Duality in the quantum Hall system

C. A. Lütken and G. G. Ross
Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom
(Received 6 May 1991; revised manuscript received 9 December 1991)

We suggest that a unified description of the integer and fractional phases of the quantum Hall system may be possible if the scaling diagram of transport coefficients is invariant under linear fractional (modular) transformations. In this model the hierarchy of states, as well as the observed universality of critical exponents, are consequences of a discrete SL $(2,\mathbb{Z})$ symmetry acting on the *parameter space* of an effective quantum-field theory. Available scaling data on the position of delocalization fixed points in the integer case and the position of mobility fixed points in the fractional case agree with the model within experimental accuracy.

I. INTRODUCTION

Both the integer and fractional plateaus observed[1,2] in the transverse conductivity σ_{xy} of the quantum Hall system are well understood in terms of variational wave functions approximating the responsible ground states.[3,4] There has been less success in extending this first quantized description to a second quantized many-body (field-theoretic) formalism capable of explaining not only the plateaus, but also the observed scaling in the transitions between the plateaus. Experiments[5] suggest that the same scaling exponents apply to transitions between both integer and fractional levels, but to date there is no explanation for this within the context of the hierarchy generation scheme[3,4] which so successfully accounts for the observed levels. Indeed, the physics of the integer quantum Hall effect can be understood in terms of noninteracting electrons, while the fractional effect involves the repulsion between electrons,[6] making it difficult to understand the connection between the scaling phenomena. This problem has even prompted some authors to suggest a completely different class of variational wave functions in order to generate the hierarchy in a way which would relate the fractional and integer effects in an obvious manner.[7]

In this paper we consider the general question of how to construct a scaling theory for the quantum Hall system which is capable of describing the extended states, found in the transitions between levels, which are responsible for the transport of electric charge through the system and, therefore, the changes in σ_{xy} and σ_{xx} between the plateaus. For a review of previous work in this direction we refer the reader to Ref. 8. Because of the mentioned "universality" of the critical exponents, our task is to construct a *single* quantum-field theory satisfying apparently incompatible constraints: on one hand it should contain an infinite set of nested phases with distinct physical properties; on the other hand the critical points characterizing the transition between the phases should in some sense be universal. (In a real system, Wigner crystallization, which has recently been observed,[9] will disallow all but a finite number of these phases, but this refinement will not be included in the model considered here.) More precisely, what we seek is a single *family* of field theories parametrized by the longitudinal (dissipative) conductivity σ_{xx} and the transverse (Hall) conductivity σ_{xy}, with the property that σ_{xy} is forced to take a fractional value when σ_{xx} vanishes, and which provides a natural explanation for why the scaling observed between these plateaus is always the same.

We propose that this can be achieved by making the partition function invariant under an infinite discrete group acting on the *parameter space* $(\sigma_{xy},\sigma_{xx})$, in such a way that it maps all critical points into each other. A symmetry of this type will organize the parameter space into regions, or equivalence classes, which are the different phases of the system. These phases are mapped into each other under the action of elements of the symmetry group, but they are not identified because the physical parameters have different values in different phases. In short, such a symmetry will classify, rather than identify, different regions of parameter space. It will, however, force the scaling equations to have the same form close to all critical points, thus explaining why the critical exponents are always the same.

In view of the existence of the hierarchy, it is clear that the phase structure of the two-dimensional parameter space of the quantum Hall system must be extremely complicated. It might, therefore, appear to be a daunting task to identify an effective-field theory with this structure, much less a microscopic theory which would give rise to such a field theory in the long-wavelength limit. In fact, while exploring Abelian lattice gauge theories with theta terms,[10] Cardy[11] discovered that a certain class of self-dual spin models has the phase diagram shown in Fig. 1, which has most, if not all, of the properties we want. The complexity of this phase diagram is a consequence of including a topological theta term in the action, which enhances the usual duality or Kramers-Wannier[12] symmetry of the partition function to an infinite discrete *non-Abelian* group, SL$(2,\mathbb{Z})$. Because SL$(2,\mathbb{Z})$ is realized by complex fractional linear (Möbius) transformations, it follows that the phases only touch the real axis at rational numbers. As was apparently first

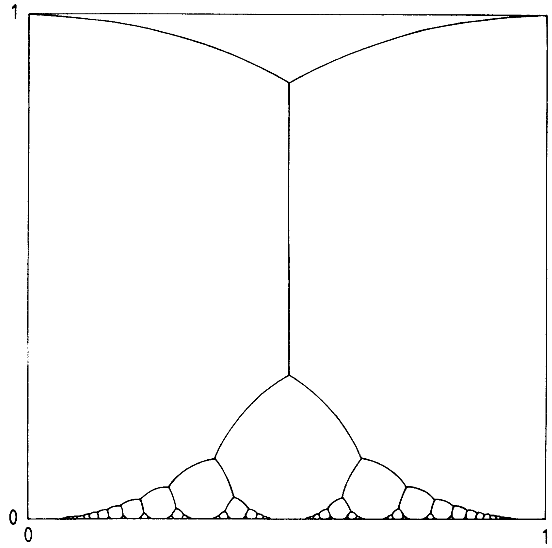


FIG. 1. SL(2,$\mathbb{Z}$) invariant phase diagram.

noted by Girvin and MacDonald[13] (see also Refs. 14 and 15), this so-called "oblique confinement"[16] scenario is reminiscent of the hierarchy generation scheme of the fractional quantum Hall effect,[4] which is designed to generate only odd-denominator filling fractions by iterating two real fractional linear transformations.

II. DUALITY

Self-duality is found in many statistical models, where it relates high- and low-temperature properties of the system. Cardy generalized the usual duality transformation to the case which includes a topological theta term coupling two copies of the Ising model.[11] While we are not suggesting that we can, at the moment, derive this model from the microscopic physics of electrons in a transverse magnetic field, it is worth exhibiting here not only because it is a prototype possessing the kind of hierarchical structure that we need; it also provides much needed physical intuition which will be useful for motivating the structure of the effective-field theories that we shall be discussing.

The partition function studied by Cardy and Rabinovici is[10,11]

$$Z_p(g^{-2},\theta)=\mathrm{Tr}\exp\left[-\frac{1}{2g^2}(\Delta_\mu\phi_a-2\pi s_{\mu a})\times(\Delta^\mu\phi^a-2\pi s^{\mu a})+ipn_a\phi^a+\frac{ip\theta}{32\pi^2}\epsilon_{\mu\nu}\epsilon_{ab}(\Delta^\mu\phi^a-2\pi s^{\mu a})\times(\Delta^\nu\phi^b-2\pi s^{\nu b})\right], \quad (1)$$

where μ and ν label the two spatial dimensions, a and b label the two models whose spins ϕ_1 and ϕ_2 reside on dual lattices, and $s_{\mu a}$ are integer-valued plaquette variables. n_a is an integer-valued link variable representing a conserved electric current, and its piece of the action constrains ϕ_a to take values in $\mathbb{Z}_p$.

Duality in this model corresponds to the exchange of electric and magnetic monopole charges, where the monopole current is given by $m_a=\frac{1}{2}\epsilon_{\mu\nu}\Delta^\mu s_a^\nu$. This is best expressed as a transformation on the complex parameter

$$z=\frac{\theta}{2\pi}+i\frac{2\pi}{pg^2}, \quad (2)$$

in terms of which the duality transformation is just a linear-fractional transformation

$$z\to S(z)=-\frac{1}{z}. \quad (3)$$

The parameter z is more convenient than Cardy's choice $\zeta=i\bar{z}$ (the bar denotes complex conjugation) because with this parametrization the duality transformation S acts on the complex upper-half plane as a conventional generator of SL(2,$\mathbb{Z}$).

The term in (1) which is proportional to θ couples the monopole current to the spin fields, and it is only nonzero in the presence of fields with nontrivial topology. It is, in fact, a discretized and dimensionally reduced version of the usual four-dimensional theta term $\theta\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}$. The angular nature of the theta term means that the action is periodic in θ, so that Z_p is also invariant under the translation

$$z\to T(z)=z+1. \quad (4)$$

Because $S^{-1}=S$, arbitrary compositions of S and T are necessarily of the form $R=T^{p_1}ST^{p_2}S\ldots ST^{p_n}$ ($p_1,p_2,\ldots,p_n\in\mathbb{Z}$), and since this is a continued fraction it can be written as a linear-fractional transformation

$$R(z)=\frac{az+b}{cz+d},\quad a,b,c,d\in\mathbb{Z}, \quad (5)$$

with unit determinant $ad-bc=1$. Furthermore, every modular transformation is of this type, so that the symmetries S and T, in fact, generate all of the modular group SL(2,$\mathbb{Z}$).

This symmetry may be given a geometrical interpretation as follows. If the fields ϕ_a are taken to have periodic boundary conditions, then in the continuum limit (1) looks like a two-dimensional nonlinear sigma model with a toroidal target space, whose action

$$(g_{ab}\delta^{\mu\nu}+it_{ab}\epsilon^{\mu\nu})\partial_\mu\phi^a\partial_\nu\phi^b$$

is parametrized by the toroidal metric $g_{ab}=g_{xx}\delta_{ab}$ and the torsion $t_{ab}=t_{xy}\epsilon_{ab}$. In two dimensions the torsion is topological, and formally similar to the theta term in (1). The parameter space is four dimensional, corresponding to the three possible ways to deform the independent components of the symmetric metric g_{ab}, and the one independent component of the torsion. Because the target space is a torus, the parameter space is automatically a complex space, so it is convenient and conventional[17] to collect the four real parameters into two complex "moduli:"

$$\tau=(g_{xy}+i\sqrt{\det g})/g_{yy}, \quad \sigma=t_{xy}+i\sqrt{\det g} \ . \tag{6}$$

τ is just the complex structure ("shape") parameter of the torus, which in our case is frozen at $\tau=i$, while the imaginary part of σ parametrizes the Kähler form ("size") of the torus, because it determines the volume of the target space. In our homogeneous and isotropic case the complexified Kähler form simplifies to $\sigma=t_{xy}+ig_{xx}$, which is just z. Furthermore, it is well known that the space of τ's of a torus is invariant under SL(2,$\mathbb{Z}$), but it is equally true for the other parameter σ of the nonlinear model, thus providing a geometric explanation of the SL(2,$\mathbb{Z}$) invariance of the parameter space of (1).

Finally, we also note that $\mathcal{Z}_p$ is invariant under the transformation

$$z\to J(z)=-\bar{z} \ , \tag{7}$$

whose physical origin is time-reversal invariance. J is the only automorphism of SL(2,$\mathbb{Z}$) which is not, itself, a (holomorphic) fractional-linear transformation, so in summary we have that Cardy's partition function is invariant under the full automorphism group of SL(2,$\mathbb{Z}$), i.e., $\mathcal{Z}_p(A(z))=\mathcal{Z}_p(z)$ where $A\in\text{Aut}\,\text{SL}(2,\mathbb{Z})$. Since J is fairly trivial we shall often take it for granted and refer to the symmetry as just SL(2,$\mathbb{Z}$).

III. PHASE DIAGRAM AND EFFECTIVE ACTIONS

If we assume a unique phase transition when θ is small or vanishes, then the only phase diagram compatible with the symmetry Aut SL(2,$\mathbb{Z}$) is the one exhibited in Fig. 1. As shown in Refs. 10 and 11, this appears to be the case for the $\mathcal{Z}_2$ and $\mathcal{Z}_3$ models, which are equivalent to two coupled Ising models and the three-state Potts model, respectively. In the former case the model is related to the eight-vertex model and may be compared with Baxter's exact solution.[18] It was also shown that the ground state of each phase, which we can label by the unique fraction p/q to which $\theta/2\pi$ converges in the strong coupling limit ($\text{Im}z\simeq g^{-2}\to 0$), can be regarded as a condensate of electric and magnetic charges in the ratio $-p/q$. Furthermore, only excitations with the same charge ratio as the ground state can appear as physical particles with short-range interaction. All other excitations incommensurate with the "vacuum charges" are confined by linear potentials, and are therefore not in the spectrum.

This is highly reminiscent[14] of the physical picture underlying the hierarchy scheme.[4] Each level in the hierarchy is regarded as a condensate of "elementary" excitations, quasielectrons and quasiholes, of the ground state of the previous level. By guessing the corresponding wave functions which, within a variational scheme, can be seen to be extremely good approximations of the true ground states, the effect of each phase transition (condensation) on the transverse conductivity σ_{xy} (which is related to the number ν of fractionally filled Landau levels by $\sigma_{xy}=\nu e^2/h$) can be derived. The resultant states are made up of electric and magnetic degrees of freedom with vorticity-to-charge ratio given by $\nu^{-1}-1$ and suggests the identification $(\theta/2\pi)^{-1}=\nu^{-1}-1$ in the strong-coupling limit.[14] When the quasihole excitations of a state responsible for a plateau in the conductivity condense, the change in σ_{xy} is some power of the transformation A: $\sigma_{xy}\to\sigma_{xy}/(1+2\sigma_{xy})$. Together with particle-hole duality B: $\sigma_{xy}\to 1-\sigma_{xy}$, this generates the whole hierarchy. Note that A and B are real fractional-linear transformations which generate a *proper subgroup* G of Aut SL(2,$\mathbb{Z}$). G is not in SL(2,$\mathbb{Z}$) because particle-hole duality involves the time-reversal symmetry J.

We see that the hierarchy scheme is both physically and mathematically similar to Cardy's self-dual models, but several points need to be addressed. Clearly, G preserves the "oddness" of the denominator of the filling fraction so, if the starting value of the iterations is odd, then the hierarchy is odd. However, the phase diagram of SL(2,$\mathbb{Z}$) also admits even-denominator fractions. This is not inconsistent, because if we restrict attention to very strong magnetic fields so that the spin degrees of freedom are frozen out, then Laughlin's analysis suggests that states with even-denominator fractions cannot be formed starting with a system of just fermions. This means that the initial conditions following from a fermionic system will necessarily be in an odd phase, and the system must remain in that phase under the renormalization-group (RG) flow that determines the macroscopic properties of the system. Thus, in strong magnetic fields, the phase diagram of Fig. 1 is consistent with G since the even self-dual phases are inaccessible to a fermionic system.

A more difficult question is why, in trying to model the quantum Hall system, we should consider self-dual models at all. At the moment we can only offer some speculations. The relevant observation may be that Baxter's exact solution of the eight-vertex model[18] shows that critical points lie in the self-dual plane. Zamolodchikov and Fatteev[19] have conjectured that the same will be true in all $\mathbb{Z}_N\times\mathbb{Z}_N$ theories. Thus, starting from a theory without duality, the RG flow may be expected to take the system at criticality to one of the self-dual models, giving rise to the phase diagram of Fig. 1. Furthermore, since the critical theories are conformally invariant, the system may be driven to low values of p in accordance with Zamolodchikov's c theorem,[20] since the lowest value of the central charge for the $\mathcal{Z}_p$ models is $c=1$ when $p=2$. A rigorous discrete symmetry of the system, perhaps associated with the crystal structure, may prevent us from reaching the $\mathcal{Z}_2$ model at the bottom, so we do not want to dismiss $\mathcal{Z}_3$ out of hand.

To summarize, we have seen that the phase diagram of Cardy's self-dual spin models is consistent with the hierarchy, and that the transverse conductivity σ_{xy} is related to $\theta/2\pi$. In the plateau region only σ_{xy} is nonzero, because the Fermi energy lies in the region of localized states for which $\sigma_{xx}=0$. In the transition region between plateaus, σ_{xx} may be nonzero when the Fermi energy crosses that of an extended state, corresponding to the disappearance of the mass gap. The appearance of such an extended state is consistent with the Coleman-Mermin-Wagner theorem[21] which prohibits Goldstone modes in two dimensions, provided it corresponds to a critical point rather than a spontaneously broken phase. At the critical point the symmetry breaking associated with a (relevant) operator vanishes, leading to an enlarged symmetry and a massless state.[22] This (de)localization phenomenon has been studied theoretically using a Gaussian random potential to describe the effects of the impurities responsible for the inelastic scattering which underlies the dissipative conductivity σ_{xx}. Using the replica trick[8] or supersymmetry,[23] an effective Lagrangian describing the behavior of the averaged electron propagation, and hence the Hall conductances, may be determined. This leads to an effective-Lagrangian description of the form

$$\mathcal{L}_{\rm eff}(\sigma_{xx},\sigma_{xy})=\sigma_{xx}\mathcal{L}_{\rm kin}+\sigma_{xy}\mathcal{L}_{\rm top} , \quad (8)$$

where $\mathcal{L}_{\rm kin}$ is the kinetic piece of the action and $\mathcal{L}_{\rm top}$ is a topological term. The properties at the critical point are largely determined by the symmetries of the effective Lagrangian, and here we investigate the implications if the renormalized form possesses the duality symmetry discussed above. $\mathcal{L}_{\rm eff}$ is of the same general form as (1) [after an SL(2,$\mathbb{Z}$) transformation to make $\theta/2\pi=\sigma_{xy}$], provided that we can regard the transport coefficient σ_{xx} as the effective coupling constant $2\pi/pg^2$ of the model. It is then clear from what has been said above that the partition function $\mathcal{Z}_{\rm eff}$ determined by $\mathcal{L}_{\rm eff}$ is invariant under the modular group acting on the *complexified* conductivity parameter

$$\sigma=\sigma_{xy}+i\sigma_{xx} , \quad (9)$$

i.e., $\mathcal{Z}_{\rm eff}(A(\sigma))=\mathcal{Z}_{\rm eff}(\sigma)$ where $A\in\mathrm{Aut\,SL}(2,\mathbb{Z})$.

As a kind of consistency check on this approach, we could ask how the effective Lagrangian would appear if instead of the conductivities we chose to parametrize $\mathcal{L}_{\rm eff}$ by the resistivities ρ_{ij} $(i,j=x,y)$ which, by definition

$$\rho_{ij}=(\sigma^{-1})_{ij} \quad (i,j=x,y) , \quad (10)$$

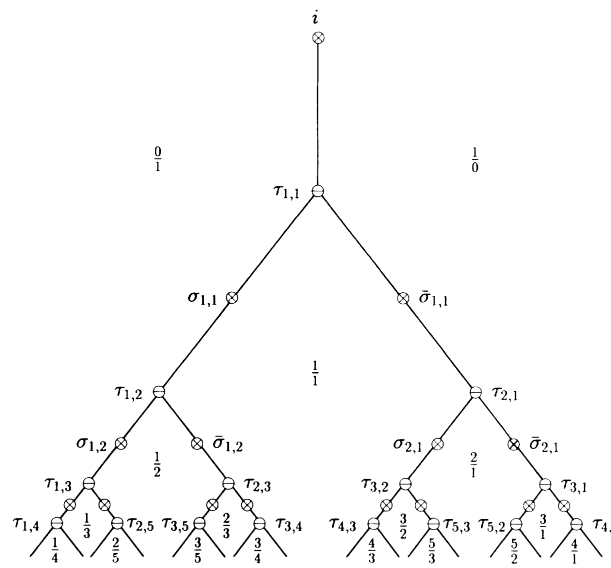
form a completely equivalent set of transport coefficients. Recall first that the Onsager relations for a homogeneous and isotropic medium in an external magnetic field (which breaks parity) imply that $\sigma_{yy}=\sigma_{xx}$ and $\sigma_{yx}=-\sigma_{xy}$, and similarly for ρ_{ij}.[24] This is why the theory depends only on one complex parameter σ, and it is equally natural to complexify the resistivities $\rho=\rho_{xy}+i\rho_{xx}$. By virtue of (10), ρ and σ satisfy

$$\rho=S(\sigma)=-\frac{1}{\sigma} . \quad (11)$$

Since we have already demanded that the theory be self-dual, it follows that ρ and σ have the same phase diagram, shown in Fig. 1.

IV. SL(2,$\mathbb{Z}$)

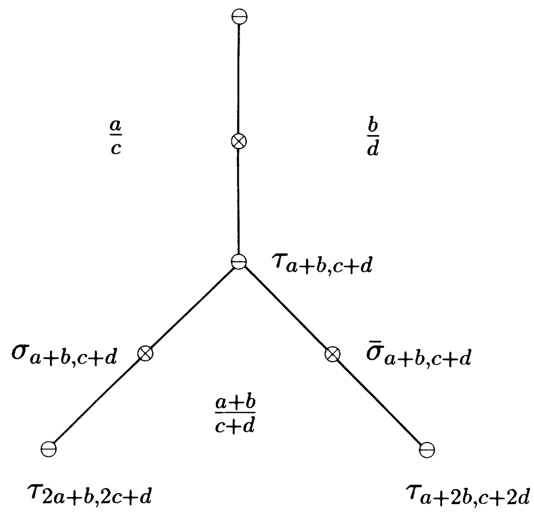
The phase diagram is determined by the "tree" of phase boundaries and it is perhaps surprising that the *only* tree that is invariant under SL(2,$\mathbb{Z}$) is the one shown in Fig. 1.[25] Translation invariance (T) ensures that it is sufficient to consider only the region $0\le\sigma_{xy}<1$, but the unification of integer and fractional effects enforced by SL(2,$\mathbb{Z}$) is best appreciated by drawing the SL(2,$\mathbb{Z}$) tree as in diagram 1, which is faithful only to the topology of the phase diagram.



Each phase is labeled by the fractional value of σ_{xy} to which the phase converges when $\sigma_{xx}\to 0$. The left-right symmetry of the diagram reflects the duality between integer and fractional phases, which is inherent in this model. We cannot have one without the other.

Furthermore, since we have no freedom in the construction of the SL(2,$\mathbb{Z}$) tree, we should be able to make completely rigid quantitative predictions which, when confronted with sufficiently precise experimental data, should unambiguously verify or falsify the model. The most important information of this type is, of course, the location of the Hall plateaus, i.e., the points on the real axis accessible from any of the phases. In mathematics they are called "parabolic fixed points of SL(2,$\mathbb{Z}$)," and as already mentioned they coincide with the rationals. They will be denoted by the symbol $\oplus$. In addition, the main set of data which must be extracted from the tree is the location of the "nodes" where the tree bifurcates and new phases appear. They are called "elliptic fixed points of SL(2,$\mathbb{Z}$) of order three" (E_3), and we denote them by $\ominus$. Finally, except for a trivial "hyperbolic fixed point" at infinity, the only other class of fixed points of SL(2,$\mathbb{Z}$) are the "elliptic fixed points of order two" (E_2), which we represent by the symbol $\otimes$.

The E_n $(n=2,3)$ fixed points are all images of $\sigma=\exp(i\pi/n)$ under SL(2,$\mathbb{Z}$) transformations, and their coordinates may be computed as follows. Consider the generic bifurcation depicted in diagram 2.



Every phase a/c is separated from a phase b/d to the right by a phase boundary which bifurcates into an additional phase $a/c \oplus b/d = (a+b)/(c+d)$ at a node labeled by $\tau_{a+b,c+d}$. If we let g and $\bar{g}$ denote the linear-fractional transformations

$$g(z) = \frac{a+(a+b)z}{c+(c+d)z}, \quad \bar{g}(z) = \frac{(a+b)+bz}{(c+d)+dz}, \tag{12}$$

then the fixed points associated with the new phase are given by

$$\sigma_{a+b,c+d} = g(i), \quad \bar{\sigma}_{a+b,c+d} = \bar{g}(i) \quad \in E_2(\otimes),$$
$$\tau_{2a+b,2c+d} = g(j), \quad \tau_{a+2b,c+2d} = \bar{g}(j) \quad \in E_3(\ominus), \tag{13}$$

where $j = \exp(i\pi/3)$. Clearly, by iterating this algorithm we can recursively generate the whole hierarchy of phases in diagram 1. For example, the three fixed points of SL(2,$\mathbb{Z}$) associated with the qth phase of the principal sequence[3] of fractional states with filling factors $1/q$ are located at

$$\tau_{1,q} = \frac{j}{1+(q-1)j}, \quad \sigma_{1,q} = \frac{i}{1+qi},$$
$$\bar{\sigma}_{1,q} = \frac{1+i}{q+(q-1)i}. \tag{14}$$

The positions of the fixed points of the charge-conjugated sequence of phases $1-1/q$ are obtained by simply reflecting these points in the line $\mathrm{Re}\sigma = \frac{1}{2}$. We also record the location of the fixed points in the "secondary sequence" of phases, which are labeled by the fractions $k/(2k+1)$ $(k=1,2,\ldots)$,

$$\tau_{k,2k+1} = \frac{(k-1)+j}{(2k-1)+2j},$$
$$\sigma_{k,2k+1} = \frac{(k-1)+ki}{(2k-1)+(2k+1)i}, \tag{15}$$
$$\bar{\sigma}_{k,2k+1} = \frac{k+i}{(2k+1)+2i},$$

since these will be of particular interest when we turn to a comparison with experiment.

V. RENORMALIZATION GROUP FLOW AND IMPLICATIONS FOR THE QUANTUM HALL EFFECT

Up to now we have considered the implications of SL(2,$\mathbb{Z}$) only for the phase diagram, but it also gives us information about the RG flow which determines the macroscopic properties of the system. Since phase boundaries are RG flow lines and the appearance of additional phases is controlled by fixed points of the RG, the scaling digram must coincide with the tree of Fig. 1. Every fixed point of SL(2,$\mathbb{Z}$) must be a fixed point of the RG, but the converse is not necessarily true. Thus, at the very least, there must be RG fixed points at the nodes (the triple points E_3) and the other elliptic fixed points of order two (E_2), and the parabolic fixed points on the real line. In order to explain the quantization of the plateaus the latter must be the only stable fixed points, since the RG flow drives the system to the strong-coupling limit at which $\sigma_{xx}=0$ and σ_{xy} is rational (these are the only points accessible starting from any of the phases above the strong coupling limit). The structure of the rest of the RG flow depends on the character of the remaining fixed points and on whether there are RG fixed points additional to those of SL(2,$\mathbb{Z}$).

The $\mathcal{Z}_3$ model is consistent with the minimal choice of fixed points given above, because it is known to have an isolated repulsive fixed point at i, so it follows in this case that the E_2 fixed points are repulsive. This, in turn, implies that the E_3 fixed points are (three-way) saddle points, assuming no additional fixed points. The $\mathcal{Z}_2$ model may be completely solved because it is related[11] to the symmetric eight-vertex model solved by Baxter.[18] It has a critical *line* along the fundamental phase boundary between i and j corresponding to the addition of an infinite number of RG fixed points. Finally, there is another simple possibility (not realized by Cardy's models) in which the E_3 fixed points are repulsive and the E_2 fixed points are ordinary saddle points. With this latter identification the flow lines of the SL(2,$\mathbb{Z}$)-invariant diagram are as shown in Fig. 2, where the broken lines are flows which end up at the attractive (real) fixed points. The $\mathcal{Z}_2$ and $\mathcal{Z}_3$ flow diagrams are similar, but in the latter case the arrows on the tree must be reversed, while in the former case no arrows can be assigned to these lines at all since they are marginal directions. The broken flow lines are always the same.

The scaling diagram in Fig. 2 is consistent with the observed hierarchy of fractional phases.[2,4] By tuning the experimental parameters, which include the magnetic field, temperature, contamination, and other microscopic quantities specific to the particular sample being used, we can dial up specific values in the σ plane from which to start the RG flow. The system will flow to the *unique* stable fixed point $(p/q,0)$ which belongs to the same phase as the starting value of σ. In very strong magnetic fields all electron spins are aligned with the field, so that the Fermi-Dirac statistics of the electrons prevent any

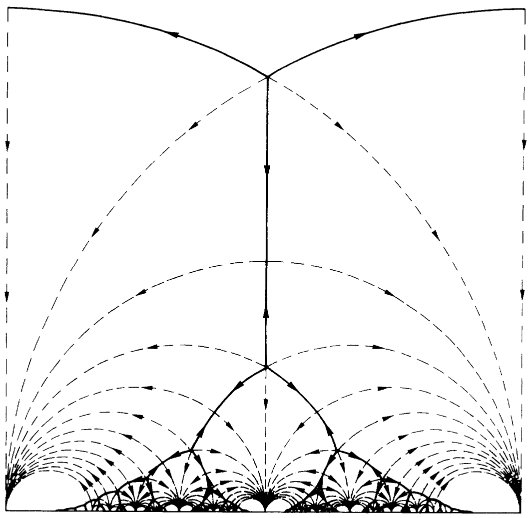


FIG. 2. The simplest assignment of fixed points and flow lines to the SL(2,$\mathbb{Z}$) invariant scaling diagram.

even wave function, appropriate for even-denominator phases by the standard argument of Laughlin,[3] from appearing. If such a system is tuned so that the initial values of σ_{xx} and σ_{xy} correspond to an even phase in Fig. 1, then the large energy associated with these (spin-flipped) phases will mean that the system prefers to be in a nondual phase not represented in this two-parameter scaling theory. However, the RG flow may be expected to take the system to critical points associated with odd phases of the self-dual plane described by Fig. 1. This is why it is usually odd values of q which are observed. At lower values of the magnetic field, where spin flips are no longer disallowed by energetics, even phases should, according to this argument, appear and some have been observed.[26]

The microscopic physics of the nondual initial configurations will similarly determine within which fixed point's domain of attraction a (dirty) configuration with large initial σ_{xx} lies. If the system is in the domain of attraction of the hyperbolic fixed point at $i\infty$, then no quantum Hall effect will be observed. Alternatively, if the microscopic physics is such that a large σ_{xx} is in the domain of attraction of one of the integer fixed points, then the integer Hall effect will still be observed for large initial values of σ_{xx}. Lacking a derivation of the self-dual effective action from a microscopic one describing electrons in a transverse magnetic field, we cannot determine which possibility is relevant.

As the starting value of σ_{xx} is pushed closer to zero, by improving the purity of the sample, say, more and more physically distinct phases become available, and in the order required by the hierarchy scheme.[4] This "nesting" of plateaus was quickly recognized as a desirable feature of a scaling theory of the quantum Hall system, and is therefore built into the scaling diagram proposed in Ref. 27, which was inspired by the integer scaling theory[28] based on the replica σ model.[29] It is gratifying to have it forced upon us by a simple discrete symmetry.

The tree structure of our RG flow diagram is different from that proposed in Ref. 27, where the branch points are trifurcations rather than bifurcations, but appears to be the only one consistent with duality. The main difference between the two is the appearance of even phases, including the 1/0 phase, in the self-dual digram. This raises the question of whether we can maintain the physical interpretation, proposed in Ref. 27, of the mobility fixed points as gap-closing transitions, i.e., as transitions in which increasing disorder destroys a phase. This "dirty limit" implies the overlapping of two of the Landau bands in the sample, so that two extended states merge into one. Hence, a Hall plateau should disappear, and this should correspond to the disappearance of a phase in the scaling diagram. Since the appearance of new phases is controlled by the mobility fixed points $\ominus$, we should be able to interpret these fixed points as gap-closing transitions. The scaling diagram in Ref. 27 was constructed by hand, so that this is obviously the case. The question is whether this interpretation can be maintained for the self-dual diagram, and we see no reason why this should not be the case. If the magnetic field is sufficiently strong to exclude the even regions, i.e., if it is not energetically favorable for the system to access the self-dual plane, then it might seem difficult to associate the closing of a mobility gap with a $\ominus$ fixed point, on the $\sigma_{xy}=\frac{1}{2}$ line, say, unless the $\frac{1}{3}$ and $\frac{2}{3}$ phases share a common border. But again, there is no contradiction since it is quite possible for two odd phases to be "nearest neighbors" in a three-dimensional flow diagram of which we have only obtained the self-dual slice. And again we cannot address this question in any detail because we lack the necessary microscopic information. Furthermore, in weaker magnetic fields the roles are reversed; while our diagram immediately admits $\ominus$ as a gap-closing transition, there is apparently no room for such phases at all in the diagram of Ref. 27.

The situation is summarized in the somewhat heuristic three-dimensional diagram in Fig. 3, which shows how the peaks, plateaus, and inflection points of the "experimental" graphs $\rho_{xx}(B)$ and $\rho_{xy}(B)$ are collated by the SL(2,$\mathbb{Z}$) tree. While the ρ plane encodes the "universal" (material-independent) aspects of the quantum Hall system, the widths of the plateaus depend on the sample, so that the scale on the B axis is arbitrary. In principle, $\rho_{xy}(B)$ is a Devil's staircase, but only a few (odd) plateaus are shown, corresponding to an experiment with finite resolution in a very strong magnetic field.

This diagram also explains the observation reported in Ref. 30, which was one of the main motivations for undertaking the work described here. In Ref. 30 it was shown that Laughlin's wave functions appear effortlessly in a second quantized (field-theoretic) formalism based on conformal invariance, which accounts both for the simple and elegant form of these wave functions, as well as their astonishing accuracy on the plateaus. In our scaling theory, these obviously correspond to attractive fixed points ($\oplus$). Since quantum-field theories are conformal at fixed points of the RG it should, therefore, be possible to recover Laughlin's wave functions directly from the action (1), by computing expectation values of strings of vertex operators which describe the quasiparticles in this

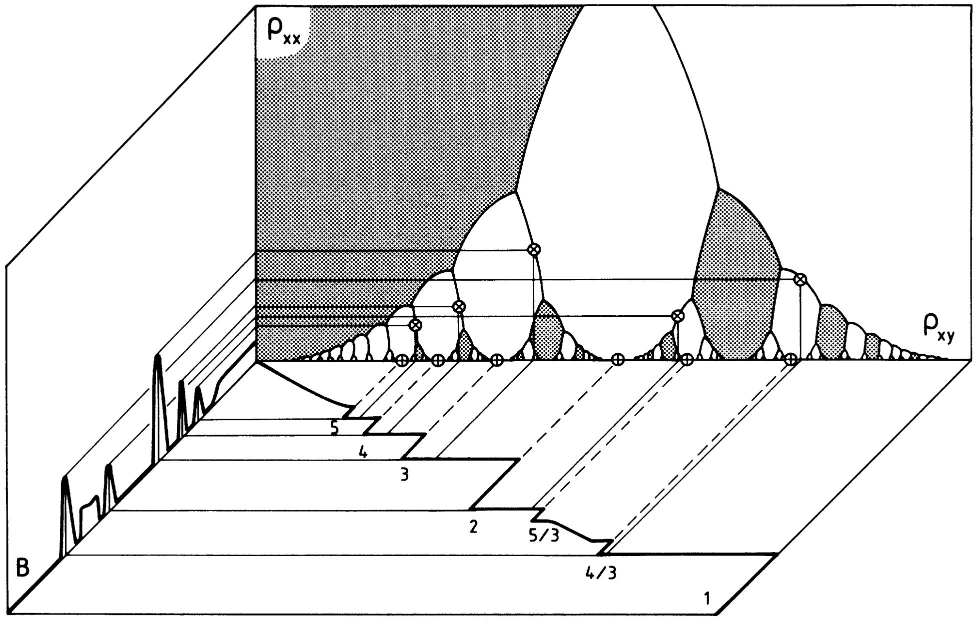


FIG. 3. Resistivities ρ_{xx} and ρ_{xy} plotted against the external magnetic field B, and compared to the SL(2,ℤ) tree, which encodes universal features of the quantum Hall effect.

theory.[31]

The transition between any two neighboring phases, integer or not, is controlled by the same *type* of fixed point. Because the RG equations describing the flow in the neighborhood of any two such points are related by SL(2,ℤ), they must determine the same critical exponents. So, while there are an unlimited number of physically distinct phases, the properties of the fixed points controlling the appearance of a new phase are always the same. (This may, perhaps, be called universality,[32] although this breaks with common usage of the word.) Hence, in this model the critical exponents must be the same for all Hall steps, as is indeed observed[5] in all the transitions studied so far.

We would also like to check the location of the fixed points in the σ plane which were computed above, but a reliable experimental determination of this information is complicated by our lack of knowledge about microscopic properties of the quantum Hall system. Since what can be measured is the flow of an initial point in the σ plane as a function of temperature, we need to translate the scaling equations into temperature-driven flows.[33,32] This is in principle possible because the effective sample size (the renormalization scale) is related to the inelastic scattering length, which can be increased by decreasing the temperature (Thouless scaling).[34] Hence, the measurement of $(\sigma_{xy},\sigma_{xx})$ for a succession of decreasing temperatures corresponds to a measurement of $(\sigma_{xy},\sigma_{xx})$ as the system follows a RG flow line.

We have found two sets of data which appear to be sufficiently accurate to make a comparison with the model. The initial scaling experiment[33] demonstrated rather strikingly that so-called "delocalization" fixed points associated with the transition between integer levels are located at $\sigma_{xy}=n+\frac{1}{2}$ $(n=0,1,2,\dots)$, in agreement with the scaling diagram in Fig. 2. Furthermore, it is apparent from the flow diagram in Fig. 3 of Ref. 33 that once the system has entered deeply into the quantum domain, so that semiclassical effects no longer distort the RG flow (solid lines in Ref. 33), then not only do the directions of the flow lines coincide with our Fig. 2, but they also appear to originate at $\sigma_{xx}=\frac{1}{2}$. If this is really the case, then the delocalization fixed points between integer phases coincide with the E_2 fixed points labeled $\sigma_{n,1}$ $(n=1,2,\dots)$ in diagram 1.

Similar temperature-driven scaling experiments have also been carried out in some of the fractional phases.[35] However, the statistics are not sufficient to provide flow diagrams of the same quality as for the flow between integer levels. The most accurately determined quantities in this case are the values of σ_{xx} at which the $\frac{1}{3}$, $\frac{2}{5}$, $\frac{3}{7}$, $\frac{4}{9}$ (and possibly also the $\frac{5}{11}$) phases first appear. Since these data presumably locate the height at which the "tip" of a fractional phase is located, we interpret these σ_{xx} values as the imaginary parts of the "mobility" (E_3) fixed points associated with the secondary sequence.

Table I compares the experimental data reported in Ref. 35 with the theoretical values $\tau_{k,2k+1}$ $(k=1,2,3,4,5)$ recorded above in (15). We see that while the values agree within the accuracy of the experiment, the theoretical values appear to be slightly but systematically lower than the measured values. This is not surprising, because the measurements were performed at $\sigma_{xy}=\frac{1}{2}$. We therefore do not expect their results to coincide exactly with any of the nearest-neighbor E_3 fixed points, but they should get closer for higher k, because Re $\tau_{k,2k+1}\rightarrow\frac{1}{2}$ when k grows, and indeed they do. The comparison could be improved by repeating the experimental determination of the mobility fixed points at the appropriate

TABLE I. Experimental data σ_{xx}^{MFP} on the location of mobility fixed points, obtained in Ref. 35 by studying temperature-driven flows in the σ plane at $\sigma_{xy}=\frac{1}{2}$, compared with the imaginary part $\sigma_{xx}^{E_3}$ of the E_3 fixed point of SL(2,$\mathbb{Z}$) associated with the appearance of the new phase. The errors are estimated in Ref. 36.

$\frac{k}{2k+1}$	σ_{xx}^{MFP}	$\sigma_{xx}^{E_3}$
$\frac{1}{3}$	0.15 ± 0.03	0.1237
$\frac{2}{5}$	0.055 ± 0.01	0.0456
$\frac{3}{7}$	0.025 ± 0.005	0.0221
$\frac{4}{9}$	0.013 ± 0.003	0.0129
$\frac{5}{11}$	$0.01\pm?$	0.0084

values of σ_{xy}.

Although the authors of Ref. 35 also attempted to locate the delocalization fixed points, their criteria appear to be somewhat arbitrary and the statistics seem inadequate at present to resolve the details of the scaling diagram. We also need a better understanding of the way in which the RG flow in very strong magnetic fields is pushed out of the (energetically) forbidden even denominator regions, particularly for transitions that cross even-denominator fractions.

VI. SUMMARY

In summary, by postulating that the quantum Hall system is invariant under modular (duality) transformations, we are led to a phase diagram and an associated RG flow diagram which are consistent with current observations of both the integer and the fractional quantum Hall effect. An important feature of this symmetry is that the critical exponents associated with transitions between integer and between fractional levels should all be the same, in agreement with existing observations. This is achieved while maintaining the original hierarchy generation scheme of Laughlin, Haldane, and Halperin. Determination of the critical exponents requires the identification of a particular model within the class possessing SL(2,$\mathbb{Z}$) symmetry.[31] The duality hypothesis is also in qualitative agreement with the observed temperature dependence of the transition between integer levels. Moreover, the predicted appearance of new fractional phases is in quantitative agreement with experiment.

ACKNOWLEDGMENTS

C.A.L. wishes to thank the Norwegian Research Council for Science and the Humanities (NAVF) for financial support, and C. Series and T. Skjelbred for discussions on SL(2,$\mathbb{Z}$). We are also grateful to I. Aitchison and J. Zuk for discussions.

[1] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
[2] D. C. Tsui, H. L. Störmer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).
[3] R. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).
[4] F. D. M. Haldane, Phys. Rev. Lett. **51**, 605 (1983); B. Halperin, *ibid.* **52**, 1583 (1984).
[5] H. P. Wei, D. C. Tsui, M. A. Paalanen, and A. M. M. Pruisken, Phys. Rev. Lett. **61**, 1294 (1988); L. Engel, H. P. Wei, D. C. Tsui, and M. Shayegan, in *Proceedings of the Eighth International Conference on Electronic Properties of Two-Dimensional Systems, Grenoble, France, 1989* [Surf. Sci. **229**, 13 (1990)].
[6] R. B. Laughlin, in *The Quantum Hall Effect,* 2nd ed., edited by R. E. Prange and S. M. Girvin (Springer-Verlag, Berlin, 1990), and references therein.
[7] J. K. Jain, S. A. Kivelson, and N. Trivedi, Phys. Rev. Lett. **11**, 1297 (1990).
[8] A. M. M. Pruisken, in *The Quantum Hall Effect* (Ref. 6).
[9] A. Khurana, Phys. Today **43** (12), 17 (1990), and references therein.
[10] J. L. Cardy and E. Rabinovici, Nucl. Phys. B **205**, 1 (1982).
[11] J. L. Cardy, Nucl. Phys. B **205**, 17 (1982).
[12] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[13] S. M. Girvin and A. H. MacDonald, Phys. Rev. Lett. **58**, 1252 (1987).
[14] A. Shapere and F. Wilczek, Nucl. Phys. B **320**, 669 (1989).
[15] S.-J. Rey and A. Zee, Nucl. Phys. B **352**, 897 (1991).
[16] G. 't Hooft, Nucl. Phys. B **190**, 455 (1981).
[17] R. Dijkgraaf, E. Verlinde, and H. Verlinde, in *Perspectives in String Theory, Copenhagen, 1987,* edited by P. DiVecchia and J. L. Peterson (World Scientific, Singapore, 1988).
[18] R. J. Baxter, Phys. Rev. Lett. **26**, 832 (1971).
[19] A. B. Zamolodchikov and V. A. Fatteev, Zh. Eksp. Teor. Fiz. **89**, 380 (1985) [Sov. Phys. JETP **62**, 215 (1985)].
[20] A. B. Zamolodchikov, Pis'ma Zh. Eksp. Teor. Fiz. **43**, 565 (1986) [JETP Lett. **43**, 730 (1986)].
[21] S. Coleman, Commun. Math. Phys. **31**, 259 (1973); N. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).
[22] I. Affleck, Nucl. Phys. B **265**, 409 (1986), and references therein.
[23] H. A. Weidenmüller, Nucl. Phys. B **290**, 87 (1987); H. A. Weidenmüller and M. R. Zirnbauer, *ibid.* **305**, 339 (1988).
[24] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media,* Course of Theoretical Physics Vol. 8 (Pergamon, Oxford, 1982).
[25] J.-P. Serre, *Trees* (Springer-Verlag, Berlin, 1980) [translated from the original French edition: *Arbres, Amalgames, SL_2* (Society of Mathematics, France, 1977)], Astérisque No. 46.
[26] R. G. Clark, R. J. Nicholas, J. R. Mallett, A. M. Suckling, A. Usher, J. J. Harris, and C. T. Foxon, in *Proceedings of the 18th International Conference on Physics of Semiconductors, Stockholm, 1986,* edited by O. Engström (World Scientific, Singapore, 1987), p. 393; R. Willett, J. Eisenstein, H. Störmer, D. Tsui, A. Gossard, and J. English, Phys. Rev. Lett. **59**, 1776 (1987).
[27] R. B. Laughlin, M. L. Cohen, J. M. Kosterlitz, H. Levine, S. B. Libby, and A. M. M. Pruisken, Phys. Rev. B **32**, 1311 (1985).
[28] D. E. Khmel'nitzkii, Pis'ma Zh. Eksp. Teor. Fiz. **38**, 454 (1983) [Sov. Phys. JETP Lett. **38**, 552 (1983)]; H. Levine and S. B. Libby, Phys. Lett. B **150**, 182 (1985); A. M. M. Pruisken, Phys. Rev. B **32**, 2636 (1985).

[29]H. Levine, S. Libby, and A. M. M. Pruisken, Phys. Rev. Lett. **51**, 1915 (1983); Nucl. Phys. B **240**, 30 (1984); **240**, 49 (1984); **240**, 71 (1984); A. M. M. Pruisken, *ibid.* **235**, 277 (1984).
[30]S. Fubini and C. A. Lütken, Mod. Phys. Lett. A **6**, 487 (1991).
[31]C. A. Lütken and G. G. Ross (unpublished).
[32]A. M. M. Pruisken, Phys. Rev. Lett. **61**, 1297 (1988).
[33]H. P. Wei, D. C. Tsui, and A. M. M. Pruisken, Phys. Rev. B **33**, 1488 (1985).
[34]D. J. Thouless, Phys. Rev. Lett. **39**, 1167 (1977).
[35]J. R. Mallett, R. G. Clark, J. J. Harris, and C. T. Foxon, in *High Magnetic Fields in Semiconductor Physics II,* edited by G. Landwehr (Springer-Verlag, Berlin, 1989), p. 132.
[36]J. R. Mallett, doctoral dissertation, Oxford University, 1989.

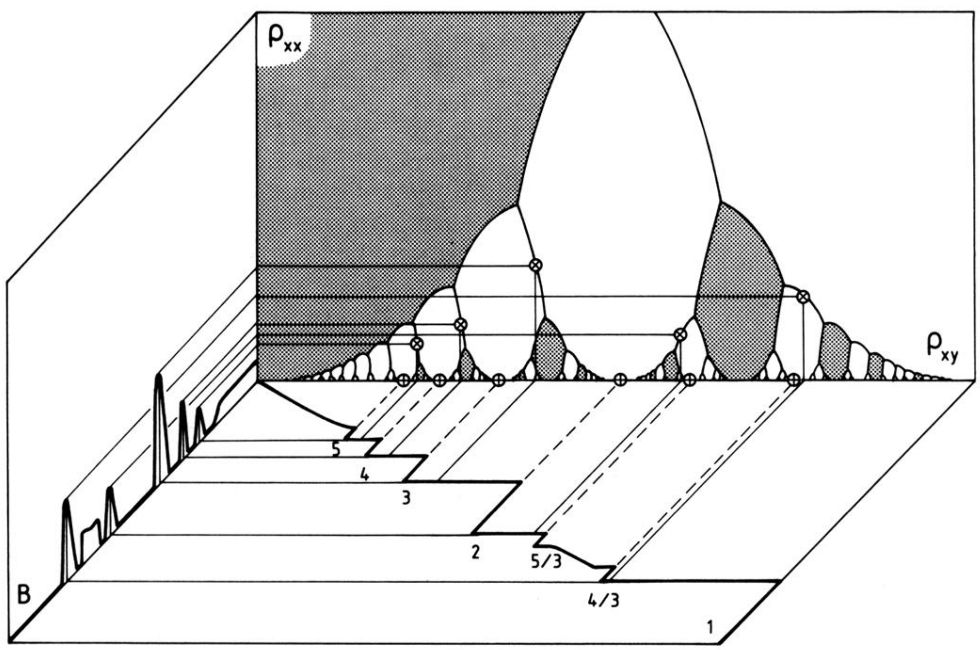


FIG. 3. Resistivities ρ_{xx} and ρ_{xy} plotted against the external magnetic field B, and compared to the SL(2,$\mathbb{Z}$) tree, which encodes universal features of the quantum Hall effect.