Fractional quantum Hall effect and quantum symmetry

G. Grensing
Institut für Theoretische Physik und Astrophysik, Universität Kiel, D-24118 Kiel, Germany
(Received 15 July 1998; revised manuscript received 1 September 1999)

Laughlin's quantum Hall wave function is obtained as an exact ground state of an N-particle Hamiltonian, in which the electrons themselves are coupled to the Chern-Simons field. The proof requires quantum field-theoretic methods in the Schrödinger picture, and clearly exhibits the relation to conformal field theory since the Knizhnik-Zamolodchikov connection shows up and the Laughlin ground state is recognized as a conformal block. A refined version of this approach can be applied for toroidal boundary conditions, i.e., the electrons are confined to the physical interior of the Hall sample. The result turns out to be a nontrivial modification of the p-fold degenerate Haldane-Rezayi wave function. Furthermore, it is shown that the degeneracy of the ground state is accounted for by the quantum group $U_q(su_2)$ with deformation parameter $q=\exp(2\pi i/p)$ and p odd; put differently, the possible quantum numbers p of the quantum Hall effect, determining the filling fractions $\nu=1/p$, can be explained by an underlying quantum symmetry.

INTRODUCTION AND SUMMARY

The quantum Hall effect[1] is one of the most fascinating phenomena which have been discovered in the last two decades.[2] But it has also proven to be one of the most difficult to understand. In addition, almost all of the more recent important developments are expected to find an application in this context.

Let us give some examples to certainly illustrate what we have in mind. First, conformal field theory[3] has its impact; in particular, the Laughlin ground state[4] can be constructed by means of vertex operator methods,[5] even though there is no explanation why these constructions do work at all. Second, Laughlin's excitations have been shown[6] to carry fractional spin and statistics, and their anyonic behavior is accounted for by the ''statistical'' Chern-Simons gauge field.[6,7] Third, infinite-dimensional Lie algebras[8] such as Kac-Moody and Virasoro algebras are involved through edge excitations,[9] so that chiral Wess-Zumino-Witten models[10] should be of relevance as well; as opposed to the bulk system, for the edge states the impact of conformal field theory is rather well founded. Fourth, according to Witten,[11] Chern-Simons theory[12,13] is deeply related to knot theory;[14] beyond this, knot theory has been revolutionized by quantum group-theoretical concepts (Refs. 15 and 16; cf. also Ref. 17). Hence one could also expect a quantum symmetry to be involved in this context; a first hint comes from the surprising observation[18] that the rather antique Landau theory in two dimensions carries a quantum group structure.

In spite of all these insights or conjectures, however, there is no conclusive answer to the one basic question of why Laughlin's celebrated trial ground state seems to be almost exact.

There have been several attempts to reach a more detailed understanding of Laughlin's variational guess, which can be motivated by following the analogy with superfluidity.[19] Numerical studies for small numbers of electrons have shown[4,20] that Laughlin's wave function is a nearly perfect ground state, being largely independent of the detailed form of the electron interaction. Furthermore, Haldane[21] has constructed a class of Hamiltonians for which Laughlin's trial wave function is exact; such an observation was also made by Trugman and Kivelson[22] for a model with short-range interaction. In the field-theoretic context, the bosonized version of Laughlin's wave function[23] was obtained by Kane *et al.*[24] and Karlhede *et al.*[25] from the effective Ginzburg-Landau theory[23,26,25] in the Gaussian approximation. In a similar vein, Lopez and Fradkin[27] found the modulus squared of the Laughlin state in the long-wavelength approximation of the fermion Chern-Simons theory[28] in which, following Jain,[29] an even number of flux quanta is attached to the electrons in order to relate the fractional to the integer quantum Hall effect. In another approach, using methods of collective field theory, Sheng *et al.*[30] also gave a derivation of the Laughlin wave function.

We take another route to answer the question about the origin of Laughlin's trial wave function. The point of departure resembles the fermion theory in that we couple the electrons themselves to the Chern-Simons connection. The reason is that in two dimensions they must necessarily follow braided paths. Hence, the constituent particles must also be subjected to the Chern-Simons field, whereas until now only the excitations have been accepted to ''feel'' the statistical interaction. We differ from the fermion field theory[28] in that an odd number of flux quanta is bound to the electrons; we could also take the approaches of Kane *et al.*[24] and Karlhede *et al.*[25] as a starting point, but we avoid to apply the questionable singular gauge[23] and leave it with the original electrons. Furthermore, we neglect the repulsive electron interaction, and so one might object that we arrive at what is called the anyon gas. This model has been devised to describe the excitations of the quantum Hall system, but it is also expected to play a decisive role in the context of high-temperature superconductivity and has been intensively studied over the years.[31] Our approach, however, differs in one crucial aspect from the anyon gas because we do not eliminate the ''redundant'' Chern-Simons connection beforehand, at the expense of obtaining a rather untractable N-particle Schrödinger equation with the nonlocal statistical interaction, for which an exact solution seems to be out of reach. Instead, we treat the Chern-Simons field as a degree of freedom which must fully be quantized from the outset. The reason is that one is deep down in the intricacies of constrained systems.[32] Though the Chern-Simons field which is involved here is Abelian so that the machinery of Becchi-Rouet-Stora quantization is not needed, nevertheless there are two subtle points which require special care. First, the Chern-Simons action is a Hamiltonian first-order action[33] in which the spatial components $A=(A_1,A_2)$ of the connection form a symplectic vector; if one identifies A_1 as a coordinate and A_2 as a momentum, rotational invariance is lost. Hence holomorphic quantization appears to be the natural choice. Second, the time component A_0 of the Chern-Simons connection is a Lagrange multiplier, the variation of which yields the constraint, and so one must decide whether one restricts the symplectic phase space at the classical level and quantizes afterwards, or quantizes first and restricts to physical states by the requirement that they are annihilated by the operator constraints.[32] Since these two processes need not commute, the two resulting quantum systems may turn out to be essentially different. The present problem, i.e., two-dimensional electrons in the presence of the Chern-Simons field, can serve as a prime example, where this happens to be the case. The elimination at the classical level before quantizing results in the standard symmetric ''statistical'' connection of the anion gas, whereas the inverse process yields, as we will show, an Abelian Knizhnik-Zamolodchikov connection,[34] which is unsymmetric. Hence the system we investigate differs from the anyon gas in a crucial aspect.

These matters form the content of Sec. I, where it is shown that if we quantize first and constrain afterwards, then techniques from the Schrödinger quantization of quantum field theory[35] make the ground state an exactly calculable quantity, which turns out to be a conformal block and reduces to Laughlin's trial wave function for odd filling fractions. Hence, unlike other alternative frameworks to quantize this theory, the approach chosen in the present paper allows an unambigious determination of Laughlin's wave function as an exact ground state, which here follows for rather ''kinematic'' reasons from the coupling of the electrons to the Chern-Simons connection. The present results also entail that conformal field theory is involved in this context in a prescribed way, whereas the observation that Laughlin's trial wave function can also be obtained by vertex operator techniques[5] is a purely ''experimental'' fact. We conclude Sec. I with a brief discussion of excitations, which emerge in a rather straightforward way in this context.

Of course, the present model is a drastic oversimplification of the quantum Hall system since the repulsive Coulomb interaction and impurities are boldly ignored. The hard problems which remain to be solved are that the formation of the ground state is stabilized by interactions, and that for[4] $\nu \sim 1/70$ [or for $\nu \sim 1/7$ (Ref. 36)] Wigner crystallization sets in.

Another basic question is related to the fact that Laughlin's wave function is defined in the entire plane only, and does not take into account the finite extent of the Hall sample. Profound work on this problem is due to Haldane and Rezayi;[37] these authors imposed toroidal boundary conditions, which is the really natural choice. But a further problem comes in since one expects from general arguments[11,13] that the total charge is bound to vanish on a compact closed surface. Hence it will be essential to overcome the barrier of total charge zero, and we shall see in Secs. II and III that a modification of the degenerate Haldane-Rezayi wave function is required to circumvent this restriction.

The result is obtained on elaborating techniques, developed by Bos and Nair[13] in a different setting. However, in the present context it will be an essential step to avoid a Wilson-like treatment of matter particles, which also obscures that it is indeed a ground state which is determined. Furthermore, a concept is developed to derive the invariance properties of the wave functionals and the invariance conditions to be imposed, being superior to the technique of inspired guess used in the literature.

Thus, also in the present context, the ground state of the quantum Hall effect turns out to be degenerate,[38] so that it is the ground-state degeneracy p which now determines the filling fraction $\nu=1/p$ for p odd. The origin of this fact has always been a matter of debate and attributed to a ''topological order,'' [39] but this is just a name. We show in Sec. IV that the quantum group $U_q(su_2)$ with deformation parameter $q=\exp(2\pi i/p)$ provides an explanation, because the p-fold degenerate ground state yields an irreducible representation of this quantum algebra.

Hence a quantum symmetry yields the organizing principle and can thus be seen—much in the same way as the spin with its underlying ''classical'' su_2 symmetry provides for an explanation of level splittings in atomic spectra—to account for the experimentally observed filling fractions $\nu=1/p$ of the quantum Hall effect, with p an odd integer. Another way to put this outcome into perspective is to draw an analogy with elementary particle theory, where the advent of a new quantum number is always associated with a new or enlarged Lie symmetry. However, in the present case we do not find an ordinary symmetry, but a quantum symmetry emerges. Section V is devoted to some concluding remarks.

I. LAUGHLIN'S WAVE FUNCTION AS AN EXACT GROUND STATE

It is shown that the Laughlin's trial wave function in the plane may be obtained from first principles, being two in number. The first derives from the observation that the space-time trajectories of electrons living in two dimensions must necessarily follow braided paths, because the particle trajectories cannot intersect due to the exclusion principle. Hence the configuration space is multiply connected, its homotopy group being given by the braid group. Path-integral methods then tell us[40] that one has also to sum over all classes of nonhomotopic paths, making itself felt in a phase factor, describing the braiding of the paths. This phase adds as a topological term to the action, which can be rewritten as a line integral over the ''statistical'' gauge field so that the particles now experience a nonlocal interaction. The crucial insight then is that the statistical gauge field can be implemented by the local coupling to the Chern-Simons field (see, e.g., Ref. 41). It is important to stress that this kind of reasoning forbids the presence of a term of Maxwell type; only the Chern-Simons term is induced.

Hence, it is entirely natural and well founded to couple the electrons to the Chern-Simons field, since they must be endowed with the degrees of freedom appropriate to the two-dimensional case. Though this may be seen as a change of attitude because only the excitations are generally accepted to carry fractional statistics, nevertheless we stick to the above assumption and explore its consequences.

Thus, the Lagrangian for two-dimensional nonrelativistic electrons with charges q_n in an external magnetic field B is taken to be

$$L=\sum_{n=1}^{N}\left(\frac{m}{2}\dot{x}_n^i\delta_{ij}\dot{x}_n^j-q_n[\mathring{A}_i(x_n)+A_i(x_n)]\dot{x}_n^i-q_nA_0(x_n)\right)+L_{\mathrm{CS}}, \tag{1.1}$$

where $\mathring{A}_i(x)=-\frac{1}{2}B\varepsilon_{ij}x^j$ with $i=1$ and 2 denotes the classical external gauge field, and (A_0,A) the Chern-Simons gauge field with Lagrangian

$$L_{\mathrm{CS}}=\frac{k}{4\pi}\int d^2x\,\varepsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho=\frac{k}{4\pi}\int d^2x\,\varepsilon^{ij}(\dot{A}_iA_j+A_0F_{ij}). \tag{1.2}$$

The integration extends over a two-dimensional domain Σ which, in the present section, is taken to be the whole plane. On obtaining the second form, it is assumed that the component A_0, which plays a special role, vanishes on the boundary of Σ. Then the surface term arising from a partial integration can safely be ignored.

The Chern-Simons part is a first-order Lagrangian, and so the kinetic term on the right-hand side of Eq. (1.2) gives the symplectic 2-form, which determines the Poisson bracket to be (cf. Ref. 33)

$$\{A_i(x),A_j(y)\}=-\frac{2\pi}{k}\varepsilon_{ij}\delta(x-y) \tag{1.3}$$

whereas the terms proportional to A_0 yield the constraint

$$\frac{k}{2\pi}F_{12}(x)=\rho(x)=\sum_n q_n\delta(x-x_n), \tag{1.4}$$

which is first class in Dirac's [32] nomenclature.

The system described by Lagrangian equations (1.1) and (1.2), and its quantization, has extensively been studied, where the generally applied strategy is to eliminate the gauge connections by means of the constraint;[42,7,43] however, some care is required in applying this elimination procedure, since it must be performed in a gauge invariant way in order to respect the commutation relations of the Chern-Simons field.[44] This approach has also been used to obtain[45] a two-dimensional analog of the Wigner-Jordan transformation.

We proceed differently, and this constitutes the second basic principle: treating the Chern-Simons connection as a quantum field so that the constraint must be imposed as a physical state condition. *A priori*, there are two possibilities: one can either restrict the classical phase space by means of the constraint in advance and quantize afterwards, or one can quantize first and restrict afterwards. The importance of this point has often been stressed[32] since these two acts need not commute,[46] and this will happen to be the case for the system under consideration. As we shall see then, when treating the Chern-Simons field in the Schrödinger picture of quantum field theory,[35] these two basic principles allow for an exact determination of the ground state wave function.

To begin with, we need to make a decision how to quantize the Chern-Simons field. One could select one component of A at will as the generalized coordinate and the remaining one as the corresponding momentum, but it is advantageous to choose holomorphic quantization with[47]

$$\hat{A}_{\bar{z}}=A_{\bar{z}},\quad \hat{A}_z=\frac{\pi}{k}\frac{\delta}{\delta A_{\bar{z}}}. \tag{1.5}$$

The Bargmann inner product for Schrödinger wave functionals $\psi[A_{\bar{z}}](x_1,\ldots,x_N)$ is then

$$\langle\psi_1|\psi_2\rangle=\int d[A_{\bar{z}},A_z]\times\exp\left(-\frac{k}{\pi}\int d^2x\,A_{\bar{z}}A_z\right)\overline{\psi_1[A_{\bar{z}}]}\psi_2[A_{\bar{z}}], \tag{1.6}$$

where, for the time being, the x_n dependence is suppressed.

The constraint, which commutes with the Hamilton operator, in the present notation reads

$$\hat{C}=i\frac{k}{\pi}(\partial_{\bar{z}}\hat{A}_z-\partial_z\hat{A}_{\bar{z}})-\rho. \tag{1.7}$$

Furthermore, the operator obtained by exponentiation

$$\hat{U}[g]=\exp\left(-i\int d^2x\,\alpha\hat{C}\right), \tag{1.8}$$

with $g=\exp(-i\alpha)\in U(1)$, yields a (proper) representation of time-independent gauge transformations, acting on the Schrödinger wave functionals as

$$\hat{U}[g]\psi[A_{\bar{z}}]=\exp(-i\omega[g,A_{\bar{z}}])\psi[g^{-1}A_{\bar{z}}], \tag{1.9}$$

where $gA_{\bar{z}}=A_{\bar{z}}+\partial_{\bar{z}}\alpha$, and ω is the 1-cocycle

$$\omega[g,A_{\bar{z}}]=-\int d^2x\,\alpha\rho+i\frac{k}{\pi}\int d^2x\,A_{\bar{z}}\partial_z\alpha-i\frac{k}{2\pi}\int d^2x\,\partial_{\bar{z}}\alpha\partial_z\alpha. \tag{1.10}$$

These gauge transformations play the role of (infinite-dimensional) Wigner symmetry transformations, and so it makes sense to require the wave functionals to be invariant with respect to these, i.e.,

$$\hat{U}[g]\psi[A_{\bar{z}}]=\psi[A_{\bar{z}}], \tag{1.11}$$

in accordance with Dirac's prescription that the constraint must annihilate physical states.

The invariance condition can be employed to determine the $A_{\bar{z}}$ dependence of the wave functional. For this purpose, we use the parametrization

$$A_{\bar{z}}=\partial_{\bar{z}}\chi, \tag{1.12}$$

where χ is complex so that gauge invariance is maintained. This relation is inverted by means of the Green's function

$P(x-y)=-4\partial_z G(x-y)$ of the operator $\partial_{\bar z}$, where $G(x-y)=-(1/4\pi)\ln|\mu(z-w)|^2$ is the standard propagator with μ the infrared cutoff; in particular, we obtain

$$\frac{\delta}{\delta A_{\bar z}(x)}=-\int d^2y P(x-y)\frac{\delta}{\delta\chi(y)}. \tag{1.13}$$

Observe that the transformation law (1.9) is a purely algebraic result; thus it holds as well if g is generalized to be an element of the complexification $U(1)^{\mathbb{C}}$. Then one can choose $\alpha=\chi$ to obtain

$$\psi[A_{\bar z}](x_1,\dots,x_N)=\exp\left(i\sum_n q_n\chi(x_n)+\frac{k}{2\pi}\int d^2x\,\partial_{\bar z}\chi\partial_z\chi\right)\psi(x_1,\dots,x_N), \tag{1.14}$$

and this is the result at which we aimed.

What remains to investigate is the Hamiltonian which, in the present notation, reads

$$\hat H=-\frac{1}{2m}\sum_n 2(\hat D_{\bar z_n}\hat D_{z_n}+\hat D_{z_n}\hat D_{\bar z_n}), \tag{1.15}$$

with the covariant derivatives

$$\hat D_{z_n}=\hat\nabla_{z_n}-\frac{1}{4\ell^2}q_n\bar z_n,\quad \hat D_{\bar z_n}=\hat\nabla_{\bar z_n}+\frac{1}{4\ell^2}q_n z_n. \tag{1.16}$$

Here we have split off the contribution of the external magnetic field, being hidden in the magnetic length $\ell=\sqrt{1/B}$ in units with $\hbar=c=e=1$. The Chern-Simons part is contained in the operators

$$\hat\nabla_{z_n}=\partial_{z_n}+i\frac{\pi}{k}q_n\int d^2x\,P(x_n,x)\frac{\delta}{\delta\chi(x)}, \tag{1.17}$$

$$\hat\nabla_{\bar z_n}=\partial_{\bar z_n}-iq_n\partial_{\bar z_n}\chi(x_n). \tag{1.18}$$

They act on the wave functionals as

$$\hat D_{z_n}\psi[A_{\bar z}](x_1,\dots,x_N)=\exp\left(i\sum_m q_m\chi(x_m)+\frac{k}{2\pi}\int d^2x\,\partial_{\bar z}\chi\partial_z\chi\right)\times D_{z_n}\psi(x_1,\dots,x_N), \tag{1.19}$$

and analogously for $\hat D_{\bar z_n}$. The operators

$$D_{z_n}=\nabla_{z_n}-\frac{1}{4\ell^2}q_n\bar z_n,\quad D_{\bar z_n}=\nabla_{\bar z_n}+\frac{1}{4\ell^2}q_n z_n \tag{1.20}$$

closely resemble Eqs. (1.16), but now we have

$$\nabla_{z_n}=\partial_{z_n}-\frac{\pi}{k}q_n\sum_m q_m P(x_n,x_m)=\partial_{z_n}-\frac{1}{k}\sum_{m\neq n}\frac{q_nq_m}{z_n-z_m}, \tag{1.21}$$

$$\nabla_{\bar z_n}=\partial_{\bar z_n}. \tag{1.22}$$

In the second version of Eq. (1.21), the term with $m=n$ is indeed absent, since both the infrared regulator μ and the short-distance cutoff ε drop out. For the former, this is obvious; for the latter, the assertion follows if we define it by means of the heat-kernel expansion to be

$$\lim_{y\to x}G(x-y)=\int_\varepsilon^{1/\mu}d\tau\langle x|e^{-\tau\Delta}|x\rangle=-\frac{1}{4\pi}\ln(\mu\varepsilon), \tag{1.23}$$

so that in flat space ε is x independent, which is the property we need. Beyond this, the definition also shows that the ultraviolet regulator has geometrical significance.

Here we make contact with conformal field theory since the operators (1.21) and (1.22) are recognized as (Abelian) Knizhnik-Zamolodchikov[34] derivatives. It is an essential point that they take an unsymmetric form because the antiholomorphic partner of the Knizhnik-Zamolodchikov connection

$$\mathcal{A}_{z_n}=-\frac{i}{k}\sum_{m\neq n}\frac{q_m}{z_n-z_m} \tag{1.24}$$

is simply absent.[50] This is a fact being well known to mathematicians (Ref. 48; see also Ref. 49), but not appreciated in the quantum Hall effect literature. In the present approach, the Knizhnik-Zamolodchikov connection appears as the relic of the Chern-Simons connection, and it is in its unsymmetry that the treatment of the Chern-Simons field as a quantum-mechanical degree of freedom, which is not eliminated beforehand at the classical level, manifests itself.

The Hamiltonian $\hat H$ acts on $\psi(x_1,\dots,x_N)$ as the operator

$$H=-\frac{1}{m}\sum_n(D_{\bar z_n}D_{z_n}+D_{z_n}D_{\bar z_n}), \tag{1.25}$$

and we restrict ourselves to the determination of the ground state with zero-point energy $E_0=\Sigma_n 1/2\omega_n$, where $\omega_n=q_nB/m$ is the cyclotron frequency. This is accomplished by requiring $D_{\bar z_n}\psi_0(x_1,\dots,x_N)=0$, which is solved by

$$\psi_0(x_1,\dots,x_N)=\exp\left(-\frac{1}{4\ell^2}\sum_n q_n|z_n|^2\right)\varphi_0(x_1,\dots,x_N), \tag{1.26}$$

with

$$\nabla_{\bar z_n}\varphi_0(x_1,\dots,x_N)=0. \tag{1.27}$$

What remains is

$$H\psi_0(x_1,\dots,x_N)=\exp\left(-\frac{1}{4\ell^2}\sum_m q_m|z_m|^2\right)\times\left(-\frac{1}{m}\sum_n\nabla_{\bar z_n}\nabla_{z_n}+E_0\right)\varphi_0(x_1,\dots,x_N), \tag{1.28}$$

so that we are done if we also require

$$\nabla_{z_n}\varphi_0(x_1,\ldots,x_N)=0. \tag{1.29}$$

The solution of the Knizhnik-Zamolodchikov equations (1.27) and (1.29) is the Abelian conformal block

$$\varphi_0(x_1,\ldots,x_N)=\mathcal{N}_0'\prod_{n<m}(z_n-z_m)^{q_nq_m/k}, \tag{1.30}$$

where no regularization problems become involved at this stage. As for the quantum Hall effect we are only interested in the properties of the electrons, and so we can take $A_{\bar z}=0$ in the wave functional; furthermore, on choosing $q_n=1$ and $k=1/p$, with p an odd integer, we thus obtain, with Eqs. (1.26) and (1.30), the Laughlin wave function, as promised.

The derivation clearly shows that this state is of minimum energy and, in addition, provides for an explanation of the main series of filling fractions $\nu=1/p$. A basic assumption in obtaining this result has been that it is not enough to simply cancel the third coordinate in order to describe quantum-mechanical particles restricted to two dimensions; they must be supplied with the degrees of freedom, being specific to the two-dimensional case, and this is accomplished by means of the Chern-Simons (quantum) field.

The conformal block φ_0 can be interpreted as the vacuum expectation value of N chiral vertex operators with ''charges'' $q_n/\sqrt{k}$, and this may be seen as an explanation of the hitherto rather accidental, fact that the holomorphic part of the Laughlin ground-state wave function can also be obtained by means of vertex operator techniques (Ref. 5; see also below).

Up to now, we could avoid regularization problems; however, they come in on adressing normalization issues. From Eq. (1.6), for the inner product we obtain

$$\begin{aligned}\langle\psi(x_1,\ldots,x_N)|\psi'(x_1,\ldots,x_N)\rangle &=\frac{\det\Delta}{|\mathcal{G}|}\int d[\bar\chi,\chi]\exp\left(-\frac{k}{\pi}\int d^2x\,\partial_{\bar z}\chi\partial_z\bar\chi\right)\\ &\quad\times\psi^*[\partial_{\bar z}\chi](x_1,\ldots,x_N)\psi'[\partial_{\bar z}\chi](x_1,\ldots,x_N),\end{aligned} \tag{1.31}$$

where we have factored out the volume $|\mathcal{G}|$ of the gauge group in order to count each gauge orbit only once. We rewrite this in the multiplicative form

$$\begin{aligned}\langle\psi(x_1,\ldots,x_N)|\psi'(x_1,\ldots,x_N)\rangle &=\psi^*(x_1,\ldots,x_N)K(x_1,\ldots,x_N)\psi'(x_1,\ldots,x_N),\end{aligned} \tag{1.32}$$

with the diagonal kernel

$$K(x_1,\ldots,x_N)=\det\Delta\int d[\phi]\exp\left(-\frac{2k}{\pi}\int d^2x\,\partial_{\bar z}\phi\partial_z\phi-2\sum_n q_n\phi(x_n)\right) \tag{1.33}$$

and where $\phi=\mathrm{Im}\,\chi$. The functional integral is recognized as the vacuum expectation value of the product of N (nonchiral) vertex operators with ''imaginary charges,'' which can be evaluated to give

$$K(x_1,\ldots,x_N)=(\det\Delta)^{1/2}\mu^{-(1/k)Q^2}\varepsilon^{-(1/k)\Sigma_nq_n^2}\times\prod_{n<m}|x_n-x_m|^{-2(q_nq_m/k)}. \tag{1.34}$$

Restricting ourselves to the ground-state wave functional, we want to achieve that the cutoffs cancel in the inner product. But we then see that choice (1.30) for φ_0 is not quite right, but (on appealing to the vertex operator analogy, alluded to above) we can correct for this by instead choosing

$$\begin{aligned}\varphi_0(x_1,\ldots,x_N)&=\mathcal{N}_0\exp\left(-\frac{2\pi}{k}\sum_{n,m}q_nq_mG(z_n-z_m)\right)\\ &=\mathcal{N}_0\mu^{(1/2k)Q^2}\varepsilon^{(1/2k)\Sigma_nq_n^2}\prod_{n<m}(z_n-z_m)^{q_nq_m/k},\end{aligned} \tag{1.35}$$

where $G(z)$ is the holomorphic part of the propagator, and $\mathcal{N}_0$ is cutoff independent. The right-hand side of Eq. (1.35) is just the inverse of the holomorphic square root of Eq. (1.34), so that the cutoff dependent contributions precisely cancel. Hence the inner product

$$\langle\psi_0(x_1,\ldots,x_N)|\psi_0(x_1,\ldots,x_N)\rangle=|\mathcal{N}_0|^2(\det\Delta)^{1/2}\times\exp\left(-\frac{1}{2\ell^2}\sum_n q_n|z_n|^2\right) \tag{1.36}$$

is a finite quantity if the determinant of the Laplacian is understood to be regularized by means of ζ-function techniques. Finally, the x_n integrations remain to be done, but for $q_n=1$ they pose no problem since they are all Gaussian.

It is noteworthy that the present approach provides for a rather unexpected resolution of the normalization problem for the Laughlin wave function. Furthermore, let us stress that if we had we attempted to regularize the wave functional itself, and not its inner product, we would have run into trouble. As is well known, the ultraviolet cutoff can be absorbed in a multiplicative renormalization of the vertex operator by normal ordering, but what is left is the infrared cutoff, which enforces the total charge $Q=\Sigma_nq_n$ to vanish. One could dispose of this restriction by means of the Dotsenko-Fateev background charge method,[51] but in the present context there is no point in so doing. We only have to guarantee the wave functional to be normalizable, and this is the reason why one can escape the conclusion that the total charge must be zero. We shall have occasion to return to this topic repeatedly in the course of the further development.

We end this section with a brief discussion of excitations, which may be obtained in a rather straightforward manner within the present approach. Let us assume the existence of charged excitations, which are designed to balance a small increase of the external magnetic field such that the filling is locked at the original value $\nu=1/p$ with $p>1$ odd. Since the energy of the fictious particles should be strictly less than the zero-point energy $\epsilon=\omega/2$ per electron, it is tempting to assume that their charge e^* is fractional; in particular, we choose $e^*=e/p$. Under the further assumption that the mass of the excitations equals that of the electrons, we have then achieved that $\omega^*=\omega/p$, and so the zero-point energy of an excitation is strictly less than that of an electron. The excitations are also subject to the coupling to the Chern-Simons field; hence we can now use the above results [see Eq. (1.26) and (1.35)], where it will pay that we have not specified the charges of the constituents from the outset. Thus, partitioning the coordinates into $(x_1,\ldots,x_N)$ with charges $e(q_n=1)$ for the electrons, and $(y_1,\ldots,y_M)$ with charges $e^*(q_m=1/p)$ for the excitations, we immediately obtain the ground-state wave function

$$\psi_{0;N,M}(x_1,\ldots,x_N;y_1,\ldots,y_M)$$
$$=\mathcal{N}_{0;N,M}\exp\left(-\frac{1}{4\ell^2}\sum_n |z_n|^2-\frac{1}{4\ell^2}\frac{1}{p}\sum_m |w_m|^2\right)$$
$$\times\prod_{n<n'}^{N}(z_n-z_{n'})^p\prod_{n,m}^{N,M}(z_n-w_m)\prod_{m<m'}^{M}(w_m-w_{m'})^{1/p}, \tag{1.37}$$

of energy $E_{0;N,M}=1/2\omega[N+(1/p)M]$, and the normalization factor is

$$\mathcal{N}_{0;N,M}=\mathcal{N}_0\mu^{[N+(1/p)M]}\varepsilon^{(1/2)[pN+(1/p)M]}. \tag{1.38}$$

This is recognized as Laughlin's trial state with N electrons and M excitations, where only the last factor is less well known; this immediately shows that, whereas the conventional statistics of the electrons is not changed by the coupling to the Chern-Simons field, the statistics of the excitations with $s^*=1/p$ is fractional.[6] Of course, we do not claim to have given any additional insight into the origin of the excitations, which remains obscure; a deeper understanding requires a second quantized description of the electrons, which will be given elsewhere.

II. ELECTRONS ON THE TORUS

We want to generalize the results obtained so far to realistic boundary conditions; thus the electrons are restricted to a rectangular domain in the plane of extensions L_1 and L_2, i.e., the torus. For the time being, we only keep the Chern-Simons field; the external magnetic field will be taken care of at the very end.

Beyond the seminal work of Haldane and Rezayi,[37] an abundant literature is available on this subject,[52] in which the classical reduction procedure is used throughout. For the present approach, the work of Bos and Nair (Ref. 13; cf. also Ref. 53), though not devoted to the quantum Hall effect, will be of special relevance. These authors treated the general case of a Riemann surface of arbitrary genus with Wilson lines inserted; however, they explicitly restricted the investigation to the case of vanishing total charge, a restriction, which is generally believed to be unavoidable on a closed surface (cf. also Ref. 11). Hence, it will be essential for what follows to circumvent this verdict, which also seems to be one of the main stumbling blocks in related investigations.

So let us return to the first part of Sec. I where all integrations are now understood to extend over the fundamental domain of the torus. But it is essential to note that we can no longer ignore boundary terms; they must all be kept. Then the basic results contained in formulas (1.1)–(1.10) remain valid with one exception; formula (1.8), which must be modified. Here we encounter a basic difference, as compared to the plane, because the general rule that the constraint $\hat{C}[\alpha]=\int d^2x\,\alpha\hat{C}$ can be identified with the generator of gauge transformations no longer applies to a system restricted to a finite domain. What is to be called the generator of gauge transformations can be inferred from the classical symmetries. The Chern-Simons part of the Lagrangian is not invariant under time-independent gauge transformations, but changes by a total time derivative only so that the generalized Noether procedure[54] yields the conserved quantity

$$\hat{Q}[\alpha]=-i\frac{k}{\pi}\int d^2x(\partial_{\bar{z}}\alpha\hat{A}_z-\partial_z\alpha\hat{A}_{\bar{z}})-\int d^2x\,\alpha\rho, \tag{2.1}$$

where again the hats signify the transition to the operator level. This differs from the constraint $\hat{C}[\alpha]$ through a crucial boundary term, which cannot be neglected since the Chern-Simons gauge field (A_1,A_2) neither vanishes on the boundary, nor may be assumed to be periodic. If we now define the operator of gauge transformations by

$$\hat{U}[g]=\exp(-i\hat{Q}[\alpha]), \tag{2.2}$$

then the transformation law of the Schrödinger wave functional takes the same form, as given in Eqs. (1.9) and (1.10) above.

But one must qualify what kinds of gauge transformations $g(x)=\exp[-i\alpha(x)]$ are allowed: we postulate that they must be doubly periodic. This requirement permits two classes: small gauge transformations with doubly periodic parameters $\alpha(x)$, and large gauge transformations

$$g_{(m_1,m_2)}(x)=\exp[-i\,\alpha_{(m_1,m_2)}(x)] \tag{2.3}$$

depending on two integers m_1 and m_2, with ''parameters''

$$\alpha_{(m_1,m_2)}(x)=2\pi\left(m_1\frac{x_1}{L_1}+m_2\frac{x_2}{L_2}\right), \tag{2.4}$$

not being continuosly connected to the identity. Accordingly, we allow for gauge fields which, in addition to the standard contribution (1.12), contain a constant term, as well as a term depending linearly on the coordinates. We make the choice

$$A_{\bar{z}}(x)=\partial_{\bar{z}}\chi(x)+a+bz, \tag{2.5}$$

where a, b, and χ take compex values and χ is required to be doubly periodic; the nonstandard b term will turn out to be of special relevance. Again, the small gauge transformations only affect χ, whereas the large gauge transformations can be absorbed in the constant term a, which is mapped into

$$a'=a+i\frac{\pi}{L_2}(m_2-\tau m_1), \tag{2.6}$$

where $\tau=iL_2/L_1$ is the modular parameter of the (rectangular) torus. The b term is left inert under both types of transformations; we shall see that it will play an essential role in obtaining values $Q\neq 0$ of the total charge.

Let us return to operator (2.2), implementing the gauge transformations on wave functionals. It is easy to see that the cocycle condition

$$\omega[hg,A_{\bar{z}}]-\omega[h,A_{\bar{z}}]-\omega[g,h^{-1}A_{\bar{z}}]=0 \tag{2.7}$$

only holds for small gauge transformations, whereas for large gauge transformations it is violated by a nonvanishing boundary term. For the operators $\hat{U}_{(m_1,m_2)}=\exp(-i\hat{Q}[\alpha_{(m_1,m_2)}])$, this entails that they do not commute,

$$\hat{U}_{(m_1,m_2)}\hat{U}_{(n_1,n_2)}=\exp[2\pi ik\times(m_1n_2-m_2n_1)]\hat{U}_{(n_1,n_2)}\hat{U}_{(m_1,m_2)}, \tag{2.8}$$

in spite of the fact that, classically, the large gauge transformations $g_{(m_1,m_2)}$ form an Abelian group. Thus we are faced with a gauge anomaly.

Of special importance for the further development will be the central extension

$$\hat{U}_{(\lambda;m_1,m_2)}=\lambda\hat{U}_{(m_1,m_2)}, \tag{2.9}$$

with $\lambda=\exp(\pi ikn)$ and $n\in\mathbb{Z}$, which is a Heisenberg-Weyl group with composition law

$$(\lambda;m_1,m_2)(\lambda';m_1',m_2')=(\lambda\lambda' e^{\pi ik(m_1m_2'-m_2m_1')};m_1+m_1',m_2+m_2') \tag{2.10}$$

Realization (2.9) is an operator-valued representation of this discrete group in the ordinary sense.

It is essential to note that small and large gauge transformations commute. Furthermore, they commute with the covariant derivatives $\hat{\nabla}_{\bar{z}_n}$ and $\hat{\nabla}_{z_n}$, and thus with the Hamiltonian as well. In addition, they are unitary with respect to the Bargmann inner product.

In the following, we need the inversion of the parametrization (2.5); this is accomplished by means of the doubly periodic propagator $G(x,y)$ with the properties[55]

$$\Delta G(x,y)=\delta(x,y)-\frac{1}{V},\quad \int d^2x\, G(x,y)=0, \tag{2.11}$$

where $V=L_1L_2$. It is given in terms of another propagator $G_0(x-y)$, only obeing the first of the above properties, which is

$$G_0(x-y)=-\frac{1}{4\pi}\ln|E(v-w)|^2+\frac{1}{2}\frac{[\mathrm{Im}(v-w)]^2}{\mathrm{Im}\,\tau}, \tag{2.12}$$

where $v=z/L_1$ and $E(v)=\theta_1(v|\tau)/\theta_1'(0|\tau)$ are the prime form; θ_1 is the odd Jacobi theta function.[56] The propagator we need then is (cf. also Ref. 57)

$$G(x,y)=G_0(x-y)-\frac{1}{V}\int d^2x'[G_0(x-x')+G_0(x'-y)]+\frac{1}{V^2}\int d^2x'd^2y'\,G_0(x'-y'), \tag{2.13}$$

and it is easy to check that it is indeed orthogonal to the zero mode. We also shall have need for

$$P(x,y)=-4\partial_z G(x,y)=P_0(x,y), \tag{2.14}$$

where it makes no difference which of the two Green's functions we use. As a last remark, the propagator G_0 is not modular invariant; this property can be supplied for on replacing the prime form E by[58]

$$F(v)=\theta_1(v|\tau)/\eta(\tau), \tag{2.15}$$

where η is the Dedekind function; however on passing to Eq. (2.13), modular invariance is lost again.

With these preliminaries out of the way, the inversion can be done, and we end up with[59]

$$\frac{\delta}{\delta A_{\bar{z}}(x)}=-\int d^2y\,P(x,y)\frac{\delta}{\delta\chi(y)}+\frac{1}{V}\frac{\delta}{\delta a}. \tag{2.16}$$

Note that the coefficient b does not get involved on the right-hand side; it is not quantized. An adequate decription of the b term appears to be that it plays the role of a classical Chern-Simons background field.

Now we would like to proceed as in Sec. I to determine the χ dependence of the wave functional by means of small gauge transformations, so we could try again to impose the requirement that $\hat{U}[g]$ acts as the identity operator on physical states. However, on so doing, one soon runs into severe consistency problems, which may be traced back for the b-term. Its contribution to the 1-cocycle reads $(k/\pi)b\int d^2x\, z\partial_z\alpha=(k/\pi)b\int d^2x\,\partial_z(z\alpha)-(k/\pi)b\int d^2x\,\alpha$, and we would prefer if we could get rid of the nasty boundary term. For the moment, this remark should suffice to motivate that we only require physical wave functionals to be invariant up to a phase, i.e.,

$$\hat{U}[g]\psi[A_{\bar{z}}]=e^{-i\phi[\alpha]}\psi[A_{\hat{z}}], \tag{2.17}$$

where the phase

$$\phi[\alpha]=i\frac{k}{\pi}b\int d^2x\,\partial_z[z\alpha(x)] \tag{2.18}$$

is linear in α, and thus respects the group law. Because $\hat{U}[g]$ acts unitarily,[60] the above condition (2.17) also respects the inner product. On passing to the complexification $U(1)^{\mathbb{C}}$, we thus obtain

$$\psi[A_{\bar{z}}](x_1,\ldots,x_N)=\exp\left(i\sum_n q_n\chi(x_n)-\frac{k}{\pi}b\int d^2x\,\chi+\frac{k}{2\pi}\int d^2x\,\partial_{\bar{z}}\chi\partial_z\chi\right)\psi[a](x_1,\ldots,x_N), \tag{2.19}$$

so that the χ and a dependences factorize. As a check, one verifies directly that this wave functional indeed obeys the condition (2.17) that we began with.

What remains to discuss is the role of the constraint, which in terms of the new variables reads as follows:

$$\hat{C}(x)=-i\frac{\delta}{\delta\chi(x)}-i\frac{k}{\pi}\partial_z\partial_{\bar{z}}\chi(x)-\sum_n q_n\delta(x-x_n)-i\frac{k}{\pi}b+\frac{i}{V}\int d^2y\frac{\delta}{\delta\chi(y)}. \tag{2.20}$$

As we have already commented upon, for a system living in a domain of finite extent, gauge invariance and the imposition of the constraint are different issues. Hence, imposing the constraint is a separate requirement, which amounts to

$$\hat{C}(x)\psi[A_{\bar{z}}]=\frac{i}{V}\int d^2y\,\frac{\delta}{\delta\chi(y)}\psi[A_{\bar{z}}]=0, \tag{2.21}$$

and this condition fixes the value of b to be[61]

$$b=i\frac{\pi}{k}\frac{Q}{V}. \tag{2.22}$$

Note then that the first two terms in the exponential on the right-hand side of Eq. (2.19) add up to give an effective charge density of the total effective charge zero.

Let us turn to large gauge transformations, the explicit form of which follows from Eqs. (1.9) and (1.10) to be

$$\begin{aligned}\hat{U}_{(m_1,m_2)}\psi[A_{\bar{z}}]=&\exp\left(\pi\frac{Q}{L_2}[(m_2-\bar{\tau}m_1)Z-\bar{Z}(m_2-\tau m_1)]\right)\\&\times\exp\left(-ikL_1(m_2-\bar{\tau}m_1)a\right.\\&+\frac{\pi}{2}\frac{Q}{L_2}(m_2-\bar{\tau}m_1)L-k\frac{\pi}{2}\frac{L_1}{L_2}(m_2-\bar{\tau}m_1)\\&\left.\times(m_2-\tau m_1)\right)\psi\left[A_{\bar{z}}-i\frac{\pi}{L_2}(m_2-\tau m_1)\right],\end{aligned} \tag{2.23}$$

with $L=L_1+iL_2$; furthermore, we have introduced the center of charge $Z=\Sigma_n q_n z_n/Q$, which always appears in the form QZ, and so makes sense as well for the case of vanishing total charge. These transformations only affect the a dependence, and thus we can take χ to be zero. At this point, one could guess that invariance under large gauge transformations should as well be imposed only up to phase (2.18), i.e., in the same manner as for small ones. But this guess is wrong, as we want to make plausible now, and for this purpose, we return once more to small gauge transformations. Its generators may be split as

$$\hat{Q}[\alpha]=\hat{C}[\alpha]+\hat{B}[\alpha], \tag{2.24}$$

with the ''boundary'' operator

$$\hat{B}[\alpha]=-i\int d^2x\,\partial_{\bar{z}}\left(\alpha\frac{\delta}{\delta A_{\bar{z}}}\right)+i\frac{k}{\pi}\int d^2x\,\partial_z(\alpha A_{\bar{z}}), \tag{2.25}$$

depending only on the Chern-Simons field and not on the matter part. For the parametrization (2.5), this operator reduces to

$$\hat{B}[\alpha]=\phi[\alpha], \tag{2.26}$$

with $\phi[\alpha]$ the phase introduced in Eq. (2.18); this is a c number only, and so it is obvious that $\hat{B}$ and $\hat{C}$ commute. We thus have revealed the origin of condition (2.17), because the imposition of the constraint just yields the invariance up to a phase. After all, condition (2.22), having been obtained from the requirement that the constraint annihilates physical states, now appears in a different perspective, since it may be seen as a statement about ''global'' gauge invariance [cf. the remark following Eq. (2.22)]. Indeed, for $\mathring{\alpha}$ an x independent gauge transformation, we have $\hat{Q}[\mathring{\alpha}]=-\mathring{\alpha}Q$ and $\phi[\mathring{\alpha}]=i(k/\pi)bV\mathring{\alpha}$, so that Eq. (2.17) yields the assertion.

As to large gauge transformations, however, things are rather different. For a proper understanding of their finesses, it is helpful to forget for a moment the matter part, i.e., we only investigate pure Chern-Simons theory on the torus; but we could also take a Riemann surface of arbitrary genus.[53] Then it is rather straightforward to show that the operators $\hat{U}_{(m_1,m_2)}$, for the special values $m_1=0$, $m_2=-1$ and $m_1=1$, $m_2=0$, coincide with the holonomy operators

$$\hat{U}_{(0,-1)}=\exp ik\int_\alpha \hat{A}\,dx,\quad \hat{U}_{(1,0)}=\exp ik\int_\beta \hat{A}\,dx, \tag{2.27}$$

where α and β denote the two independent homology cycles on the torus. Thus, for a pure Chern-Simons theory the generators of large gauge transformations have deep geometrical topological significance (cf. Ref. 11). Hence one expects the operators $\hat{U}_{(m_1,m_2)}$ to be of comparable importance as well if matter particles are present: however, then such a Wegner-Wilson type interpretation is no longer available.

Returning to the problem at hand, for large gauge transformations, the boundary term (2.25) does not degenerate into a c number; in particular, one verifies that the operators $\exp -i\hat{B}[\alpha_{(m_1,m_2)}]$ and $\hat{U}_{(m_1,m_2)}$ act identically on wave functionals, as one expects, and so we may leave it with the latter. Furthermore, it would be inconsistent to require the operators $\hat{U}_{(m_1,m_2)}$ to act as the identity operator on physical wave functionals, because they need not commute for different values of (m_1,m_2). Taking k to be rational, i.e., $k=k_1/k_2$, with k_1 and k_2 coprime integers, we can avoid a contradiction if we only require $\hat{U}_{k_2(m_1,m_2)}$ to act as an identity up to an m-dependent phase. Specifically, we choose (cf. also Ref. 52)

$$\begin{aligned}\hat{U}_{k_2(m_1,0)}\psi[A_{\bar{z}}]&=e^{-2\pi i\eta_1 m_1}\psi[A_{\bar{z}}],\\ \hat{U}_{k_2(0,m_2)}\psi[A_{\bar{z}}]&=e^{-2\pi i\eta_2 m_2}\psi[A_{\bar{z}}],\end{aligned} \tag{2.28}$$

with $\eta_{1,2}\in[0,1]$, since it is this choice, which prevents the conflict with the (global) gauge anomaly.

Now it amounts to a lengthy but straightforward calculation to solve these conditions in terms of Jacobi θ functions[56] with characteristics α, $\beta\in\mathbb{R}$,

$$\begin{aligned}\theta\begin{bmatrix}\alpha\\ \beta\end{bmatrix}(w|\sigma)=\sum_{n=-\infty}^{+\infty}&\exp\{\pi i\sigma(n+\alpha)^2\\&+2\pi i(n+\alpha)(w+\beta)\},\end{aligned} \tag{2.29}$$

where w, $\sigma\in\mathbb{C}$ and $\mathrm{Im}\,\sigma>0$. There are k_1k_2 linear independent solutions of Eq. (2.28), which may be written as

$$\psi_l[a](x_1,\ldots,x_N)=\exp\left[-\frac{k}{2\pi}V\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)^2 + iQ\overline{Z}\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\right] \times\exp\left[-\frac{i}{2}Q\overline{L}\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\right] \times\theta\begin{bmatrix}\frac{1}{k_1k_2}\left(l+\eta_2+\frac{Q}{2}k_2\right)\\ \eta_1+\frac{Q}{2}k_2\end{bmatrix} \times\left[-i\frac{k_1}{\pi}L_2\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\middle| k_1k_2\tau\right] \times\phi(x_1,\ldots,x_N), \tag{2.30}$$

with $l=1,\ldots,k_1k_2$. For $Q=0$, Eq. (2.30) coincides with an analogous result in Ref. 13 by specializing to the case of genus one there.

However, this is not the final form, we still have to discuss the Hamiltonian

$$\hat{H}=-\frac{1}{2m}\sum_n 2(\hat{\nabla}_{\bar{z}_n}\hat{\nabla}_{z_n}+\hat{\nabla}_{z_n}\hat{\nabla}_{\bar{z}_n}), \tag{2.31}$$

where the covariant derivatives are given by

$$\hat{\nabla}_{\bar{z}_n}=\partial_{\bar{z}_n}-iq_n\left(\partial_{\bar{z}_n}\chi(x_n)+a+i\frac{\pi}{k}\frac{Q}{V}z_n\right), \tag{2.32}$$

$$\hat{\nabla}_{z_n}=\partial_{z_n}+i\frac{\pi}{k}q_n\int d^2xP(x_n,x)\frac{\delta}{\chi(x)}-i\frac{\pi}{k}\frac{q_n}{V}\frac{\delta}{\delta a}. \tag{2.33}$$

If these are applied to the wave functionals (2.19) and (2.30), we have to commute them through in front of ϕ; on so doing, it proves to be advantageous to define

$$\phi(x_1,\ldots,x_N)=\exp\left(\frac{\pi}{k}\frac{Q^2}{V}|Z|^2-\frac{\pi}{2k}\frac{Q^2}{V}Z^2 -\frac{\pi}{k}\frac{Q}{V}\sum_n q_n|z_n|^2 +\frac{\pi}{2k}\frac{Q}{V}\sum_n q_nz_n^2\right)\varphi(x_1,\ldots,x_N), \tag{2.34}$$

so that the covariant derivatives act on φ in the simplified forms

$$\nabla_{\bar{z}_n}\varphi(x_1,\ldots,x_N)=\partial_{\bar{z}_n}\varphi(x_1,\ldots,x_N), \tag{2.35}$$

$$\nabla_{z_n}\varphi(x_1,\ldots,x_N)=\left[\partial_{z_n}-\frac{q_n}{k}\partial_{z_n}\sum_m q_m \times\ln E\left(\frac{1}{L_1}(z_n-z_m)\right)\right]\varphi(x_1,\ldots,x_N). \tag{2.36}$$

They may appropriately be addressed to as the Knizhnik-Zamolodchikov derivatives on the torus. Accodingly, the final form of the wave functionals reads

$$\psi_l[A_{\bar{z}}](x_1,\ldots,x_N)=\exp\left[i\int d^2x\left(\rho-\frac{Q}{V}\right)\chi +\frac{k}{2\pi}\int d^2x\,\partial_{\bar{z}}\chi\partial_z\chi\right] \times\exp\left(-\frac{k}{2\pi}Va^2-iQ(Z-\overline{Z})a -\frac{\pi}{k}\frac{Q}{V}\sum_n q_n|z_n^2|+\frac{\pi}{2k}\frac{Q}{V}\sum_n q_nz_n^2\right) \times\exp\left[-\frac{i}{2}Q\overline{L}\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\right] \times\theta\begin{bmatrix}\frac{1}{k_1k_2}\left(l+\eta_2+\frac{Q}{2}k_2\right)\\ \eta_1+\frac{Q}{2}k_2\end{bmatrix} \times\left[-i\frac{k_1}{\pi}L_2\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\middle| k_1k_2\tau\right] \times\varphi(x_1,\ldots,x_N). \tag{2.37}$$

The Hamiltonian $\hat{H}$ acts on φ as the operator

$$H\varphi=-\frac{1}{m}\sum_n(\nabla_{\bar{z}_n}\nabla_{z_n}+\nabla_{z_n}\nabla_{\bar{z}_n})\varphi, \tag{2.38}$$

with the derivatives as shown.

It is a straightforward matter now to determine the minimum energy state of the reduced Hamiltonian (2.38) by following the same logic as in the plane. This is the solution of the equations

$$\nabla_{z_n}\varphi_0(x_1,\ldots,x_N)=0,\quad \nabla_{\bar{z}_n}\varphi_0(x_1,\ldots,x_N)=0 \tag{2.39}$$

and is obtained to be

$$\varphi_0(x_1,\ldots,x_N)=\mathcal{N}_0\prod_n \varepsilon^{q_n^2/2k}\prod_{n<m}E\left(\frac{1}{L_1}(z_n-z_m)\right)^{q_nq_m/k}, \tag{2.40}$$

where the normalization constant $\mathcal{N}_0$ is cutoff independent since there is no infrared regulator on a compact surface. Hence the ground states of the system turn out to be exactly solvable.

Let us specialize the above result to the case relevant for the quantum Hall effect. Because the wave functional must be completely antisymmetric in the electron coordinates, this requires $k_1=1$ and $k_2=p$ to be odd integers. At this point we come into contact with the work of Haldane and Rezayi,[37] in which Laughlin's ansatz is generalized to the torus; this was also discussed by Cristofano *et al.*[5] by means of vertex operator techniques.[62] If we define $\psi_{0;l}[A_{\bar z}]|_{\chi=a=0}(x_1,\ldots,x_N)=\psi_{o;l}(x_1,\ldots,x_N)$, we expect these wave functions to verify the Haldane-Rezayi ground states. However, the explicit form

$$\psi_{o;l}(x_1,\ldots,x_N)=\exp\left(-\frac{\pi}{k}\frac{Q}{V}\sum_n q_n|z_n|^2+\frac{\pi}{2k}\frac{Q}{V}\sum_n q_n z_n^2+\frac{\pi}{2k}\frac{Q}{V}\overline{L}\sum_n q_n z_n\right)\times\theta\begin{bmatrix}\frac{1}{p}\left(l+\eta_2+\frac{Q}{2}p\right)\\ \eta_1+\frac{Q}{2}p\end{bmatrix}\times\left(p\frac{1}{L_1}\sum_n q_n z_n\middle|p\tau\right)\varphi_0(x_1,\ldots,x_N)\tag{2.41}$$

reveals that our result differs, among other things, in a decisive exponential prefactor, one can get rid of only if the total charge $Q=\Sigma_n q_n$ is required to vanish. Hence the degenerate Haldane-Rezayi ground state appears to be restricted to the sector of total charge zero, and so should hardly be related to the quantum Hall effect.

One could object that the final results (2.37) and (2.40) an ''gauge dependent'' in the sense that the b term in Eq. (2.5) can be modified. But we could as well have chosen an ''unsymmetric gauge,'' for which the parametrization takes the form

$$A_{\bar z}(x)=\partial_{\bar z}\chi(x)+a+i\frac{\pi}{k}\frac{Q}{V}(z-\bar z).\tag{2.42}$$

Then one can follow similar steps as before to obtain the analog of result (2.37), which reads

$$\psi_l[a](x_1,\ldots,x_N)=\exp\left(-\frac{k}{2\pi}Va^2-iQ(Z-\bar Z)a+\frac{\pi}{2k}\frac{Q}{V}\sum_n q_n(z_n-\bar z_n)^2\right)\times\exp\left[-QL_2\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\right]\times\theta\begin{bmatrix}\frac{1}{k_1k_2}(l+\eta_2+Qk_2)\\ \eta_1\end{bmatrix}\times\left[-i\frac{k_1}{\pi}L_2\left(a+i\frac{\pi}{k}\frac{Q}{V}Z\right)\middle|k_1k_2\tau\right]\times\varphi(x_1,\ldots,x_N),\tag{2.43}$$

with the χ dependence remaining unaltered. Inspection shows, however, that there are again additional exponential prefactors for $Q\neq 0$. As we see it, there is no special advantage in working with these wave functionals. This is essentially different from the situation one encounters in the one-particle Landau theory, where toroidal boundary conditions can only be imposed in the unsymmetric gauge.[63]

III. MODULAR INVARIANCE AND RENORMALIZATION

The torus can be realized as the quotient $\mathbb{C}/(\mathbb{Z}+\tau\mathbb{Z})$, with $\mathbb{C}$ its universal covering space. The lattice $\mathbb{Z}+\tau\mathbb{Z}$ derives from large diffeomorphisms $z\to z+n_1+\tau n_2$, and this interpretation is in line with the notion that it should make no difference, physically, which one of the tori labeled by n_1 and n_2 one selects. These diffeomorphisms are generated by modular transformations $M\in SL(2,\mathbb{Z})$, and so the request for modular invariance is of direct physical relevance.

Though the coordinates x_n are modular parameters as well, we restrict ourselves to the pure a dependence. The θ functions depend on a through the variable $\bar v=-(i/\pi)(L_2/p)a$ with $1>v_1\geqslant 0$ and $L_2/L_1>v_2\geqslant 0$. Accordingly, the variables in the general definition (2.29) are to be identified as $w=pv$ and $\sigma=p\tau$, where we have suppressed the complex conjugation of v, and

$$\alpha=\frac{1}{p}\left(l+\eta_2+\frac{Q}{2}p\right),\quad \beta=\eta_1+\frac{Q}{2}p.\tag{3.1}$$

Note that it is v and τ, which are the modular parameters of the true torus.

Under a modular transformation

$$M=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\tag{3.2}$$

they are transformed into[57]

$$w'=w/(c\sigma+d),\quad \sigma'=(a\sigma+b)/(c\sigma+d)\tag{3.3}$$

and

$$\beta'=a\beta-b\alpha+\tfrac{1}{2}ab,\quad \alpha'=-c\beta+d\alpha+\tfrac{1}{2}cd.\tag{3.4}$$

It is a known result that the θ functions behave under these transformations as

$$\theta\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}(w'|\sigma')=\varepsilon_M e^{\pi i\phi_M(\alpha,\beta)}(c\sigma+d)^{1/2}e^{\pi i[c/(c\sigma+d)]w^2}\times\theta\begin{bmatrix}\alpha\\ \beta\end{bmatrix}(w|\sigma),\tag{3.5}$$

with

$$\phi_M(\alpha,\beta)=-\alpha\beta+(a\beta-b\alpha)(-c\beta+d\alpha)+(-c\beta+d\alpha)ab\tag{3.6}$$

and ε_M a complicated eighth root of unity, which only depends on M.

The modular group is generated by two elements M_1 and M_2, and we begin with $M_1=\left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right)$, transforming the θ function into

$$\theta\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}(w'|\sigma')=\theta\begin{bmatrix}-\beta\\ \alpha\end{bmatrix}\left(v'\,\middle|\,\frac{1}{p}\tau'\right), \tag{3.7}$$

with $v'=v/\tau$ and $\tau'=-1/\tau$. However, it is not the result we want since the right-hand side must depend on $(pv'|p\tau')$, but this can be corrected to give

$$\theta\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}(w'|\sigma')=\sum_{l'} e^{2\pi i(1/p)[l'-\eta_1-(Q/2)p]l}\times\theta\begin{bmatrix}\frac{1}{p}\left(l'-\eta_1-\frac{Q}{2}p\right)\\ \eta_2+\frac{Q}{2}p\end{bmatrix}(pv'|p\tau'), \tag{3.8}$$

so that the arguments of the θ function now only contain the modular transformed values of the variables of the true torus. For the second generator $M_2=\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$, a similar reasoning yields

$$\begin{aligned}\theta\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}(w'|\sigma')&=\theta\begin{bmatrix}\alpha\\ \beta-\alpha+\frac{1}{2}\end{bmatrix}(pv|p\tau+1)\\ &=\exp\left\{\pi i(p-1)\frac{1}{p}\left(l+\eta_2+\frac{Q}{2}p\right)\right.\\ &\quad\left.\times\left[\frac{1}{p}\left(l+\eta_2+\frac{Q}{2}p\right)-1\right]\right\}\\ &\quad\times\theta\begin{bmatrix}\frac{1}{p}\left(l+\eta_2+\frac{Q}{2}p\right)\\ \eta_1-(l+\eta_2)+\frac{p}{2}\end{bmatrix}[pv|p(\tau+1)].\end{aligned} \tag{3.9}$$

What must finally be achieved is that the θ-function characteristics, appearing on the right-hand sides of Eqs. (3.8) and (3.9), only take values in the same range, as do the characteristics α and β given in Eq. (3.1); this restricts the admissible values of the phases. The computation shows that this requirement is met for $\eta_1=\eta_2=0$ if Q is odd, and $\eta_2=\frac{1}{2}=-\eta_1$ if Q is even, including the value zero.

The remaining part of this section is devoted to normalization problems. As we have seen, in the plane there is one ground state, which is also normalizable since the cutoff dependence cancels. On the torus, however, the ground state is degenerate, and it will prove essential to ascertain that the p linearly independent wave functionals can indeed be chosen to be orthonormal.

For the proof, we return to the Bargmann inner product. Because the χ and a dependences factorize, we can write Eq. (1.6) as the product

$$\begin{aligned}&\langle\psi_l(x_1,\ldots,x_N)|\psi_{l'}(x_1,\ldots,x_N)\rangle\\ &\quad=J_{ll'}(x_1,\ldots,x_N)K(x_1,\ldots,x_N).\end{aligned} \tag{3.10}$$

Here the first factor is the ordinary integral

$$\begin{aligned}J_{ll'}=&\int d[\overline{a},a]\exp\left(-\frac{k}{\pi}\int d^2x\left|a+i\frac{\pi}{k}\frac{Q}{V}Z\right|^2\right)\\ &\times\overline{\psi_l[a]}\psi_{l'}[a],\end{aligned} \tag{3.11}$$

and the second factor the functional integral

$$\begin{aligned}K(x_1,\ldots,x_N)=&\det{}'\Delta\int d[\phi]\exp\left(-\frac{k}{2\pi}\int d^2x\,\partial\phi\partial\phi\right.\\ &\left.-2\int d^2x\,\phi\rho_{\rm eff}\right),\end{aligned} \tag{3.12}$$

where the prime on $\det\Delta$ denotes the omission of the zero mode; furthermore, $\phi=\mathrm{Im}\,\chi$ and

$$\rho_{\rm eff}(x)=\sum_n q_n\delta^2(x-x_n)+\frac{1}{2}\left(\frac{Q}{L_1}\delta(x^2)+\frac{Q}{L_2}\delta(x^1)\right)-2\frac{Q}{V}. \tag{3.13}$$

We begin with the latter one, which could be interpreted as the vacuum expectation value of N (nonchiral) vertex operators with ''imaginary charges,'' were it not for the two additional terms in Eq. (3.13). The third derives from the b term in Eq. (2.19), whereas the second in brackets has its origin in the nasty boundary term $\phi[\chi]$ [see Eq. (2.18)]; remember that in Sec. II some work was required to get rid of it, but now we cannot avoid coming across this boundary term again. However, in the present context, both additional terms serve an important purpose because the second can be interpreted as an additional charge density of total charge Q, being concentrated on the boundary of the fundamental domain of the torus, and the third is a background charge so that

$$\int d^2x\,\rho_{\rm eff}(x)=0 \tag{3.14}$$

holds. Hence the theory by itself manages to make the quantum field-theoretic system, as described by the functional integral (3.12), overall neutral. It is in this way that the verdict, according to which the total charge on a closed surface must vanish, is avoided. The functional integration can now be done with the result

$$\begin{aligned}K(x_1,\ldots,x_N)=&(\det{}'\Delta)^{1/2}\exp\left(\frac{2\pi}{k}\int\int d^2x\,d^2y\,\rho_{\rm eff}(x)\right.\\ &\left.\times G(x,y)\rho_{\rm eff}(y)\right),\end{aligned} \tag{3.15}$$

where again it makes no difference which of the two propagators we choose. The argument of the exponential is an ambiguous expression; we will discuss regularization issues below.

What remains is the a integral (3.11), which can be reduced to the orthogonality relations of the θ functions. A rather long calculation is required to show that all the individual exponential prefactors in Eq. (2.37) serve their special purpose in obtaining the result

$$J_{ll'}(x_1,\ldots,x_N)=\mathcal{N}\delta_{ll'}\exp\biggl(-\frac{2\pi}{k}\frac{Q}{V}\sum_n q_n|z_n|^2+\frac{\pi}{2k}\frac{Q}{V}\sum_n q_n(z_n^2+\overline{z}_n^2)\biggr)\times\exp\biggl(\frac{\pi}{2k}\frac{Q^2}{V}(L\overline{Z}+\overline{L}Z)-\frac{\pi}{2k}\frac{Q^2}{V}\times(Z-\overline{Z})^2\biggr)|\varphi(x_1,\ldots,x_N)|^2, \tag{3.16}$$

with $\mathcal{N}$ an l-independent normalization constant. On combining Eqs. (3.16) and (3.15) according to Eq. (3.10), where boundary contributions are ignored, one finds

$$\exp\biggl(\frac{2\pi}{k}\int\int d^2x\,d^2y\,\rho(x)G_0(x,y)\rho(y)\biggr)J_{ll'}(x_1,\ldots,x_N)=\mathcal{N}(\det{}'\Delta)^{1/2}\exp\biggl(-\frac{\pi}{k}\frac{Q}{V}\sum_n q_n|z_n|^2+\frac{\pi}{2k}\frac{Q}{V}\times\sum_n q_n(L_{\overline{z}_n}+\overline{L}_{z_n})\biggr)\varepsilon^{-(1/k)\Sigma_n q_n^2}\prod_{n\neq m}\biggl|\times E\biggl(\frac{1}{L_1}(z_n-z_m)\biggr)\biggr|^{-q_nq_m/k}|\varphi(x_1,\ldots,x_N)|^2\delta_{ll'}\,. \tag{3.17}$$

In particular, for the ground-state wave function (2.40), the factors with the product of prime forms and the ultraviolet regulator are cancelled, and so only the square root of the regularized determinant of the Laplacian together with the exponential prefactor remain. Thus the ground states have a finite, cutoff-independent, norm, and can be assumed to be orthonormal, because for $q_n=1$ the final x_n integrations are (incomplete) Gaussian integrals which can be done, at least in principle.

IV. QUANTUM SYMMETRY

There is a relic of large gauge transformations, which will prove to be of crucial importance in the following. This comes about since the invariance under large gauge transformations has not fully been exploited so far, and the operators $\hat{U}_{(m_1,m_2)}$ still act as symmetries of the system.

We simplify the discussion by limiting ourselves to a case relevant for the quantum Hall effect. Hence we choose $q_n=1$ for all n, i.e., $Q=N$, and restrict the Chern-Simons coupling constant to values $k=1/p$, with p an odd integer. Omitting the χ dependence, the ground states [see Eq. (2.37)] then read

$$\psi_{0;l}[a](x_1,\ldots,x_N)=\exp\biggl(-\frac{k}{2\pi}Va^2-i\sum_n(z_n-\overline{z_n})a-\frac{\pi}{k}\frac{N}{V}\sum_n|z_n|^2+\frac{\pi}{2k}\frac{N}{V}\sum_n z_n^2\biggr)\times\exp\biggl[-\frac{i}{2}N\overline{L}\biggl(a+i\frac{\pi}{k}\frac{1}{V}\sum_n z_n\biggr)\biggr]\times\theta\begin{bmatrix}\frac{1}{p}\left(l+\eta_2+\frac{Np}{2}\right)\\ \eta_1+\frac{Np}{2}\end{bmatrix}\times\biggl[-i\frac{L_2}{\pi}\biggl(a+i\frac{\pi}{k}\frac{1}{V}\sum_n z_n\biggr)\biggg|p\tau\biggr]\times\varphi_0(x_1,\ldots,x_N). \tag{4.1}$$

As generators of the discrete Heisenberg-Weyl group [cf. Eqs. (2.9) and (2.10)], the two operators

$$S=\hat{U}^{-1}_{(0,1)},\quad T=\hat{U}^{-1}_{(1,0)} \tag{4.2}$$

are selected, obeying the relation

$$TS=qST, \tag{4.3}$$

with $q=\exp(2\pi i/p)$. On using known properties of θ functions, one can show by means of the explicit form (2.23) that they act on the ground-state wave functions as

$$S\psi_{0;l}[a](x_1,\ldots,x_N)=q^{l+\eta_2}\psi_{0;l}[a](x_1,\ldots,x_N), \tag{4.4}$$

$$T\psi_{0;l}[a](x_1,\ldots,x_N)=q^{\eta_1}\psi_{0;l-1}[a](x_1,\ldots,x_N). \tag{4.5}$$

Hence the transformation of the variable a may be rewritten so as to result in a unitary transformation of the basis of ground-state wave functions, and we can legitimately set a equal to zero.

These (Verlinde-type) operators are the building blocks of a quantum enveloping algebra,[15] as we will demonstrate now.[64] On passing from the discrete Heisenberg-Weyl group to its group algebra, it makes sense to form linear combinations of the generating elements S and T and products thereof. In particular, we choose

$$J_{\pm}=\frac{q^{\mp 1/2}S-q^{\pm 1/2}S^{-1}}{q-q^{-1}}T^{\mp 1},\quad K=-S^2, \tag{4.6}$$

and it is straightforward to prove that these operators obey the defining relations of the quantum algebra $U_q(su_2)$, i.e.,

$$[J_+,J_-]=\frac{K-K^{-1}}{q-q^{-1}},\quad KJ_{\pm}K^{-1}=q^{\pm 2}J_{\pm}. \tag{4.7}$$

On the p-dimensional basis of ground states $|0;l\rangle$, they act as

$$J_{\pm}|0;l\rangle=\left[l\pm\frac{1}{2}\right]|0;l\pm1\rangle,\quad K|0;l\rangle=-q^{2l}|0;l\rangle \tag{4.8}$$

for $\eta_1=\eta_2=0$,[65] where $[x]=(q^x-q^{-x})/(q-q^{-1})$. As we know from Sec. III, the basis can be chosen to be orthonormal; furthermore, the operators of large gauge transformations are unitary with respect to the Bargmann inner product. Accordingly, the quantum algebra generators $J_\pm$ and K are Hermitian in the sense $J_\pm^\dagger=J_\mp$ and $K^\dagger=K^{-1}$. Representation (4.8) is indeed compatible with these Hermiticity properties.

Up to now, we have not assumed the algebra, defined by Eq. (4.7), to be equipped with a quasitriangular structure. If the deformation parameter is a pure phase, or even a root of unity, it is a subtle problem to provide for a $*$ structure of $U_q(su_2)$ as a quasitriangular Hopf algebra. The standard definition (Ref. 66, see also Ref. 15) only covers the case that q takes real values. This problem can be solved[63] by means of a modification of the standard $*$-structure. Hence it may be taken for granted that the representation (4.8) is indeed unitary for $U_q(su_2)$ as a quasitriangular Hopf algebra, and not as an algebra only.

Let us mention that the representations encountered in the present context are somewhat special because they have quantum dimension 0, i.e., $\mathrm{tr}\,K=0$, and as such are often regarded as ‘‘unphysical’’ (cf. Ref. 16).

Having provided for a quantum group structure of the ground-state wave functionals, we would like to obtain something similar for the wave functions (2.40) and (2.41) as well. This can be achieved by absorbing the transformation of the variable a in the center of charge coordinate Z by replacing the z_n simultaneously by $z_n'=z_n-(L_1/pN)(m_1-\tau m_2)$, and one finds

$$\begin{aligned}\hat{U}_{(m_1,m_2)}\psi_{0;l}[a](x_1,\dots,x_N)&=\exp\Bigg(-\frac{\pi}{L_2}\sum_n \overline{z_n'}(m_2-\tau m_1)+\frac{\pi}{2}\frac{N}{L_2}(m_2-\overline{\tau}m_1)L\\&\quad-\frac{\pi}{2p}\frac{L_1}{L_2}(m_2-\overline{\tau}m_1)(m_2-\tau m_1)\Bigg)\\&\quad\times\psi_{0;l}[a](x_1',\dots,x_N').\end{aligned}\tag{4.9}$$

We are enabled now to set $a=0$, thus obtaining for the generators

$$S\psi_{0;l}(x_1,\dots,x_N)=\exp\left(\frac{\pi}{L_2}\sum_n \overline{z_n'}-\frac{\pi}{2}\frac{N}{L_2}L-\frac{\pi}{2p}\frac{L_1}{L_2}\right)\times\psi_{0;l}(x_1',\dots,x_N'),\tag{4.10}$$

with $z_n'=z_n+(L_1/pN)$, and

$$T\psi_{0;l}(x_1,\dots,x_N)=\exp\left(-i\frac{\pi}{L_1}\sum_n \overline{z_n'}-i\frac{\pi}{2}\frac{N}{L_1}L-\frac{\pi}{2p}\frac{L_2}{L_1}\right)\psi_{0;l}(x_1',\dots,x_N'),\tag{4.11}$$

with $z_n'=z_n-i(L_2/pN)$. The effect on the x_n coordinates may be looked upon as providing the torus with a lattice structure, where L_1/pN and L_2/pN are the lattice constants.[67]

As a check, the ‘‘quantum plane’’ relations (4.3) are easily verified for the new version [Eqs. (4.10) and (4.11)] of the generators, from now on being taken as the definition for the operators of large ‘‘diffeomorphisms.’’ An acid test consists in showing that indeed they respect the condition [see Eq. (2.28)]

$$\begin{aligned}\hat{U}_{p(m_1,m_2)}\psi_{0;l}(x_1,\dots,x_N)&=\exp[-2\pi i(m_1\eta_1+m_2\eta_2)\\&\quad-\pi i p m_1 m_2]\psi_{0;l}(x_1,\dots,x_N),\end{aligned}\tag{4.12}$$

which may also be understood as a quasiperiodicity requirement.

Finally, let us turn to the case we are ultimately interested in, the quantum Hall effect. We may be brief since most of the work has already been done. On adding the classical B field, we use the background field method.[68] Then the small and large gauge transformations leave the classical B part untouched so that all the results of Sec. II concerning the χ and a dependences, remain valid. Only the determination of the ground-state wave function requires modification, but we can follow similar steps as in Sec. I, to obtain

$$\begin{aligned}\varphi_0(x_1,\dots,x_N)&=\mathcal{N}_0\prod_n \varepsilon^{Np/2}\\&\quad\times\exp\left(-\frac{1}{4\ell^2}\sum_n |z_n|^2\right)\\&\quad\times\prod_{n<m}E\left(\frac{1}{L_1}(z_n-z_m)\right)^p.\end{aligned}\tag{4.13}$$

After all, the ground-state wave functions of the quantum Hall effect with natural boundary conditions thus turn out to be

$$\begin{aligned}\psi_{0;l}(x_1,\dots,x_N)&=\exp\Bigg(-\pi\frac{Np}{V}\sum_n |z_n|^2+\frac{\pi}{2}\frac{Np}{V}\sum_n z_n^2\\&\quad+\frac{\pi}{2}\frac{Np}{V}L\sum_n z_n\Bigg)\theta\begin{bmatrix}\frac{1}{p}\left(l+\frac{Np}{2}\right)\\ \frac{Np}{2}\end{bmatrix}\\&\quad\times\left(p\frac{1}{L_1}\sum_n z_n\middle| p\tau\right)\varphi_0(x_1,\dots,x_N),\end{aligned}\tag{4.14}$$

where, for definiteness (see Sec. III), we have chosen N to be odd.

The quantum group properties, having been established above, remain valid up to one qualification. Only the quasiperiodicity properties (4.10) and (4.11) take a slightly different form, because there is an additional contribution from

$$\varphi_0(x_1,\ldots,x_N)=\exp\Bigg(-\frac{\pi}{2}\frac{1}{L_2}\sum_n\,[(m_2-\bar{\tau}m_1)z_n+\bar{z}_n \times(m_2-\tau m_1)]+\frac{\pi}{2p}\frac{L_1}{L_2}(m_2-\bar{\tau}m_1) \times(m_2-\tau m_1)\Bigg)\varphi_0(x_1',\ldots,x_N'). \quad (4.15)$$

Here we have identified Np with the degree of degeneracy of Landau's theory, i.e., $Np=V/2\pi\ell^2$. But again, it is the action of the operators S and T [cf. Eqs. (4.4) and (4.5)]

$$S|0;l\rangle=q^l|0;l\rangle,\quad T|0;l\rangle=|0,l-1\rangle, \quad (4.16)$$

which counts since this result entails Eq. (4.8), which is the essential property. Unitarity, however, remains an issue since our attempts to orthonormalize the wave functions (4.14) have been unsuccessful. To summarize, the p-fold degenerate ground state of the quantum Hall effect can be equipped with a quantum symmetry;[69] hence the quantum ''group'' $U_q(sl_2)$, with q an odd root of unity, turns out to be the spectrum generating algebra of the quantum Hall effect.

V. CONCLUSION

Let us give a resume of what has been shown. There is no dispute that Laughlin's variational wave function captures the essential properties of the quantum Hall system. Hence the assumption made in Sec. I that the electrons themselves must be coupled to the Chern-Simons quantum field for rather ''kinematical'' reasons is indeed the missing link since, in the absence of interactions, it opens the way for a consistent derivation as an exact ground state. But the quantum number p, which determines the filling fraction $\nu=1/p$, remains more or less unexplained.

A realistic theory of the quantum Hall effect, however, must take into account the finite extent of the Hall sample, and for this reason we have investigated the generalization to the torus. Thus the final results [Eqs. (4.13) and (4.14)] for the Hall ground-state wave functions, differing in the exponential prefactor and in the characteristics of the θ function from what has been proposed by Haldane and Rezayi[37] and others, should be an improvement of the Laughlin ground state for natural boundary conditions.

At this point one is faced with the, at first sight, strange fact that the ground state is p-fold degenerate; hence it is the degree of degeneracy, which now determines the filling fraction. But, at second sight, a quantum group turns out to provide for the explanation since it discriminates between the different ground states, and thus the organizing principle behind is a quantum symmetry, as opposed to a standard Lie group symmetry.

We may revert the argument on taking the quantum group to be given a *priori*. Then the possible quantum numbers are obtained by classifying the relevant representations of the deformation $U_q(su_2)$. This investigation has been done[63] with the result that, for an odd root of unity $q=\exp(2\pi i/p)$, there is just one unitary and irreducible representation with vanishing quantum dimension for each given value of p. It is in this sense, that a quantum symmetry provides for an explanation of the quantum numbers for the experimentally observed values of the filling fraction $\nu=1/p$, with p an odd integer.

Of course, many problems remain. The most serious omission is that we have not touched at all the effect of the repulsive Coulomb interaction, which should lead to a stabilization of the system at the filling fractions $\nu=1/p$, i.e., give rise to the formation of the Hall plateaus. Furthermore, the above treatment of excitations is insufficient; as will be shown elsewhere, they may be given a satisfactory explanation, which also provides for a representation of the braid group on the torus,[70] and thus yields the relation to knot theory. The ultimate aim, however, will be a second quantized description of the matter particles at nonzero temperature, with an eye toward supplying for a derivation of the effective Chern-Simons action,[23,26,25] which has been proposed as an analog of the Ginzburg-Landau effective action[71] of superconductivity, from microscopic properties.

[1] For a review, see *The Quantum Hall Effect*, edited by R. E. Prange and S. M. Girvin (Springer, New York, 1990); *Quantum Hall Effect*, edited by M. Stone (World Scientific, Singapore 1992); A. Karlhede, S. A. Kivelson, and S. L. Sondhi, in Correlated Electron Systems (*Jerusalem Winter School for Theoretical Physics*), edited by V. J. Emery (World Scientific, Singapore, 1993).

[2] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980); D. C. Tsui, H. L. Störmer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

[3] A. Belavin, A. Polyakov, and A. Zamolodchikov, Nucl. Phys. B **241**, 33 (1984); for a review, see P. Ginsparg, in *Fields, Strings and Critical Phenomena, Les Houches 1988*, edited by E. Brezin and J. Zinn-Justin (North-Holland, Amsterdam, 1990).

[4] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983); for a review, see R. B. Laughlin, in *The Quantum Hall Effect* (Ref. 1).

[5] S. Fubini, Int. J. Mod. Phys. A **5**, 3553 (1990); S. Fubini and C. A. Lütken, Mod. Phys. Lett. A **6**, 457 (1991); G. Cristofano, G. Maiella, R. Musto, and F. Nicodemi, Phys. Lett. B **262**, 88 (1991); Nucl. Phys. B **33C**, 119 (1993); G. Moore and N. Read, *ibid.* **360**, 362 (1991); X. G. Wen and Y. S. Wu, *ibid.* **419**, 455 (1994); M. A. I. Flohr, Mod. Phys. Lett. A **11**, 55 (1996).

[6] D. Arovas, J. R. Schrieffer, and F. Wilzcek, Phys. Rev. Lett. **53**, 722 (1984).

[7] For a review, see *Fractional Statistics and Anyon Superconductivity*, edited by F. Wilczek (World Scientific, Singapore, 1990).

[8] For a review, see P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1986).

[9] X. G. Wen, Phys. Rev. B **41**, 12 834 (1990); M. Stone, Ann. Phys. (N.Y.) **207**, 38 (1991); A. Capelli, G. V. Dunne, C. A. Trugenberger, and G. R. Zemba, Nucl. Phys. B **398**, 531 (1993); J. Fröhlich and U. M. Studer, Rev. Mod. Phys. **65**, 733 (1993).

[10] R. Floreanini and R. Jackiw, Phys. Rev. Lett. **59**, 1873 (1987); J. Sonnenschein, Nucl. Phys. B **309**, 752 (1988); M. Stone, *ibid.* **327**, 399 (1989).

[11] E. Witten, Commun. Math. Phys. **121**, 351 (1989).

[12] G. Moore and N. Seiberg, Phys. Lett. B **220**, 422 (1989); S. Elitzur, G. Moore, A. Schwimmer, and N. Seiberg, Nucl. Phys. B **326**, 108 (1989); G. Dunne, R. Jackiw, and C. Trugenberger, Ann. Phys. (N.Y.) **194**, 197 (1989).

[13] M. Bos and V. P. Nair, Int. J. Mod. Phys. A **5**, 959 (1990).

[14] E. Guadagnini, M. Martinelli, and M. Mintchev, Nucl. Phys. B **336**, 581 (1990); J. M. F. Labistada, P. M. Llatas, and A. V. Ramallo, Nucl. Phys. B **348**, 651 (1991).

[15] For a review, see V. Chari and A. Pressley, *A Guide to Quantum Groups* (Cambridge University Press, Cambridge, 1994); S. Majid, *Foundations of Quantum Group Theory* (Cambridge University Press, Cambridge, 1995).

[16] L. Alvarez-Gaumé, C. Goméz, and G. Sierra, Nucl. Phys. B **330**, 347 (1990); C. Goméz and G. Sierra, Nucl. Phys. B **352**, 791 (1991).

[17] A. Lerda and S. Sciuto, Nucl. Phys. B **401**, 613 (1993); R. Caracciolo and A. R. Monteiro, Phys. Lett. B **308**, 58 (1993).

[18] I. I. Kogan, Int. J. Mod. Phys. A **9**, 3887 (1994).

[19] S. M. Girvin, in *The Quantum Hall Effect* (Ref. 1).

[20] F. D. M. Haldane, in *The Quantum Hall Effect* (Ref. 1).

[21] F. D. M. Haldane, Phys. Rev. Lett. **51**, 605 (1983).

[22] S. A. Trugman and S. Kivelson, Phys. Rev. B **31**, 5280 (1985).

[23] S. M. Girvin and A. H. MacDonald, Phys. Rev. Lett. **58**, 1252 (1987).

[24] C. L. Kane, S. Kivelson, D. H. Lee, and S. C. Zhang, Phys. Rev. B **43**, 3255 (1991); S. C. Zhang, Int. J. Mod. Phys. B **6**, 25 (1992).

[25] A. Karlhede *et al.*, in *The Quantum Hall Effect* (Ref. 1).

[26] S. C. Zhang, T. H. Hansson, and S. Kivelson, Phys. Rev. Lett. **62**, 82 (1989); N. Read, *ibid.* **62**, 86 (1989).

[27] A. Lopez and E. Fradkin, Phys. Rev. Lett. **96**, 2126 (1992).

[28] A. Lopez and E. Fradkin, Phys. Rev. B **44**, 5246 (1991); **47**, 7080 (1993); B. I. Halperin, P. A. Lee, and N. Read, *ibid.* **47**, 7312 (1993); S. H. Simon and B. Halperin, *ibid.* **50**, 1807 (1994); **50**, 1823 (1994); see also E. Fradkin, *Field Theory of Condensed Matter Systems* (Addison-Wesley, Redwood City, CA, 1991).

[29] J. K. Jain, Phys. Rev. B **40**, 8079 (1989); Adv. Phys. **41**, 105 (1992).

[30] D. N. Sheng, Z. B. Su, and B. Sakita, Int. J. Mod. Phys. B **5**, 1715 (1991).

[31] C. B. Hanna, R. B. Laughlin, and A. L. Fetter, Phys. Rev. B **40**, 8745 (1989); Y. H. Chen, F. Wilczek, E. Witten, and B. I. Halperin, Int. J. Mod. Phys. B **3**, 1001 (1989); A. L. Fetter and C. B. Hanna, Phys. Rev. B **45**, 2335 (1992).

[32] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University Press, New York, 1964); A. Hanson, T. Regge, and C. Teitelboim, and K. Sundermeyer, *Constrained Dynamics* (Springer, Berlin, 1982); A. Ashtekar, *Lectures on Non-Perturbative Canonical Gravity*, (World Scientific, Singapore, 1991); R. Jackiw, Nucl. Phys. B **18A**, 107 (1991); M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems* (Princeton University Press, Princeton, 1992).

[33] L. Faddeev and R. Jackiw, Phys. Rev. Lett. **60**, 1692 (1988); R. Jackiw, in *Constraint Theory and Quantization Methods*, edited by F. Colomo, L. Lusanna, and G. Marmo (World Scientific, Singapore, 1994).

[34] V. G. Knizhnik and A. B. Zamolodchikov, Nucl. Phys. B **247**, 83 (1984).

[35] R. Jackiw, in *Field Theory and Particle Physics*, edited by O. Eboli *et al.* (World Scientific, Singapore, 1990); B. Hatfield, *Quantum Field Theory of Point Particles and Strings* (Addison-Wesley, Redwood City, CA, 1992).

[36] R. Morf and N. d'Ambrumenil, Phys. Rev. Lett. **74**, 5116 (1995); J. K. Jain and R. K. Kamilla, Phys. Rev. B **55**, R4895 (1997); **55**, 9824 (1997); **55**, R13 417 (1997).

[37] F. D. M. Haldane and E. H. Rezayi, Phys. Rev. B **31**, 2529 (1985).

[38] R. Tao and Y. S. Wu, Phys. Rev. B **30**, 200 (1984); J. E. Avron and R. Seiler, Phys. Rev. Lett. **54**, 259 (1985); J. E. Avron and L. G. Yaffe, *ibid.* **56**, 2084 (1986); Q. Niu and D. J. Thouless, Phys. Rev. B **35**, 2118 (1987); X. G. Wen and Q. Niu, *ibid.* **41**, 9377 (1990); X. G. Wen, E. Dagotto, and E. Fradkin, *ibid.* **42**, 6110 (1990); A. Kol and N. Read, *ibid.* **48**, 8890 (1993).

[39] X. G. Wen, Int. J. Mod. Phys. B **4**, 239 (1990).

[40] M. G. G. Laidlaw and C. M. DeWitt, Phys. Rev. D **3**, 1375 (1971); J. S. Dowker, J. Phys. A **5**, 936 (1972); **18**, 3531 (1985); J. M. Leinaas and J. Myrheim, Nuovo Cimento **37**, 1 (1977); Y. S. Wu, Phys. Rev. Lett. **52**, 2103 (1984).

[41] D. P. Arovas, in *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989).

[42] G. W. Semenoff, Phys. Rev. Lett. **61**, 517 (1988).

[43] R. Jackiw and S. Y. Pi, Phys. Rev. D **42**, 3500 (1990).

[44] D. Boyanowsky, E. T. Newman, and C. Rovelli, Phys. Rev. D **45**, 1210 (1992).

[45] M. Lüscher, Nucl. Phys. B **326**, 557 (1989); E. Fradkin, Phys. Rev. Lett. **63**, 322 (1989).

[46] J. D. Romano and R. S. Tate, Class. Quantum Grav. **6**, 1487 (1989); K. Schleich, *ibid.* **7**, 1529 (1990).

[47] The conventions are $z=x^1+ix^2$, $\partial_z=1/2(\partial_1-i\partial_2)$ and $A_{\bar z}^-=1/2(A_1+iA_2)$.

[48] T. Kohno, Ann. Inst. Fourier **37**, 139 (1987); V. G. Drinfel'd, in *Problems of Modern Quantum Field Theory*, edited by A. A. Belavin, A. U. Klimyk, and A. B. Zamolodchikov (Springer, Berlin, 1889).

[49] E. Verlinde, in *Correlated Electron Systems* (Ref. 1).

[50] Had we followed well-trodden paths and eliminated the Chern-Simons field by means of the constraint, we would have obtained the symmetric form, i.e., the anyon result.

[51] V. S. Dotsenko and V. A. Fateev, Nucl. Phys. B **240**, 312 (1984).

[52] R. Iengo and K. Lechner, Nucl. Phys. B **364**, 551 (1991); Phys. Rep. **213**, 179 (1992); C. L. Ho and Y. Hosotani, Int. J. Mod. Phys. A **7**, 5797 (1992); A. Fayyazuddin, Mod. Phys. Lett. A **8**, 3185 (1993); M. Bergeron, D. Eliezer, and G. Semenoff, Phys. Lett. B **311**, 137 (1993).

[53] G. Grensing, Phys. Lett. B **419**, 258 (1998).

[54] C. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) **59**, 42 (1970).

[55] A. M. Polyakov, *Gauge Fields and Strings* (Harwood, London, 1987).

[56] D. Mumford, *Tata Lectures on Theta*, Progress in Mathematics Vol. 28 (Birkhäuser, Boston, 1983); J. Igusa, *Theta Functions* (Springer, Berlin, 1972).

[57] E. Verlinde and H. Verlinde, Nucl. Phys. B **288**, 357 (1987).

[58] S. Randjbar-Daemi, A. Salam, and J. A. Strathdee, Int. J. Mod. Phys. A **2**, 667 (1987).

[59] Since the notation $\psi[A_{\bar z}^-]|_{\chi=0}(x_1,\ldots,x_n)=\psi[a](x_1,\ldots,x_n)$ is used, we prefer, for typographical reasons, to write ∂_a as a functional derivative.

[60] Consequently, the operator of small gauge transformations, as well as the operator of large gauge transformations (2.23) below, automatically provide for a holomorphic square root in the sense of Ref. 13.

[61] It is noteworthy that the classical analysis is misleading because it yields $\text{Im}\, b = 1/2(\pi/k)(Q/V)$, which is wrong by a factor 1/2 and, moreover, leaves the real part of b undetermined.

[62] It deserves mention that the Haldane-Rezayi ground-state wave functions may be read off from Eq. (3.25) of Bos and Nair[13] for genus 1, a result which is explicitly restricted to the case of total charge zero. Compare also a corresponding footnote of Bergeron *et al.*[52]

[63] G. Grensing, J. Math. Phys. **39**, 6153 (1998).

[64] Another case of interest is the pure Abelian Chern-Simons theory on a Riemann surface of genus g. This theory is exactly solvable, and the solutions can be shown[53] to span the representation space of the quantum algebra $U_q(su_{g+1})$; its generators may be expressed in terms of the holonomy operators [see Eq. (2.27)] for the $2g$ independent homology cycles.

[65] For arbitrary values of the phases, the representations generally belong to the cyclic series. But there is one important exception: if the phases take the values $\eta_1 = -\eta_2 = 1/2$, then we also have modular invariance (see Sec. III), and the representation is again equivalent to Eq. (4.8), which is of highest weight.

[66] T. Masuda, K. Mimachi, Y. Nakagami, M. Noumi, Y. Saburi, and K. Ueno, Lett. Math. Phys. **19**, 187 (1990).

[67] Related operators appear in the one-particle Landau theory on the torus,[18,63] where they derive from magnetic translations; operators (4.10) and (4.11) can be seen as the generalizations to the N-particle case.

[68] G. 't Hooft, Acta Universitatis Wratislaviensis **368**, 345 (1976); B. S. De Witt, in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford University Press, Oxford, 1981); L. F. Abbott, Nucl. Phys. B **185**, 189 (1981).

[69] A connection between noncommutative geometry and the quantum Hall effect was pointed out by J. Bélissard, in *K-Theory of C*-algebras in Solid State Physics*, Springer Lecture Notes in Physics 257 (Springer, Berlin, 1986), and A. Connes, *Noncommutative Geometry* (Academic, New York, 1994); but the investigation of these authors is rather different from the present approach, and restricted to the integer effect.

[70] T. Einarsson, Phys. Rev. Lett. **64**, 1995 (1990).

[71] L. P. Gorkov, Zh. Éksp. Teor. Fiz. **36**, 1918 (1959) [Sov. Phys. JETP **9**, 1364 (1959)]; A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1963).