary the conductance y ; that is, we vary the degree of disorder. This process is shown by a dashed straight line in the diagram. When moving along this line, we can cross the separatrix and reach point B in a flow line that describes an insulator, because it tends to the point $y = 0$ as $T \rightarrow 0$. To specify the position of a point in line AB, we can use a single parameter, which is precisely what we call the control parameter. With another control parameter, the angle of intersection of the separatrix can be different. For example, we can initiate the crossover from a metallic to a nonmetallic flow trajectory by changing the electron concentration and, thus, simultaneously changing the effective interaction Θ and the conductance.

Changing the control parameter allows passing from one flow trajectory to another, and changing the temperature (size) allows traveling along a flow trajectory. But because the interaction Θ can change when any of the two parameters changes, factorization is absent; that is, we cannot suppose

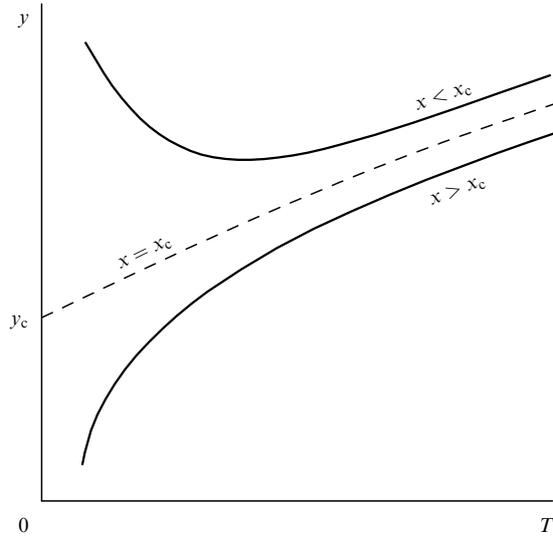


Figure 13. Qualitative scheme for the evolution of the conductance $y(T)$ curves with a control parameter in the model in [25].

that the length ξ depends only on δx and the length L_ϕ depends only on temperature. The consequence of this ‘mixing’ of variables is clearly visible in the flow diagram in Fig. 12. As we move along the separatrix toward the quantum critical point (QCP), the conductance decreases and tends to y_c . Hence, we can draw the following important conclusion, which is qualitatively shown in Fig. 13: the separatrix in the set of temperature dependences is not a horizontal line, as in Fig. 5.

An indication of the presence of a metal–insulator transition in a 2D gas was first obtained in the inversion layer of a field-effect transistor on a Si surface [26]. The presence of the transition was questioned for a long time, because it was in conflict with the concepts formulated in [7] and because the transition was not reproduced in other

materials. However, the uniqueness of silicon was found to be related to a high electron mobility, which allows performing experiments at a very low electron density, where the electron–electron interaction is especially important. With this fact, it was possible to interpret the experimental data using not the theory in [7], which was developed for noninteracting electrons, but in accordance with the model in [25]. As a result, the presence of a metal–insulator transition was supported and a finite slope of the separatrix, which was predicted in [25], was obtained (see Fig. 14a borrowed from [27]).

As is seen from the flow diagram in Fig. 12, the finite slope of the separatrix in the set of the temperature dependences of σ or R of a system of 2D electrons is controlled by the angle at which the separatrix approaches the QCP in the flow diagram. If the tangent to the separatrix is normal to the abscissa axis at the QCP due to any specific reason, the separatrix in the set of temperature dependences has a zero derivative at $T = 0$. Thus, the horizontal position of the separatrix in the set of the temperature dependences of conductivity in the vicinity of the quantum phase transition results from the symmetry of the flow diagram of a specific system and is not an inherent property of all 2D systems. Conversely, a finite slope of the separatrix is not an indispensable consequence of a two-parameter flow diagram.

An inclined separatrix complicates scaling, i.e., the reduction of measurements performed along different flow lines to one universal curve by changing the scales. Nevertheless, the scaling of the resistance $R(T)$ data is possible. Figure 14b shows the scaling carried out in the metallic region of the transition displayed in Fig. 14a. The three lower curves in Fig. 14a are replotted in the coordinates ρ/ρ_{\max} (instead of ρ) and $\rho_{\max} \ln(T/T_{\max})$ (instead of T ; here, ρ_{\max} is given in dimensionless units), and the values ρ_{\max} and T_{\max} correspond to the position of the maximum in each of the experimental curves, which are seen to merge into one curve and to coincide with the theoretical curve. This last curve was plotted using the calculations in [25] and the interaction

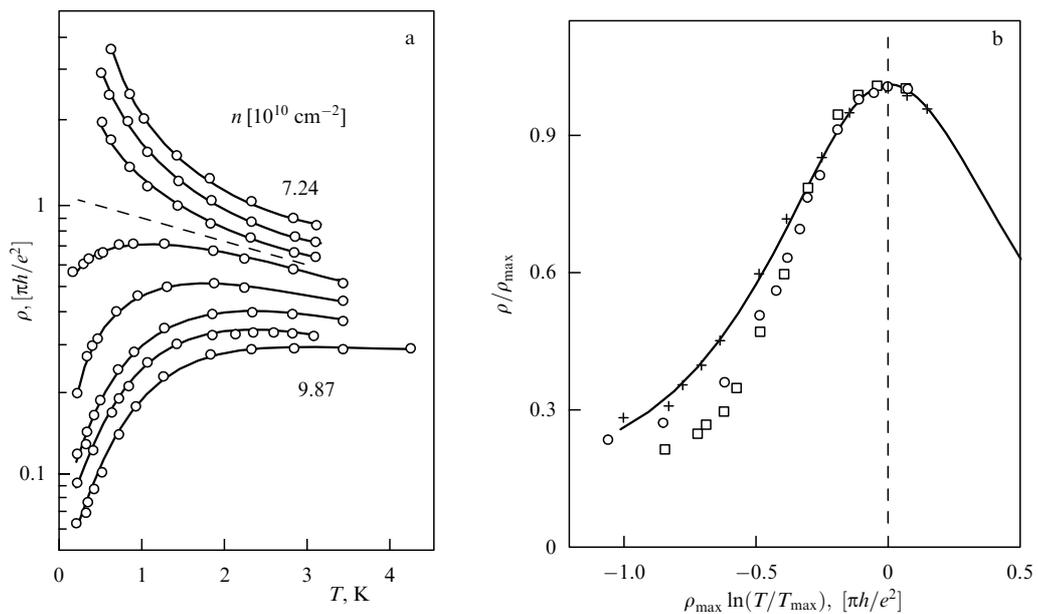


Figure 14. (a) Temperature dependences of the resistivity $\rho(T)$ for a 2D gas in a field-effect silicon transistor in a concentration range containing the metal–insulator transition [27]. (b) Scaling of the three lower curves, which correspond to the concentrations 9.87 (\square), 9.58 (\circ), and $9.14 \times 10^{10} \text{ cm}^{-2}$ ($+$) [27]. Theoretical results are shown by a solid line.

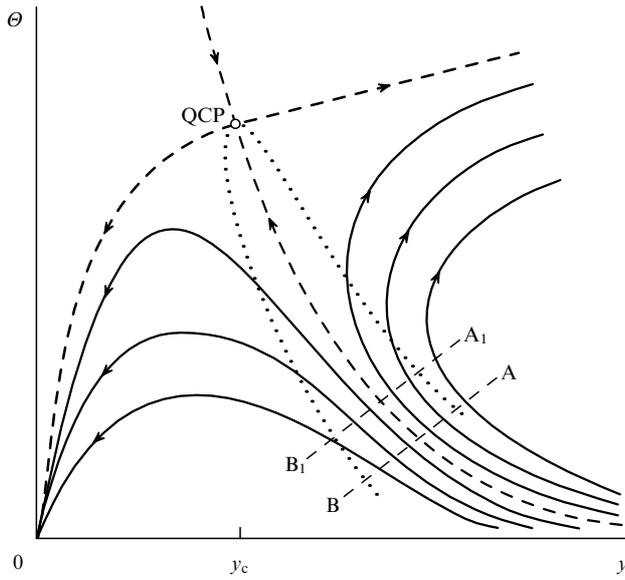


Figure 15. Critical region (indicated by a dotted curve) in the flow diagram of an interacting 2D electron gas.

parameter Θ determined in the same sample from the magnetoresistance data obtained in a field parallel to the 2D plane.

An analogous problem of an inclined separatrix is also often encountered during the experimental processing of the $R_x(T)$ temperature dependences (x is a control parameter) in the vicinity of superconductor–insulator quantum phase transitions. In Ref. [28], a procedure was proposed for ‘correcting’ the curves by introducing the additional linear term

$$R_x(T) \rightarrow R_x(T) - \alpha T, \quad \alpha = \left. \frac{\partial R_{x_c}(T)}{\partial T} \right|_{T=0} \quad (67)$$

into each of them to make the separatrix horizontal and to subsequently perform the standard scaling for 2D systems using scaling variable (35). The idea of this procedure is to compensate for the slope of the separatrix in a flow diagram. However, the correctness of this procedure has not been proved theoretically.

Formally, the (x, T) phase plane also has a meaning for a two-parameter flow diagram. However, the critical neighborhood of the transition can also be directly plotted in the diagram (see Fig. 15). It should be noted that the two-parameter scaling becomes one-parameter in the region adjacent to the separatrix inside the critical region, where flow trajectories are parallel. Indeed, as noted above, one parameter is sufficient to specify the position of a point on line AB, and this parameter does not change when line AB shifts parallel to the separatrix, e.g., to position A_1B_1 , since the flow lines are parallel to each other. As we move along the separatrix toward the QCP, near-separatrix trajectories turn aside alternately, and the strip in which the flow trajectories are parallel, as well as the critical region, narrows.

6. Quantum transitions between the different states of a Hall liquid

States in the plateaus of the quantum Hall effect are the specific phase states of a 2D electron gas with special

transport properties described by the longitudinal (σ_{xx}) and transverse (σ_{xy}) conductivities

$$\sigma_{xx} \rightarrow 0 \text{ as } T \rightarrow 0, \quad \sigma_{xy} = i \frac{e^2}{2\pi\hbar}, \quad i = 0, 1, 2, 3, \dots \quad (68)$$

Such phase states of a 2D electron gas are quantum Hall liquids with different quantum Hall numbers i determined by the values of the Hall conductivity σ_{xy} , Eqn (68), in the plateaus,

$$i = \frac{\sigma_{xy}}{e^2/2\pi\hbar}, \quad i = 1, 2, 3, \dots \quad (69)$$

The transitions from one plateau to another induced by a change in the magnetic field or the electron concentration are clearly visible in experimental curves. As can be seen from Fig. 16, σ_{xy} jumps are accompanied by narrow σ_{xx} peaks. These jumps are quantum phase transitions, and they should fit both theoretical versions compared in Section 5.

Figure 16 only displays the plateaus of the integer quantum Hall effect, which is considered below. The integer quantum Hall effect can also be realized in a noninteracting electron gas; therefore, it can be described without taking the interaction into account.

To construct the flow diagram of a 2D system of *noninteracting electrons* in a strong magnetic field, we need two conductance components that are equivalent to conductivity components σ_{xx} and σ_{xy} . Correspondingly, Eqn (37) transforms into the set of two equations

$$\begin{aligned} \frac{d \ln \sigma_{xx}}{d \ln L} &= f_1(\sigma_{xx}, \sigma_{xy}), \\ \frac{d \ln \sigma_{xy}}{d \ln L} &= f_2(\sigma_{xx}, \sigma_{xy}). \end{aligned} \quad (70)$$

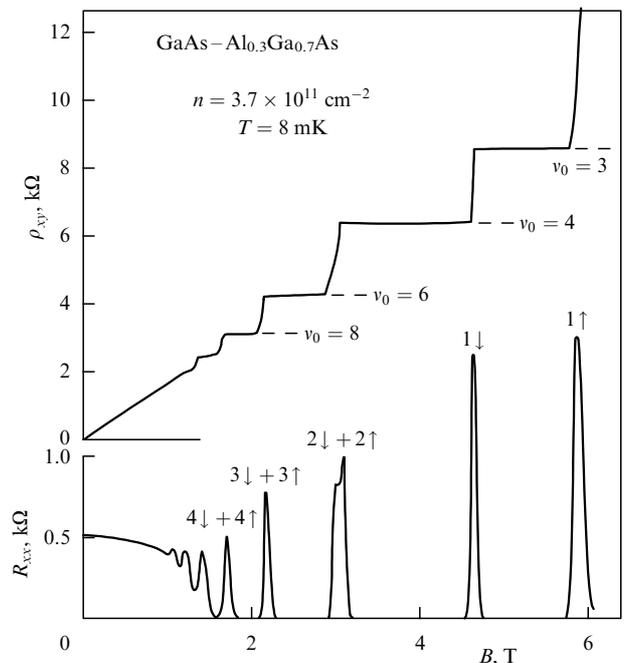


Figure 16. The magnetoresistance R_{xx} and the Hall resistivity ρ_{xy} versus the magnetic field B in a GaAs–Al_xGa_{1-x}As heterostructure at $T = 8$ mK [29]. The electron density in the 2D layer is $3.7 \times 10^{11} \text{ cm}^{-2}$ and the mobility is $\mu = 4.1 \times 10^4 \text{ cm}^2 (\text{V s})^{-1}$.

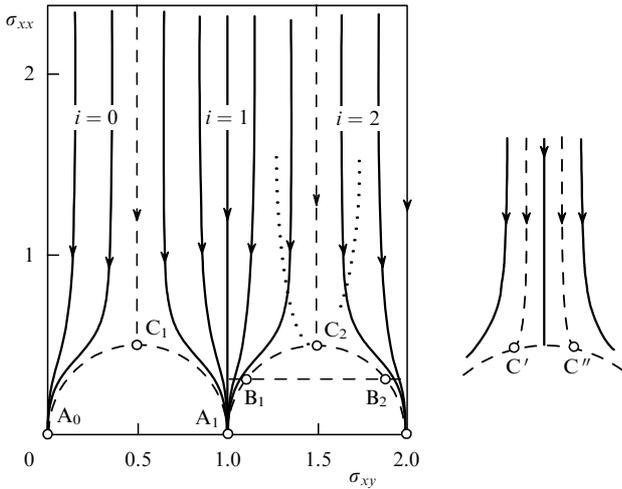


Figure 17. Flow diagram for a 2D electron gas in a strong magnetic field [30]. The coordinates are represented by the conductivity tensor components σ_{xy} and σ_{xx} in dimensionless units $e^2/2\pi\hbar$. Separatrices are indicated by dashed lines; A_i are stationary singular points and C_i are unstable singular points, which are quantum transition points similar to QCP in Fig. 12. The dashed line indicates a critical region near C_2 . Horizontal dashed line B_1 – B_2 indicates an image-point jump as the control parameter changes. The right-hand side of the figure shows the hypothetical flow diagram that corresponds to a split phase transition and the appearance of a metallic phase.

Eliminating the variable L from Eqn (70), we find a relation between σ_{xx} and σ_{xy} , which can be displayed as curves in the $(\sigma_{xy}, \sigma_{xx})$ plane [30]. This yields a flow diagram for the 2D system of noninteracting electrons in a strong magnetic field (Fig. 17). The flow lines in this diagram are separated by separatrices periodically repeated along the σ_{xy} axis. This is again a two-parameter diagram, but due to a strong magnetic field rather than to interaction. As in the previous flow diagrams, we can move along arrows in flow trajectories by either increasing the sample size L or decreasing the temperature at large L , i.e., increasing L_ϕ .

In the region below points C_i , the separatrix is split such that two equivalent points located at the same height (B_1 and B_2) appear in it. The flow lines inside the semicircles in Fig. 17 are omitted because they are thought to exist separately from the flow lines outside these semicircles: as the control parameter changes in the region below points C_i , the motion leads to a jump between points B_1 and B_2 and to a sharp increase in σ_{xy} . Strictly speaking, the jump should occur along a line located outside the semicircle and bypassing the point C_i from above. However, for simplicity, this jump is indicated by a horizontal dashed line.

The control parameter in the quantum Hall effect regime is usually given by the electron concentration or the magnetic field. Their effect on the state of a real system depends on the random field of impurities and other defects, which transforms discrete Landau levels into minibands and specifies the energy structure and the character of wave functions in them. Because a magnetic field determines the magnetic length $r_B = (\hbar c/|e|B)^{1/2}$ as a characteristic scale, we can speak about two limit types of random potential, a potential with large-scale fluctuations of characteristic sizes $\zeta \gg r_B$ and a short-range potential with $\zeta \ll r_B$. In the long-period potential model, an energy value ε_c exists near the center of each Landau miniband such that a delocalized electron wave function corresponds to this value. If the wave function is

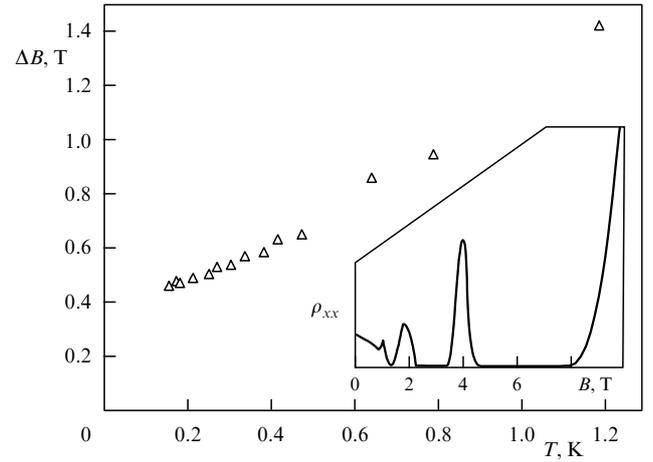


Figure 18. Temperature dependence of the peak width of the longitudinal resistivity ρ_{xx} of a 2D electron gas in the GaAs/AlGaAs heterojunction during the $2 \rightarrow 1$ transition in the magnetic field about 4 T [32]. The carrier mobility and concentration at $T = 1.5$ K are $\mu = 34,000$ cm^2 $(\text{V s})^{-1}$ and $n = 1.4 \times 10^{11}$ cm^{-2} . The inset shows the $\rho_{xx}(B)$ curve recorded at 150 mK.

strictly delocalized only at $\varepsilon_F = \varepsilon_c$ and the random field lifts the degeneracy of energy levels, a change in the electron concentration results in a jump-like transformation of the system from one phase state into another through an isolated energy state with a delocalized wave function at the Fermi level. This behavior is implied in the flow diagram in Fig. 17, where each isolated metallic state corresponds to a peculiar separatrix.

The actual width $\delta\varepsilon$ of the energy range with delocalized wave functions depends on finer processes, such as tunneling between two semiclassical trajectories that are close to each other in the vicinity of the saddle point (magnetic breakdown). In essence, $\delta\varepsilon$ is the energy uncertainty of any delocalized state. Another source of increasing the $\delta\varepsilon$ range is the finiteness of the lengths L and L_ϕ . If the $\delta\varepsilon$ range is finite, a separatrix is split into two parallel lines and the phase transition is split into two transitions: a metallic state with a partly filled layer of extended states at the Fermi level should appear between two Hall-liquid states whose indices i differ by unity. The corresponding hypothetical flow diagram is shown on the right-hand side of Fig. 17.

At first glance, it seems that experiment can distinguish between these two hypothetical possibilities. For definiteness, we assume that the electron concentration n changes in experiment (considerations for a change in the magnetic field are similar). As the concentration changes, the Fermi level moves along the energy scale. When states at the Fermi level are delocalized, the σ_{xx} conductivity is finite, and the σ_{xy} conductivity is in the intermediate region between two plateaus. Therefore, the temperature dependence of the concentration range of the intermediate region (i.e., the σ_{xx} peak width, the $\partial\sigma_{xy}/\partial n$ derivative at the center of the intermediate region, etc.) extrapolated to $T = 0$ must determine the energy range $\delta\varepsilon$ of the delocalized states.

However, the experimental results were found to be ambiguous. On the one hand, the comprehensive experiments in [31, 32] give a finite value of $\delta\varepsilon$. For example, Fig. 18 shows the transition width measured from the ρ_{xx} resistivity peak width of a 2D gas in the GaAs/AlGaAs heterojunction. The measured function is seen to be reliably

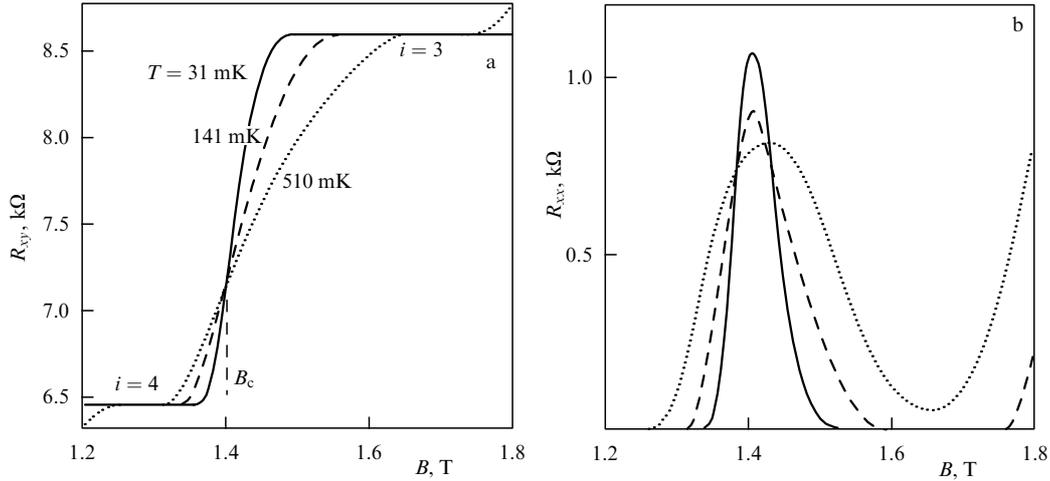


Figure 19. Transverse (R_{xy}) and longitudinal (R_{xx}) resistances of an $\text{Al}_x\text{Ga}_{1-x}\text{As} - \text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ heterostructure, $x = 0.85\%$, at various temperatures. The critical magnetic field of the 4–3 transition determined from the intersection point of the $R_{xy}(T)$ curves is $B_c = 1.40$ T [33].

extrapolated to $\Delta B_0 \approx 0.35$ as $T \rightarrow 0$. However, as we see below, many experiments give the opposite result: as the temperature decreases, the transition width tends to zero (see, e.g., Fig. 20). The factor that determines $\delta\varepsilon$ has not yet been revealed. In any case, there is no unique relation between $\delta\varepsilon$ and mobility. The statistical characteristics of the random potential, which are poorly controlled, are likely to play a key role.

We now move to experiments that do not exhibit a finite energy layer with delocalized states. Although the flow diagram of $i \rightarrow i + 1$ transitions between different states of a Hall liquid has two parameters (see Fig. 17), it is symmetric with respect to both the $\sigma_{xy} = (i + 1/2)(e^2/2\pi\hbar)$ and $\sigma_{xy} = i(e^2/2\pi\hbar)$ axes. Therefore, we may use the version of the general theory of quantum phase transitions that is based on Eqns (25) and (26) and assumes that scaling formulas of type (34) are valid. Because we consider only 2D electron systems, all resistances have the same dimension [Ω] and must have the form

$$R_{uv} = F_{uv} \left(\frac{L\varphi}{\xi} \right) = F_{uv} \left(\frac{\delta x}{T^{1/zv}} \right) \quad (71)$$

in the vicinity of the transition. Here, the u and v subscripts stand for the x and y coordinates, F_{uv} is an unknown function, and the argument of the arbitrary F function is written using the last form in Eqn (35).

In contrast to the case of metal–insulator transitions, we analyze experimental results instead of calculating or predicting the values of v and z . As an example, Fig. 19 shows the magnetic-field dependences at various temperatures of the longitudinal (R_{xx}) and transverse (R_{xy}) resistances of a Hall bar in a GaAs-based heterostructure [33].

According to Eqn (71), the scaling variable

$$u = \frac{\delta x}{T^{1/zv}} \quad (72)$$

is identically zero at all temperatures in the case where the control parameter takes the critical value; correspondingly, we have

$$R_{uv}(x_c, T) = \text{const}. \quad (73)$$

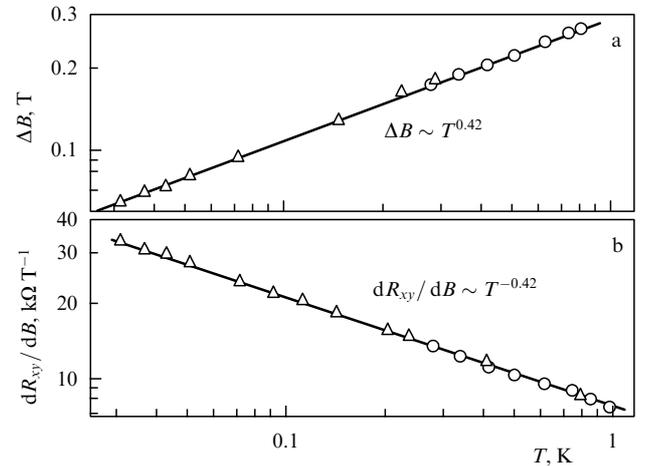


Figure 20. Determination of the critical index $\kappa = 1/zv$ for transition 4–3 using the $R_{xx}(T)$ peak width (a) and matching the slopes of the intersecting $R_{xy}(T)$ curves (b). The data were obtained from the sample used in Fig. 19 in a solution refrigerator (triangles) and a liquid ^3He cryostat (circles) [33].

Therefore, *separatrix (73) must be horizontal, and all isotherms $R_{uv}(x, T = \text{const})$ must intersect at one point $x = x_c$* . This is the first test of the applicability of Eqn (71).

We first focus on the $R_{xy}(T)$ curves. As can be seen from Fig. 19, the $R_{xy}(T)$ isotherms obtained for an $\text{Al}_x\text{Ga}_{1-x}\text{As} - \text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ heterostructure with $x = 0.85\%$ [33] do intersect at one point, $B_c = 1.40$ T. Near the intersection point, all the $R_{xy}(T)$ curves can be expanded into a series and replaced by straight lines $(\partial R_{xy}/\partial B)_{B_c}(B - B_c)$. As the slopes of these lines change from $(\partial R_{xy}/\partial B)_{B_c}$ to $(\partial R_{xy}/\partial B)_{B_c}/T^\kappa$, where $\kappa = 1/zv$, *all the straight lines must merge into one line*. The choice of the value of κ at which the relation

$$\frac{(\partial R_{xy}/\partial B)_{B_c}(T)}{T^\kappa} = \text{const} \quad (74)$$

holds is the second step in the application of the scaling procedure, and the possibility of this choice is the second condition for the applicability of the theory. To choose κ , we plot $(\partial R_{xy}/\partial B)$ versus T in a log–log scale (Fig. 20).

Formally, Eqn (71) can be applied to both longitudinal (R_{xx}) and transverse (R_{xy}) resistances. However, the peak height depends on the temperature; that is, the point of the maximum does not satisfy condition (73). Leaving aside the reasons for this fact, we can use the longitudinal resistance data for scaling analysis by accounting for the integral property of the $R_{xx}(B)$ functions in the vicinity of the transition, namely, the peak half-width ΔB determined by a certain algorithm. In Fig. 20, ΔB is determined as the distance between the two maxima of the $(\partial R_{xx}/\partial B)$ derivative.

As can be seen from Fig. 20, an analysis of both families of functions gives the same value of the critical index $\kappa = 0.42$, which is an additional argument for our scaling procedure.

This value of the critical index $\kappa = 0.42$ was repeatedly obtained in heterostructures made from various materials. However, contrary to expectations, this value is not universal: other values in the range from 0.2 to 0.8 were also detected in a number of experiments on various heterostructures. This scatter calls for explanations, since scaling relations and critical indices are usually universal.

At $T = 0$, transition occurs when the condition $\varepsilon_F = \varepsilon_c$ is satisfied and therefore the difference $\delta\varepsilon = |\varepsilon_F - \varepsilon_c|$ is the only ‘internal’ control parameter of the system, with the correlation length ξ depending on it in accordance with a power law:

$$\xi \propto (\delta\varepsilon)^{\nu_1}. \quad (75)$$

From this standpoint, the magnetic field B or the 2D electron concentration n , which depends on the gate voltage V_g , are ‘external’ control parameters x ,

$$\delta x \equiv |B_c - B| \propto (\delta\varepsilon)^{\nu_2} \quad \text{or} \quad (76)$$

$$\delta x \equiv |n_c - n| \propto |V_{gc} - V_g| \propto (\delta\varepsilon)^{\nu_2}.$$

The exponent ν_2 is the same in both cases. This is supported by the fact that the $R_{xx}(B)$ and $R_{xx}(V_g)$ experimental curves recorded using the same sample under equivalent conditions differ only in the scales on the abscissa axis. Eventually, we can write Eqn (25) as

$$\xi \propto (\delta x)^{\nu}, \quad \nu = \nu_1 \nu_2, \quad (77)$$

where the relation between the control parameter δx and the correlation length ξ consists of the following two links: ξ depends on the position of the Fermi level ε_F with respect to the delocalized level ε_c , and the difference $\varepsilon_F - \varepsilon_c$ in turn depends on δx . Correspondingly, according to Eqn (77), the index ν is the product of ν_1 and ν_2 . The relation between ξ and $\delta\varepsilon$ and the related index ν_1 are likely to be universal and the same for all transitions between different quantum Hall liquids. But the index ν_2 is determined by the density of states $g(\varepsilon)$ in the vicinity of the energy ε_c , and hence depends on the specific features of the random potential; this potential can be long- or short-range, statistically symmetric or asymmetric with respect to the mean value, and so on. The authors of [33], whose curves are used in this section, just studied the effect of specific features of the random potential on κ .

The $\Delta B(T)$ curve in Fig. 20 differs radically from the curve in Fig. 18; in the latter case, a free term ΔB_0 was introduced for the experimental data to be approximated by a power function. Nevertheless, we can also perform scaling analysis of the experimental data in this case using the second hypothetical version of the flow diagram (see Fig. 17). With

this interpretation, the presence of the ΔB_0 term means that in the $2\Delta B_0$ range of the control parameter B , the image point in the flow diagram moves across the metallic-phase corridor and falls on a separatrix not at B_0 corresponding to a maximum of $\sigma_{xx}(B)$ and the derivative $\partial\sigma_{xy}/\partial B$ but at $B_0 + \Delta B_0$. This problem is discussed in detail in review [34]. Here, we only note that Fig. 18 actually contains this scaling analysis. By extrapolating the $\Delta B(T)$ dependence to $T = 0$, we can determine the delocalized-state layer width in units of the magnetic field $2\Delta B_0$ and find that $\Delta B(T) - \Delta B_0$ depends linearly on T . This means that $\kappa = 1$ in this experiment. The same value of κ was obtained earlier in [31].

7. Conclusion

All the considered cases of metal–insulator transitions were found to be adequately described by flow diagrams. The list of theoretical works that have successfully used this technique begins with work [7], where a theoretical model for noninteracting electrons in a zero magnetic field was developed. The last achievement in this field is the construction of a flow diagram for a 2D model system of interacting electrons and the demonstration of the possibility of a metal–insulator transition in this system [25].

As regards the general theory of quantum phase transitions, a phase transition in a 3D system of noninteracting electrons demonstrates that, in principle, this theory can be used to describe localized–delocalized electron transitions. Neither conductance, which is used as a physical quantity specifying the state of the system, nor disorder, which is the main control parameter, are substantial obstacles for this theory. However, as usual, various particular cases require theoretical versions of various degrees of complexity. For example, the version described in this review cannot be applied to the model in [25].

The relative role and possibilities of both theoretical approaches are demonstrated in describing the integer quantum Hall effect. The flow diagram in Fig. 17 is very convenient for the discussion of the *types of transitions* that are possible in a system and for the formulation of questions to be experimentally checked. Many of these questions are still open. For example, it is unclear which of the versions of the flow diagram in Fig. 17 is realized in reality and whether the transition between the states with quantum indices i and $i \pm 1$,

$$i \rightleftharpoons i \pm 1, \quad (78)$$

is split [i is determined by Eqn (69)]. There are also problems related to the topology of the flow diagram. According to Fig. 17, transitions where quantum number i changes by more than unity are impossible [35, 36]. However, when theorists interpret many experimental data, they state that such transitions occur (see, e.g., review [34] and the references therein).

The general theory of quantum phase transitions does not consider the problem of the relative position of various transitions in the phase plane. This theory describes the critical neighborhood of *one* certain transition. If we assume that (a) a transition exists and (b) the factorization $\zeta = \zeta(\delta x)$, $L_\phi = L_\phi(T)$ occurs in its critical region, then the resistivity can be described by Eqns (34)–(36) (see Sections 2.3 and 5.3). Then, we can determine the transition point $x = x_c$ and critical indices by processing the $R(x, T)$ curves. This

procedure was performed in paper [33], which was discussed in Section 6. The value of the results obtained in [33] is to become clear when the questions formulated above and motivated by flow diagrams are answered.

Acknowledgments

We thank A Finkel'shtein and D Khmel'nitskii for the useful discussions.

This work was supported by the Program for the Support of Leading Scientific Schools (project no. NSh-5930.2006.2) and the Russian Foundation for Basic Research.

References

- Sachdev S *Quantum Phase Transitions* (Cambridge: Cambridge Univ. Press, 1999)
- Sondhi S L et al. "Continuous quantum phase transitions" *Rev. Mod. Phys.* **69** 315 (1997)
- Vojta T "Quantum phase transitions", in *Computational Statistical Physics* (Eds K H Hoffmann, M Schreiber) (Berlin: Springer, 2002); cond-mat/0010285
- Thouless D J *Phys. Rep.* **13** 93 (1974)
- Patashinskii A Z, Pokrovskii V L *Fluktuatsionnaya Teoriya Fazovykh Perekhodov* (Fluctuation Theory of Phase Transitions) 2nd ed. (Moscow: Nauka, 1982) [Translated into English (Oxford: Pergamon Press, 1979)]
- Goldenfeld N *Lectures on Phase Transitions and the Renormalization Group* (Reading, Mass.: Addison-Wesley, 1992)
- Abrahams E et al. *Phys. Rev. Lett.* **42** 673 (1979)
- Fisher M E "Renormalization group theory: its basis and formulation in statistical physics" *Rev. Mod. Phys.* **70** 653 (1998)
- Lee P A, Ramakrishnan T V "Disordered electronic systems" *Rev. Mod. Phys.* **57** 287 (1985)
- Gantmakher V F *Elektrony v Neuporyadochemykh Sredakh* (Electrons in Disordered Media) 2nd ed. (Moscow: Fizmatlit, 2005); *Electrons and Disorder in Solids* (New York: Oxford Univ. Press, 2005)
- Altshuler B I, Aronov A G "Electron–electron interaction in disordered conductors", in *Electron–Electron Interactions in Disordered Systems* (Modern Problems in Condensed Matter Sciences, Vol. 10, Eds A L Efros, M Pollak) (Amsterdam: North-Holland, 1985)
- Imry Y *J. Appl. Phys.* **52** 1817 (1981)
- Al'tshuler B L, Aronov A G *Pis'ma Zh. Eksp. Teor. Fiz.* **37** 349 (1983) [*JETP Lett.* **37** 410 (1983)]
- Dodson B W et al. *Phys. Rev. Lett.* **46** 46 (1981)
- Imry Y, Ovadyahu Z *J. Phys. C: Solid State Phys.* **15** L327 (1982)
- Maliapaard M C et al. *Phys. Rev. Lett.* **61** 369 (1988)
- Teplinskii V M, Gantmakher V F, Barkalov O I *Zh. Eksp. Teor. Fiz.* **101** 1698 (1992) [*Sov. Phys. JETP* **74** 905 (1992)]
- Shlimak I et al. *Phys. Rev. Lett.* **77** 1103 (1996); *J. Phys.: Condens. Matter* **9** 9873 (1997)
- Castner T G "Hopping conduction in the critical regime approaching the metal–insulator transition", in *Hopping Transport in Solids* (Modern Problems in Condensed Matter Sciences, Vol. 28, Eds M Pollak, B Shklovskii) (Amsterdam: North-Holland, 1991)
- Larkin A I, Khmel'nitskii D E *Zh. Eksp. Teor. Fiz.* **83** 1140 (1982) [*Sov. Phys. JETP* **56** 647 (1982)]
- Dobrosavljević V et al. *Phys. Rev. Lett.* **79** 455 (1997)
- Bergmann G *Phys. Rep.* **107** 1 (1984)
- Hikami S, Larkin A I, Nagaoka Y *Prog. Theor. Phys.* **63** 707 (1980)
- McMillan W L *Phys. Rev. B* **24** 2739 (1981)
- Punnoose A, Finkel'stein A M *Science* **310** 289 (2005)
- Kravchenko S V et al. *Phys. Rev. B* **50** 8039 (1994)
- Anissimova S et al. *Nature Phys.* **3** 707 (2007); cond-mat/0609181
- Gantmakher V F et al. *Pis'ma Zh. Eksp. Teor. Fiz.* **71** 231 (2000) [*JETP Lett.* **71** 160 (2000)]
- Ebert G et al. *Solid State Commun.* **44** 95 (1982)
- Khmel'nitskii D E *Pis'ma Zh. Eksp. Teor. Fiz.* **38** 454 (1983) [*JETP Lett.* **38** 552 (1983)]
- Shashkin A A, Dolgoplov V T, Kravchenko G V *Phys. Rev. B* **49** 14486 (1994)
- Balaban N Q, Meirav U, Bar-Joseph I *Phys. Rev. Lett.* **81** 4967 (1998)
- Wanli Li et al. *Phys. Rev. Lett.* **94** 206807 (2005)
- Shashkin A A *Usp. Fiz. Nauk* **175** 139 (2005) [*Phys. Usp.* **48** 129 (2005)]
- Kivelson S, Lee D-H, Zhang S-C *Phys. Rev. B* **46** 2223 (1992)
- Huckestein B *Rev. Mod. Phys.* **67** 357 (1995)