$\mathcal{F}$-invariance and its application to the quantum Hall effect

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus
prof. dr J.J.M. Franse
ten overstaan van een door het college voor promoties ingestelde
commissie in het openbaar te verdedigen in de Aula der Universiteit
op donderdag 20 mei 1999 te 10.00 uur.

door

Boris Škorić

geboren te Amsterdam

Promotor: prof. dr A.M.M. Pruisken

Promotiecommissie: prof. dr B.L. Altshuler
prof. dr P.J. van Baal
prof. dr ir F.A. Bais
dr M.A. Baranov
prof. dr A.M. Finkelstein
prof. dr K. Schoutens
prof. dr Ch.G. van Weert

Faculteit der Wiskunde, Informatica, Natuur- en Sterrenkunde
Instituut voor Theoretische Fysica

This thesis is based on the following papers:

- A.M.M. Pruisken, M.A. Baranov and B. Škorić: "(Mis)handling gauge invariance in the theory of the quantum Hall effect I: Unifying action and the $\nu = 1/2$ state", cond-mat/9712322, to appear in Phys. Rev. B.
- M.A. Baranov, A.M.M. Pruisken and B. Škorić: "(Mis)handling gauge invariance in the theory of the quantum Hall effect II: Perturbative results", cond-mat/9712323, to appear in Phys. Rev. B.
- A.M.M. Pruisken, B. Škorić and M.A. Baranov: "(Mis)handling gauge invariance in the theory of the quantum Hall effect III: The instanton vacuum and chiral edge physics", cond-mat/9807241, to appear in Phys. Rev. B.
- B. Škorić and A.M.M. Pruisken: "The fractional quantum Hall effect: Chern-Simons mapping, duality, Luttinger liquids and the instanton vacuum", cond-mat/9812437, submitted to Nucl. Phys. B.

Cover design by Jonathan Ross.
Image: the Har crater on Callisto, taken by NASA's Galileo probe on June 25, 1997.

ISBN 90-5776-022-3

Contents

ABSOLUTELY CORRECT, snapped Death.
CLARITY OF THOUGHT. REALISTIC APPROACH.
VERY IMPORTANT IN A JOB LIKE OURS.

– Terry Pratchett, *Mort*

Chapter 1

Introduction

1.1 Subject

The quantum Hall effect (qHe) is a remarkable phenomenon that has already led to the awarding of two Nobel prizes. These are well deserved: the discovery of the Hall conductance quantisation in a two dimensional electron gas by von Klitzing [39] in 1980 [1] came as a big surprise and necessitated a complete revision of the theory of electronic transport and Anderson localisation. Then, just when the dust had settled and everybody believed that two dimensional electron systems were truly understood, this illusion was shattered in 1982 by the discovery[2] of the fractional quantum Hall effect (fqHe) by Tsui, Störmer and Gossard [73]. An enormous number of publications, both experimental and theoretical, has followed in the years since these discoveries.

From a theoretical point of view the fqHe has proven to be quite challenging; new concepts have been introduced and ideas have been borrowed from high energy physics and even string theory. Whereas the integer effect can be understood in terms of the localised vs. extended character of the single particle states of noninteracting 2D electrons, what is needed for the fractional effect is a full understanding of Coulomb interactions. A successful theory for the fqHe has to take into account all the relevant ingredients at the same time: disorder *and* interactions *and* an electromagnetic field, a highly nontrivial combination. The Coulomb interactions invalidate the single-particle formalism which is conventionally used to study electronic disorder, while the disorder breaks the translational symmetry that is usually exploited in tackling the interaction problem. The strong magnetic field further complicates the matter. It is therefore not surprising that such a theory has not immediately emerged. What *has* emerged over the years is a collection of less ambitious theories and phenomenological models for the fqHe that, with varying success, try to tackle the problem by concentrating on a subset of the ingredients. Three of the most successful and best known of these are Laughlin's wave function [41], Jain's composite fermion picture [34] together with the Chern-Simons approach [21, 32, 38, 46, 47, 81, 85, 23, 79, 84], and effective edge theories [24, 25, 26, 8, 77]. Each of these approaches captures an essential part of the physics. However, they all fail or at least get highly ambiguous when one tries to include the missing ingredient(s).

The holy grail of quantum Hall theory would be a unified description in which the three above theories are incorporated at the outset. Also, it would have to comprise the

[1] Nobel prize 1985

[2] 1998 Nobel prize to Tsui, Störmer and Laughlin

iqHe and fqHe in a single formalism; the similarity of the integral and fractional effect simply begs for that. How to start the search? A crucial observation is that disorder has always been among the missing ingredients. This is quite depressing, because we know that it plays a fundamental role. Even though the fqHe occurs in samples that are very clean, it is still the disorder that causes the plateaus in the first place. Another disappointing fact is that none of the phenomenological approaches has anything to say on the subject of plateau transitions. For instance, what are the relevant quasiparticles at a transition between fractional plateaus?

The only theories that decently treat disorder, the nonlinear sigma model [74, 45, 58] (for free particles) and the Finkelstein theory [22] (for interacting particles), are applicable only to the integer effect. At least, that was the state of the art in 1995. When I started my PhD work, Pruisken and Baranov had just begun to unravel a hidden structure in Finkelstein's theory. Even though the theory does not contain the electromagnetic field, it turned out to have a symmetry akin to $U(1)$ gauge invariance, namely an invariance under a spatially constant shift of the plasmon field. This symmetry, which was dubbed '$\mathcal{F}$-invariance', together with a new truncation procedure of Matsubara frequency space, allowed us to include $U(1)$ gauge fields in the Finkelstein theory. This was an important development, because it opened up the possibility of using statistical or Chern-Simons gauge fields to 'map' the iqHe to the fqHe, thereby finally providing us with the long sought after unified description. In order to make sure that we had our hands on the right approach to the problem and weren't deluding ourselves with wishful thinking, we first had to check whether $\mathcal{F}$-invariance really is a good symmetry. This was by no means self evident, because of the new and unusual structure of what we call the '$\mathcal{F}$-algebra' of $U(1)$ transformations in truncated frequency space. The true trial by fire that we subjected the new symmetry to was a perturbative renormalisation group calculation to establish the renormalisability of the Finkelstein theory to two-loop order. As is well known in gauge theories, renormalisation procedures that do not respect all the symmetries in the problem lead to divergences. What we found was that the Finkelstein theory is indeed renormalisable at two loop level when the $\mathcal{F}$-algebra is used. Furthermore, the infrared behaviour of the theory can only be extracted from $\mathcal{F}$-invariant correlation functions.

Having proved that $\mathcal{F}$-invariance is a genuine symmetry, we were on solid footing to start putting our unified theory to further use. We constructed a combined action for both the bulk and the edge and studied the limiting cases of vanishing edge or bulk density of states. In this exercise the central role became apparent which the topological term plays, via the $\mathcal{F}$-algebra, in the description of edge states. Another important result was the clarification of the effects of disorder and Coulomb interactions at the edge, especially in the case of counter-flowing modes. Our work on edge states and their interplay with bulk states also explains the results of several recent, much discussed experiments. As a corrolary of our approach we have found, in our bulk theory as well as at the edge, that the composite fermion mapping procedure is only well-defined as long as the conditions for $\mathcal{F}$-invariance are met.

The construction of an $\mathcal{F}$-invariant unifying action and the study of its implications form the subject of this thesis. In this introductory chapter I first give a short overview of the basics of quantum Hall experiments and then of the three phenomenological approaches to the fqHe, listing their strengths and weaknesses. This is followed by a discussion of disorder, where I review the most important theoretical insights up to 1995.

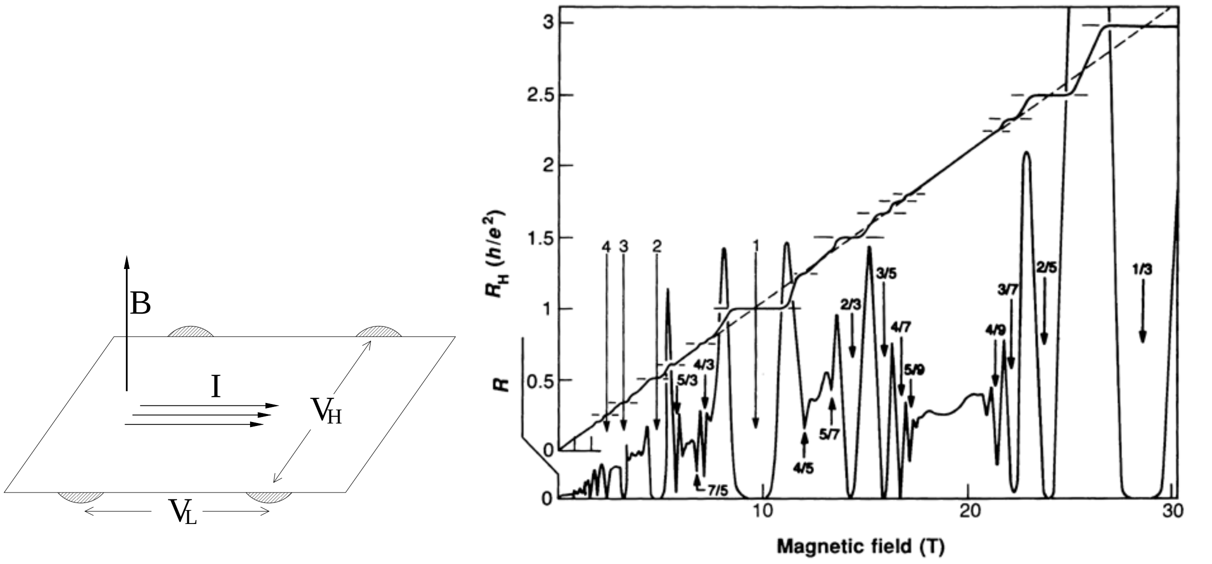

Figure 1.1: *Left: artist impression of a quantum Hall measurement. Right: Hall resistance and longitudinal resistance as a function of magnetic field. Values of ν are indicated with arrows. Figure taken from* [19].

1.2 Overview

1.2.1 Quantum Hall basics

The qHe occurs in two-dimensional electron gases at low temperatures ($\approx$100mK) and high magnetic fields (of the order of 10T). Von Klitzing's original experiments used Si/SiO_2 interfaces to constrain the degrees of freedom of the electrons to two dimensions. Most of the experiments since then have been done on GaAs/$Al_xGa_{1-x}As$ interfaces, because of the high mobility in these samples. A typical quantum Hall measurement is sketched in Fig. 1.1. A small current is sent through the electron gas. The longitudinal voltage V_{L} and the transverse or Hall voltage V_{H} are measured. This gives the longitudinal and Hall resistances $R_{\mathrm{L}} = V_{\mathrm{L}}/I$ and $R_{\mathrm{H}} = V_{\mathrm{H}}/I$. The state of a sample is characterised by its *filling fraction* $\nu = n_e h/(eB)$, with n_e the 2D electron density and B the perpendicular magnetic field strength. The filling fraction counts the number of filled Landau levels or, equivalently, the number of electrons present per magnetic flux quantum h/e. Depending on the sample quality, the inverse Hall resistance shows plateaus at integer and/or fractional multiples of e^2/h. A plateau $R_{\mathrm{H}}^{-1} = \frac{p}{q} \cdot \frac{e^2}{h}$ (with p an integer and q an odd integer) is situated around filling fraction $\nu = p/q$. While R_{H} is constant, the R_{L} curve goes to zero. The width of the plateaus depends on the amount of disorder. Very dirty samples have wide plateaus and only show the iqHe. With increasing cleanness the plateaus get narrower and more fractions start appearing. The quantisation of R_{H} in the iqHe can be as accurate as one part in 10^8. This fact has been exploited to define a standard of resistance based on the qHe. The qHe has also been used to perform an independent measurement of the fine structure constant α which appears in quantum electrodynamics. This is possible because α is given by $\alpha = \frac{1}{2}\mu_0 c \cdot \frac{e^2}{h}$, where the permeability of the vacuum μ_0 and the speed of light c are exact numbers.

Apart from the transport measurements described above, many other interesting experiments have been done involving tunnelling, thermoelectric power, spin configurations called skyrmions, double layer systems, hysteresis effects, the breakdown of the qHe as a function of current, temperature and disorder etc. Of these only tunnelling experiments will be discussed in this thesis. The electron spin will be completely neglected, since it will always be assumed that the magnetic field is strong enough to cause total spin polarisation. Also, nothing will be said about the $\nu = 5/2$ state.

A notational peculiarity has to be mentioned. Plateaus are customarily labelled with the value of ν instead of R_{H}^{-1}, even though ν varies while R_{H} stays constant.

1.2.2 Laughlin's wave function

In 1983, Laughlin [41] wrote down a variational wave function that represents the ground state of a clean system of interacting electrons at filling fractions $\nu = 1/(2p+1)$, with p a non-negative integer.

$$\Psi(z_1, \cdots, z_n) \propto \prod_{i<j} (z_i - z_j)^{2p+1} \exp -\tfrac{1}{2} \sum_k |z_k|^2 \tag{1.2.1}$$

(The $z_1, \cdots, z_n$ are the coordinates of the particles in the complex plane). He arrived at this result solely by demanding that the wave function is built from single particle functions in the lowest Landau level only (i.e. that the factor in front of the exp has to be analytic), that it is completely antisymmetric and that it is an eigenfunction of the total angular momentum. Laughlin's famous wave function shows that the Coulomb interaction forces the electrons in fqH systems to condense into a new type of *incompressible quantum fluid*. The elementary excitations of the fluid are quasiparticles with *fractional charge and statistics*. The charge of such a particle is $1/(2p+1)$ times the electron charge and the statistical angle is given by $\theta = \pi/(2p+1)$. (The electron has $\theta = \pi$). The excitation with unit charge has $\theta = 2p+1$.

For his work on the fqHe, especially for the introduction of the incompressible quantum liquid and quasiparticles with fractional charge and statistics, Laughlin shared the 1998 Nobel prize with two of the discoverers of the fqHe.

1.2.3 Composite fermions

In 1989 Jain introduced the beautifully simple picture of *composite fermions* [34]. In this picture the fqHe can be regarded as the occurrence of the iqHe in terms of composite particles that consist of an electron plus an even number $2p$ of magnetic flux quanta h/e attached to it. Having bound a part of the magnetic flux, the composite particles live in a reduced effective magnetic field $B_{\mathrm{eff}} = B - 2pn_e \cdot h/e$, with n_e the electron density. In this way a fractional ν for electrons can be related to an effective integer filling fraction m of composites,

$$\tfrac{1}{m} = \tfrac{1}{\nu} - 2p \quad ; \quad \nu = \tfrac{m}{2pm+1}. \tag{1.2.2}$$

The fractional fillings obtained in this way go by the name of 'the Jain hierarchy'. Almost all experimentally observed fractions fall in this hierarchy. In (1.2.2) the ν and p are positive. The effective integer filling m can be negative, in which case B_{eff} and B point in opposite directions.

Flux-charge composites in general can have any statistical angle and are therefore a realisation of *anyons* [81, 82, 44]. The name 'composite fermion' stems from the fact that the combination of unit charge and an even number of flux quanta yields a statistical angle that is an odd multiple of π, resulting in a 'fermionic' phase factor of -1 under particle exchange. Laughlin's wave function finds a natural interpretation as the ground state of composite fermions that completely fill the lowest Landau level.

Jain's proposal assumes that electron-electron repulsion is responsible for the formation of the composites and that the residual Coulomb interaction between composite fermions is very weak, so that they can be regarded as almost free particles. However, this statement has never been made quantitative. Jain's picture does not predict which type of composite will form under what circumstances, nor does it make a statement on the type of interaction needed to create composite fermions in the first place. Finally, it is assumed that composite fermions get localised by disorder in the same way as electrons, but there is no hint whether and how disorder affects the nature of the composites themselves. Even so, the composite fermion picture has proven very successful, certainly more so than what could be reasonably expected. For instance, large portions of conductance data as a function of ν in the fractional regime can be mapped very well onto data from the integer regime (when properly rescaled). At even denominator filling fractions it really seems as if there are particles in a vanishing B_{eff}, as shown by e.g. thermoelectric power experiments [72] at these fractions.

Statistical gauge fields

There exists a very nice way of implementing the flux attachment idea in 2+1 dimensional field theories, called the 'Chern-Simons trick'. Chern-Simons actions were originally studied in the context of massive gauge fields [16, 70]. It was found that they can also be used to transmute the statistics of particles. The Chern-Simons trick is based on the following observation [46]. Suppose that we have an action $S[\psi, A_\mu]$ for a fermionic or bosonic field ψ coupled to the electromagnetic field A_μ, then the statement is that *no observable is altered* if one makes the replacement

$$S[\psi, A_\mu] \longrightarrow S[\psi, A_\mu + a_\mu] + \tfrac{1}{2p}\tfrac{1}{4\pi}\int \mathrm{d}^3x \; \varepsilon^{\mu\lambda\kappa} a_\mu \partial_\lambda a_\kappa \tag{1.2.3}$$

(the last term is called Chern-Simons action) and integrates out a_μ. The parameter p has to be an integer for this statement to hold. On a mean field level, (1.2.3) attaches to every unit charge a number $2p$ of flux quanta associated with $\nabla \times \vec{a}$. The a_μ is called the 'statistical gauge field', since for arbitrary non-integer p the attachment of a_μ-flux changes the statistics of ψ. The statement (1.2.3) is made useful as follows. We want to derive a theory S_ν for a filling fraction ν that is an element of Jain's hierarchy, but the only thing we have is an action S_m for the corresponding integer filling m. Now suppose that we already have the sought after S_ν and perform the replacement (1.2.3). Then the resulting mean field theory will be precisely S_m, coupled to the *fluctuations* of a_μ according to (1.2.3). By integrating out these fluctuations the S_ν is reobtained. Only 're'obtained is not the right word in this case, since we didn't actually know S_ν to begin with.

1.2.4 Effective edge theories; chiral Luttinger liquids

The third important development, the formulation of effective edge theories, pioneered by Wen's chiral boson model [77], concentrates on the edge excitations of quantum Hall systems. Effective edge theories take the form of a chiral Fermi liquid in the integer regime or a chiral Luttinger liquid [49, 53] in the fractional regime. There are several ways in which it can be seen from the phenomenology of the Hall plateaus that edges play a prominent role when the Hall conductivity is quantised. One of these ways is the so-called hydrodynamic approach, which observes that the only possible gapless excitations of a 'droplet' of incompressible quantum fluid are ripples on its edge. The magnetic field constrains the ripples to move in one direction only; what results is a 1+1D effective theory in terms of relativistic chiral edge bosons. (The adjective 'relativistic' stems from the linear dispersion relation). A second, very interesting way is the Chern-Simons (C-S) approach. Fröhlich *et al.* [24, 25, 26] noted that the plateau equation of motion for the current density, as imagined in a sample free of disorder, is precisely the saddlepoint equation of a Chern-Simons action. Let us for simplicity consider the plateau around a Laughlin fraction 1/odd. The 2+1D current density j^μ satisfies

$$j^\mu = -\sigma_{\mathrm{xy}} \varepsilon^{\mu\lambda\kappa} \partial_\lambda A_\kappa, \tag{1.2.4}$$

where A_μ is the electromagnetic potential and σ_{xy} is the quantised Hall conductance. The associated C-S action is given by

$$S = \tfrac{1}{4\pi} \int \mathrm{d}^3x \; \varepsilon^{\mu\lambda\kappa} \left(-\tfrac{1}{\sigma_{\mathrm{xy}}} g_\mu \partial_\lambda g_\kappa + 2 g_\mu \partial_\lambda A_\kappa \right), \qquad j^\mu \propto \varepsilon^{\mu\alpha\beta} \partial_\alpha g_\beta. \tag{1.2.5}$$

The C-S action is topological; it has zero Hamiltonian. It was first shown by Witten [83] that the only dynamics of a general nonabelian C-S action occurs on the boundary of the integration manifold and takes the form of a WZW model. It can be shown that in the simple abelian case given above, integration over g results in a chiral boson action of the same form as in the hydrodynamic picture

$$S = \tfrac{1}{4\pi} \sigma_{\mathrm{xy}} \oint \mathrm{d}t\mathrm{d}x \; \partial_x \varphi (\partial_t \varphi - v \partial_x \varphi). \tag{1.2.6}$$

(Here the A_μ has been set zero). The velocity v is the speed of sound in the hydrodynamic approach. In the C-S approach it enters as a constant in a gauge fixing constraint that has to be imposed on the boundary. In both cases its value is arbitrary due to the lack of knowledge of the microscopic details, with the only restriction that v has to have the same sign as σ_{xy} in order for the Hamiltonian to be bounded from below.

Arbitrary filling fractions

The scenario sketched above has been generalised to more complicated filling fractions. Several incompressible quantum fluids can be taken or, equivalently, several independently conserved currents j. This leads to a generalisation of (1.2.6) with a number of bosonic fields φ equal to the number of independently conserved currents. The bosons are coupled to one another via a coupling matrix that one generally calls K and which satisfies $\sum_{ij} K^{-1}_{ij} = \sigma_{\mathrm{xy}}$. The velocity v becomes an arbitrary velocity matrix V, with $\mathrm{tr}\, V > 0$.

$$S = \tfrac{1}{4\pi} \oint \mathrm{d}t\mathrm{d}x \; \left(K^{-1}_{ij} \partial_x \varphi_i \partial_t \varphi_j - V_{ij} \partial_x \varphi_i \partial_x \varphi_j \right). \tag{1.2.7}$$

The Jain hierarchy states (1.2.2) are accommodated in this framework by putting $K_{ij} = 2p + \delta_{ij}\mathrm{sgn}(m)$, where K has size $|m| \times |m|$. The theory (1.2.7) has a rich structure featuring Virasoro, Kac-Moody and $W_{1+\infty}$ algebras. Its particle spectrum, obtained from representation theory of these infinite algebras, contains the same composite fermions and 'fractional' quasiparticles as the bulk C-S theory. This shows that the *gapful* excitations in the bulk are related in a simple way to the *gapless* edge excitations. There is also a simple relation between chiral bosons and the Laughlin wave function. The ground state (1.2.1) is obtained by taking the expectation value with respect to (1.2.6) of a product of vertex operators which represent composite fermions. The conformal invariance of the Luttinger liquid model has been successfully used as a computational tool in the calculation of experimentally accessible quantities like temperature dependent current-voltage curves for tunnelling processes [14, 35, 78] and the shot noise signature of charge $1/3$ quasiparticles at $\nu = 1/3$ [69, 55, 20]. The Luttinger liquid also provides an interesting model system for exploring the thermodynamics of particles with unusual properties such as generalised exclusion statistics [29].

Problems

In spite of their successes, the phenomenological edge theories suffer from numerous problems and ambiguities. The derivation from (1.2.4) to (1.2.6) concentrates on a single physical quantity, the conductivity tensor, which is constant over the whole width of a plateau. There is no guarantee that the conductivity is the correct quantity to look at. Quantities like the density of states, which are known to change as a function of filling fraction, are not 'visible' in this approach. As a consequence, several important questions cannot be addressed. What, if any, is the effect of disorder on edge states? What determines the parameters V_{ij}? Do Coulomb interactions between edge states and localised bulk states play a role? And do interactions *within* the edge have any importance?

The lack of understanding becomes painfully clear when the $\sigma_{\mathrm{xy}} = 2/3$ plateau is considered. In the Jain hierarchy (1.2.2) this plateau corresponds to $m \approx -2$, $p = 1$, i.e. an effective filling fraction around 2 (in a reversed magnetic field) for composite fermions that consist of an electron plus two flux quanta. The 2×2 K-matrix has eigenvalues of opposite sign, which leads to the occurrence of counter-flowing edge modes. There has been a great deal of controversy over the nature of these counter-flowing modes. Problems were reported [36, 37, 30] concerning the dependence of the Hall conductance itself(!) and of the tunnelling density of states on the arbitrary parameters V_{ij}. Scattering operators between modes were introduced *by hand* in order to solve these problems. Even worse, our work has shown that an aspect which was thought to be well understood, the positive trace of V, is in fact wrong. In the case of counter-flowing modes, the velocity matrix has a negative trace and it is the *Coulomb interaction* that stabilises the Hamiltonian.

The most spectacular failing of the effective edge models is the incorrect prediction for the tunnelling exponent in experiments where electrons are tunnelled from a Fermi liquid into the edge of a fqHe sample [27]. The predicted curve as a function of filling fraction has plateaus just like the conductance, but the experiment shows a continuous curve as a function of ν. In the qHe, theory has lagged behind the experiments for a long time. The fact that one of the very few theoretical results obtained *in advance* of the experiment has turned out to be so far off the mark really drives home the limits of the phenomenological edge model approach. Within the K-matrix context, the obvious thing is to state that the neutral modes disappear. However, the formalism itself offers no justification for this statement, and the microscopic origins of the disappearance can only be guessed at. Our work shows that it is the Coulomb interaction between edge and bulk states that causes the continuous behaviour of the tunnelling exponent as a function of ν.

1.2.5 Disorder

Localisation and delocalisation

The iqHe is now understood as a disorder effect. A theory of noninteracting electrons in a strong magnetic field and a random scalar potential suffices to explain most transport phenomena.

In the absence of disorder, the electrons occupy Landau levels at energies $E_n = \hbar\omega_c(n+1/2)$, with ω_c the cyclotron or Larmor frequency eB/m_e, where m_e is the effective mass of the electron. The degeneracy of a Landau level is given by the number of magnetic flux quanta h/e that penetrate the sample. Every state in a Landau level is 'extended' and contributes the same amount of conductance, resulting in a linear relation between the Hall conductance and the filling fraction.

The effect of disorder is to cause localisation. The delta function peaks in the density of states (d.o.s.) that occur at the Landau energies E_n in a clean system get broadened into a 'Landau band' when disorder is added. The new states at $E \neq E_n$ are all localised. Localisation is a generic effect of all kinds of random potentials. According to the conventional wisdom before the discovery of the qHe, any amount of disorder, however small, localises all states in two dimensions [1]. (At least, in a renormalisation group sense, i.e. when very large length scales are considered). The big surprise of 1980 was that in a strong magnetic field extended states should exist at the center of each Landau band in order for the qHe to exist! The existence of these extended states leads to the following picture for the behaviour of σ_{ij} as a function of filling fraction. As long as the number of extended states below the Fermi level does not change, the Hall conductance is constant. This explains the Hall plateaus around integer filling. When the Fermi energy crosses a Landau energy, extra extended states start contributing to the conductance, resulting in a sharp increase of $\sigma_{\rm xy}$. The quantisation of $\sigma_{\rm xy}$ in units of e^2/h is elegantly explained by Laughlin's flux argument [57]. Finally, the $\sigma_{\rm xx}$ curve follows from the derivative $\mathrm{d}\sigma_{\rm xy}/\mathrm{d}\nu$. The important point is that the quantity of physical interest is not the density of states, as is usually the case in condensed matter physics, but the nature of these states. The Hall quantisation is caused not by a density of states gap, but by a *mobility gap*, an energy interval containing no extended states.

The mental picture of localised vs. extended states is so convenient that it is often used even for interacting electrons, where it strictly speaking does not apply.

Most of the insight into the mobility gaps and the existence of extended states has come from topological quantum field theory, more precisely the idea of an *instanton vacuum* [59, 45, 58]. Topological field theory provides the right language for the renormalisation group description of the relevant concepts such as localisation lengths, percolation and mobility gaps.

Nonlinear sigma model

The only formalism that can combine disorder and interactions is field theory. In field theories the random potential is always assumed to have a Gaussian distribution, since this is the only distribution that can be dealt with in path integrals. However, the effective actions are largely determined by symmetries alone, thus ensuring that the long distance physics is independent of the type of disorder.

Let us review in historical order the important developments in the application of field theoretical methods to the disorder problem. The first advance was Wegner's nonlinear sigma model [74] for noninteracting particles in zero magnetic field. It captures and extends the results of Abrahams *et al.* [1] and Langer and Neal [40], who used diagrammatic expansions for impurity scattering.

At first glance, a field theoretical approach only seems to obscure the matter. First, one has to give up the idea of a wave function, and replaces it by a field $\psi(\vec{x})$. This field has two components, advanced and retarded, with frequencies $\pm\eta$. Then the field is replicated [17] so that averaging over disorder becomes possible.[3] (The quantity to be averaged is $\ln Z$ with Z the grand canonical partition sum. Writing $\langle \ln Z\rangle = \lim_{N_r\to 0} N_r^{-1}[\langle Z_r^N\rangle - 1]$, it is clear that the problem can be reduced to averaging a power of Z. This is done by taking N_r exact copies of the system). The fact that the replica limit $N_r\to 0$ has to be taken does not exactly help our intuition. Finally, in the ensuing steps that lead to the sigma model, namely integration over disorder, introduction of a Hubbard-Stratonovich matrix field and integration over the massive modes of this new field, the original field ψ is mangled almost beyond recognition. What is left of it is a matrix field $Q^{\alpha\beta}_{pp'}$ with two replica indices ($\alpha, \beta = 1,\cdots,N_r$) and two frequency indices ($p, p' =$ advanced/retarded). It can be loosely interpreted as an average Green's function $\langle\psi^\alpha_p\psi^\beta_{p'}\rangle$. The nonlinear sigma model is a low temperature, low momentum effective theory in terms of the Q-field,

$$S_{\mathrm{nl}\sigma}[Q] = -\tfrac{1}{8}\sigma^0_{\mathrm{xx}}\int \mathrm{d}^2x\ \mathrm{tr}\,(\nabla Q)^2 + \eta\pi\rho\int \mathrm{d}^2x\ \mathrm{tr}\,\Lambda Q. \tag{1.2.8}$$

Here ρ is the (exact) density of states at the Fermi level and Q is a hermitian matrix defined as

$$\begin{aligned} &Q = T^{-1}\Lambda T \qquad && T\in U(2N_r) \qquad && \Lambda^{\alpha\beta}_{pp'} = \delta^{\alpha\beta}\delta_{pp'}\ \mathrm{sgn}(p) \\ &Q^\dagger = Q && Q^2 = \mathbb{1} && \mathrm{tr}\ Q = \sum_{p,\alpha} Q^{\alpha\alpha}_{pp} = 0, \end{aligned} \tag{1.2.9}$$

where the sign of the advanced(retarded) frequency counts as $+1(-1)$. The Q-matrix is invariant under local transformations $T\to UT$ with $U\in U(N_r)\times U(N_r)$.

The conceptual and computational difficulties of this approach are illustrated by the fact that Wegner did not identify the meaning of the parameter in the kinetic term $(\nabla Q)^2$. It was later realised that the expansion procedure that yields (1.2.8) leads to an expression in terms of Green's functions that is exactly the mean field expression for the dissipative conductance, σ^0_{xx}. In view of all the massaging that the theory has gone through, it is quite amazing that two important physical quantities, σ^0_{xx} and ρ, pop up as parameters.

The formalism requires a different view of the disorder problem than the single particle picture. The appropriate language is that of the renormalisation group (RG), spontaneously broken symmetries and asymptotic freedom.

The phenomenon of localisation at large distances arises as a restoration at large length scales of a continuous symmetry which is spontaneously broken at short distances, in analogy with the 2D Heisenberg ferromagnet. The frequency η is the analogue of the magnetic field and the expression $\pi\rho\mathrm{tr}\ \Lambda\langle Q\rangle$ corresponds to the magnetisation. As long as we are not considering a density of states gap, the 'magnetisation' is always nonzero, even for $\eta\to 0$. The presence of η forces Q in the 'direction' $Q=\Lambda$. The action (1.2.8) at $\eta=0$ is invariant under global $U(2N_r)$ transformations $T\to TU$, $U\in U(2N_r)$, but a nonzero $\langle Q\rangle$ breaks the symmetry. The variable T is in fact the Goldstone mode. From

[3] Supersymmetry can also be used. However, it is incompatible with interactions, so it will not be discussed.

the ψ-field theory it can be seen that the insulating phase has a higher symmetry, namely *local* $U(2N_r)$. The kinetic term breaks this symmetry. It can only be restored by $\sigma_{xx}\downarrow 0$ under the RG. So although there are strong similarities with ferromagnets, the analogy is not perfect. The symmetry that gets restored is a local one, and the restoration is not by a vanishing tr $\Lambda\langle Q\rangle$ but by σ_{xx} flowing to zero.

The concept of asymptotic freedom enters as follows. The 'nonlinearity' due to $Q^2=\mathbb{1}$ can be taken care of by parametrising Q in such a way that it contains square roots containing a field called, say, V. An expansion of the kinetic term in powers of V yields a kinetic term tr $(\nabla V^\dagger)(\nabla V)$ plus fourth order and higher order terms in V. The coupling constant of the fourth order self-coupling term is $t\propto 1/\sigma_{xx}$.[4] At short length scales t is very small. Under the RG it flows to infinity at large distances. This is similar to the scaling of the strong interaction in QCD.

Renormalisation in $2+\varepsilon$ dimensions has shown that in two dimensions and lower σ_{xx} renormalises to zero, i.e. the system becomes insulating. A metal-insulator transition can exist only in dimensions higher than two. The conventional wisdom of Anderson localisation clearly does not give any indication of the existence of a quantum Hall effect.

Topology

The qHe requires a scaling theory for both parameters σ_{xx} and σ_{xy}. Such a theory was developed by Pruisken *et al.* [45, 58, 61, 63]. The action (1.2.8) was extended to include a *topological term*

$$\begin{aligned} S_{\mathrm{nl}\sigma} &\to S_{\mathrm{nl}\sigma}+\sigma^0_{xy}S_{\mathrm{top}} \qquad (1.2.10)\\ S_{\mathrm{top}}[Q] &= \tfrac{1}{8}\int \mathrm{d}^2x\ \mathrm{tr}\ \varepsilon_{ij}Q\partial_iQ\partial_jQ=\tfrac{1}{2}\oint \mathrm{d}\vec{x}\cdot(\Lambda T\nabla T^{-1}). \end{aligned}$$

To date, the topological term has remained one of the most difficult chapters in the theory of Anderson localisation. It can be written as a boundary expression and it disappears in perturbative expansions, yet it completely changes the bulk theory. It causes a second order phase transition at the band centers and is responsible for the integer quantisation of σ_{xy}. The way in which it manages to do this is as follows. The topological term can be written as $S_{\mathrm{top}}[Q]=2\pi i\cdot q[Q]$. The $q[Q]$ is called the topological charge. If spherical boundary conditions are imposed, i.e. $T\in U(N_r)\times U(N_r)$ at the edge, so that $Q=\Lambda$ everywhere at the boundary, then the edge can be contracted to a point and the topological charge is integer quantised as a consequence of the homotopy result

$$\pi_2\left[\tfrac{U(2N_r)}{U(N_r)\times U(N_r)}\right]=\mathbb{Z}. \qquad (1.2.11)$$

A change of $q[Q]$ can be interpreted as a change of the number of electrons in the sample. At weak coupling spherical boundary conditions have to be imposed because of finite action requirements. At strong coupling there is no such requirement and the situation is more subtle, involving edge currents. Nevertheless, the boundary condition $Q=\Lambda$ was imposed, resulting in a separation of the Q-integration manifold into distinct topological sectors. The minimum energy configuration $Q(\vec{x})$ in a certain topological sector is called an instanton. It was shown that instantons are responsible for the renormalisation of σ_{xy}. By using the dilute instanton gas approximation, beta functions were obtained in the

[4] Phrases like 'weak coupling' and 'strong coupling' will always refer to the parameter t.

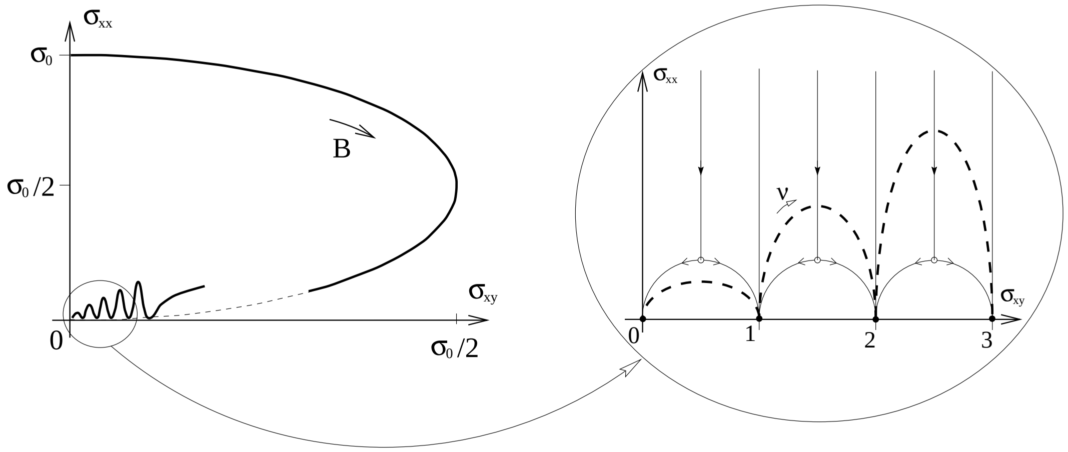

Figure 1.2: *Left: Sketch of the mean field conductances as a function of magnetic field. Right: The inset is the strong field limit or quantum Hall regime. The dashed curve gives the mean field conductance. Arrows indicate the RG flow.*

weak coupling regime. The results can be visualised in the scaling diagram shown in Fig. 1.2. Apart from the stable fixed points ($\bullet$) at ($\sigma_{\rm xx}=0, \sigma_{\rm xy}=$integer), which give rise to the Hall plateaus, there are also unstable fixed points ($\circ$) at half-integer values of $\sigma_{\rm xy}$. These are of great importance, since they describe a delocalisation of the wave functions or a 'massless' phase at the Landau band centers. A $\sigma^0_{\rm xy}$ slightly smaller than 1/2 will flow past the unstable fixed point at $\sigma_{\rm xy}=1/2$ and eventually end up in the stable fixed point at $\sigma_{\rm xy}=0$. A value of $\sigma^0_{\rm xy}$ slightly larger than 1/2 flows into $\sigma_{\rm xy}=1$. Since a sample is never truly infinite, there will be small regions around half-integer $\sigma^0_{\rm xy}$ whose RG flows keep 'hanging' in the vicinity of an unstable fixed point. This precisely corresponds to the transitions between Hall plateaus: $\sigma_{\rm xy}$ is not quantised and $\sigma_{\rm xx}$ does not vanish at the center of a Landau band.

The flow of the two conductances toward the stable fixed points is an extension of the description of localisation as a dynamic symmetry restoration. Since the topological charge changes by an integer under a local $U(2N_r)$ transformation, the theory can only be $U(2N_r)$ invariant if $\sigma_{\rm xy}$ is integer quantised while at the same time $\sigma_{\rm xx}=0$. This is an elegant way of understanding both conductance quantisations in the qHe as a manifestation of the same symmetry principle.

It was predicted [62] that the conductances and the localisation length ξ have the following scaling behaviour,

$$\sigma_{ij}(L,B) = g_{ij}([L/\xi]^{1/\nu}) \quad ; \quad \xi \propto |B-B^*|^{-\nu} \tag{1.2.12}$$

where L is the sample size, g_{ij} are regular functions, B^* is the critical magnetic field and ν is the critical exponent for the localisation length. The beta functions at strong coupling are not known, so there is no exact result for the critical exponent. The scaling relations (1.2.12) were confirmed by the measurements of H.P. Wei *et al.* [75, 76] on $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{As/InP}$ heterostructures. An important thing to notice is that scaling is only predicted for uncorrelated disorder. If there are potential fluctuations on the scale of the sample size itself, the scaling regime is never reached. The dopants in $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{As/InP}$ heterostructures are situated inside the 2D electron gas and give rise to short range disorder. In the very popular $\mathrm{GaAs/Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ samples on the other hand, the aluminium

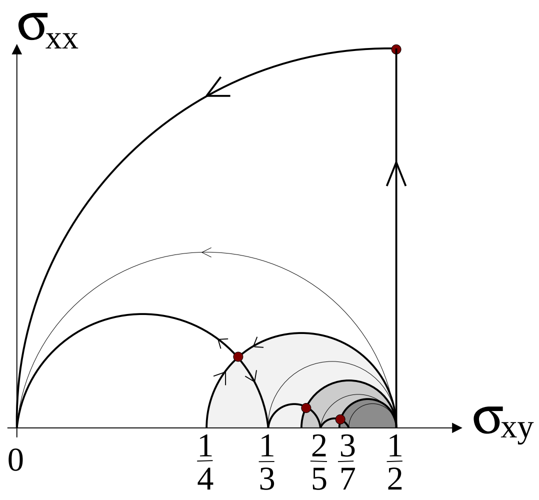

Figure 1.3: *Blowup of the interval* $0<\sigma_{\rm xy}<\frac{1}{2}$. *Shown here are the flow lines relevant for the Jain states* $p=1$, $m>0$. *The domains that flow to* $\frac{1}{3}$, $\frac{2}{5}$, $\frac{3}{7}$ ... *are indicated in different shades of grey.*

atoms are situated at a distance from the interface and cause long range potential fluctuations. Experiments on this kind of samples [71, 4] show behaviour different from (1.2.12).

Interactions

The sigma model (1.2.10) does not contain Coulomb interactions. In 1983 Finkelstein formulated an extension of (1.2.8) that does include them [22]. One of the consequences of interactions is that two frequency components (advanced/retarded) no longer suffice. One needs the full set of (fermionic) Matsubara frequencies $\frac{2\pi}{\beta}(n+\frac{1}{2})$, where β is the inverse temperature. For fully spin polarised electrons the effective action takes the form

$$S=-\tfrac{1}{8}\sigma^0_{\rm xx}\int \mathrm{d}^2x\ \mathrm{tr}\,(\nabla Q)^2+\tfrac{\pi}{\beta}z\int \mathrm{d}^2x\ \left[\sum_{\alpha klmn}Q^{\alpha\alpha}_{kl}Q^{\alpha\alpha}_{mn}\delta_{k-l,n-m}+4\sum_{\alpha n}nQ^{\alpha\alpha}_{nn}\right]. \quad (1.2.13)$$

The Q field in (1.2.13) is defined in a way similar to (1.2.9) but now in full frequency space, i.e. $T\in U(N_r\cdot N_{\rm f})$ where $N_{\rm f}$ stands for the (infinite) number of distinct Matsubara frequencies. The contribution in (1.2.13) containing the Kronecker delta is called the 'singlet' interaction term and represents the zero momentum limit of the interactions. The parameter z is called the singlet interaction amplitude. It represents the specific heat normalised with respect to the free particle case [10, 11]. The last term is a generalisation of the frequency term in (1.2.8) to the full frequency space.

With a one loop momentum shell calculation, Finkelstein showed that in the interacting case the conductance $\sigma_{\rm xx}$ still flows to zero. The z renormalises, but the relative parameter '4' between the interaction term and the frequency term does not change. This fact is related to the conservation of the total number of particles.

The next step forward in the sigma model theory was the inclusion of topological effects in the Finkelstein theory by Pruisken and Baranov [64]. They showed that the topological term, when added to (1.2.13), still leads to well defined instantons and that the scaling diagram (Fig. 1.2) still holds in the presence of interactions.

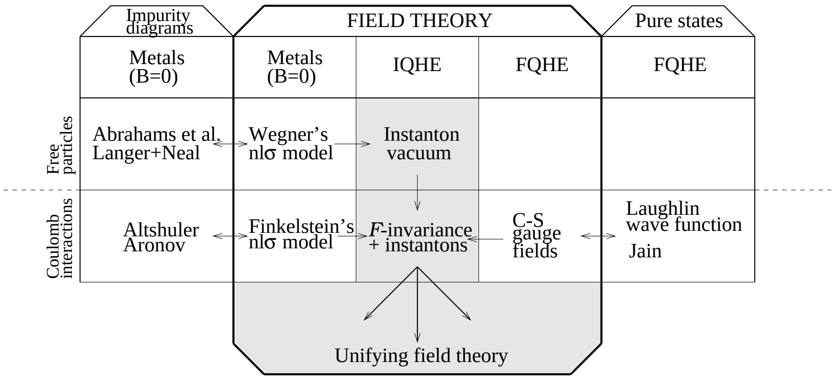

Figure 1.4: *Theories leading to a unified description of 2D metals at $B \approx 0$ and the quantum Hall effect(s).*

The fractional effect

The developments listed above have led to a deeper understanding of the iqHe. However, they do not describe the fractional effect. In Fig. 1.2 all the RG flow lines stop at integer quantised $\sigma_{\rm xy}$. It was conjectured [42] that the fqHe can be fitted into Fig. 1.2 under the 'arches' (see Fig. 1.3). There would have to be stable fixed points at ($\sigma_{\rm xx} = 0, \sigma_{\rm xy}$ quantised) and unstable ones between them at $\sigma_{\rm xx} \neq 0$ with arches as in the iqHe. This fqHe part of the scaling diagram cannot ordinarily be reached by RG, because the starting point for renormalisation, σ^0_{ij}, always lies above the 'fractional' region. A Chern-Simons mapping procedure is needed in order to change the σ^0_{ij} into effective conductances lying inside the fqH arches. However, the flux mapping procedure requires the coupling of Q to gauge fields, which is absent in (1.2.13).

It has been conjectured [51, 50, 38] that the C-S mapping of the conductances is a realisation of $\mathrm{SL}(2,\mathbb{Z})$ (or, more precisely, a subgroup of it). The statement is that the transformation rule for the complex conductance $\sigma := \sigma_{\rm xy} + i\sigma_{\rm xx}$ is an analytic continuation of the case with $\sigma_{\rm xx} = 0$,

$$\sigma \to \sigma/(2p\sigma + 1). \tag{1.2.14}$$

1.3 Organisation of this thesis

Fig. 1.4 shows the connections between our work and the theories presented in the previous sections. On the one hand we have Wegner's nonlinear σ model and the Finkelstein theory, field theories that encompass and extend the diagrammatic techniques of [1, 40, 2]. The inclusion of a topological term yields a good understanding of the iqHe. On the other hand we have Laughlin's wave function and Jain's composite fermion idea, which in field theory language are described by Chern-Simons actions. Connecting the left and right side of the figure there is the $\mathcal{F}$-invariance, enabling one to apply the C-S flux attachment to the theory of the iqHe.

The derivation of the unified description and the results that we have obtained with it are presented in this thesis as follows. Every chapter consists of a slightly modified paper. In chapter 2 I present the derivation of the $\mathcal{F}$-invariant action and discuss the intricacies of the $\mathcal{F}$-algebra. The results of a simple linear response calculation are shown.

The subject of chapter 3 is renormalisation of the Finkelstein theory in $2+2\varepsilon$ dimensions. The background field method is employed to obtain one-loop results. A two-loop calculation of the free energy is presented and scaling functions are derived. A new quantity, the quasiparticle density of states, is examined.

In chapter 4 edge states are discussed in the iqHe regime. It is shown that the topological term, or more precisely the fact that there are fluctuations around quantised topological charge, gives rise to edge excitations, and that Q-field theory on the edge is equivalent to a theory of chiral edge bosons. The difference between short range and long range disorder is investigated. A description of smooth disorder in terms of spatially separated 'edge' channels is proposed and used to explain the non-scaling of conductances and the $1/\nu$ tunnelling exponent.

The Chern-Simons mapping is discussed in chapter 5. The mapping is first performed on the bulk theory of chapter 2 and then on the edge theory of chapter 4. A classification is made of quasiparticles and tunnelling exponents in the fqH regime. An explicit relation is found between edge theories with short-ranged and long-ranged disorder. Finally, plateau transitions between fractional plateaus are investigated.

Chapter 2

$\mathcal{F}$-invariant bulk theory

2.1 Introduction

In this chapter we construct the unifying Q-field theory. We start in section 2.2 from the path integral description of fermions in 2+1 dimensions coupled to a gauge field A_μ. We introduce disorder averaging and replicas, and derive a gauge invariant effective action in terms of Q-matrices in infinite Matsubara frequency space. In section 2.2.3 we discuss the complicating effects of the truncation of frequency space. Due to the loss of the $U(1)$ algebra it seems impossible to write down a consistent theory that contains A_μ.

We next introduce the concept of $\mathcal{F}$-invariance, which is an invariance under a spatially constant $U(1)$ transformation of the Q-field, and devise a new truncation procedure that conserves this invariance. Our handling of frequency space seems to be the only consistent way of dealing with truncated $U(1)$ transformations. In section 2.2.4 we list some algebraic relations that are a consequence of our procedure. In section 2.3 we present the $\mathcal{F}$-invariant effective sigma model action in truncated frequency space, first without A_μ. The fact that $\mathcal{F}$-invariance resembles electromagnetic $U(1)$ gauge invariance allows us to next include A_μ and obtain a theory that is both $\mathcal{F}$- and gauge invariant. For the coupling it turns out to be necessary to have a detailed knowledge of edge excitations, since the topological term is a boundary term. We borrow a result from chapter 4 to write down the full coupling.

In section 2.4 we integrate out the Q-field at tree level and obtain an effective action for A_μ. We use it for a linear response calculation and study the transport properties of the system at short and long distances.

2.2 Q-field formalism; the fermionic path integral

We have tried everything else. There is no other explanation: Gaussian integrals are wrong!
– a Canadian theorist

2.2.1 Preliminaries

We are interested in the disorder average of the logarithm of the grand canonical partition function Z,

$$Z = \mathrm{tr}\ e^{\beta(\mu N - H)} \tag{2.2.1}$$

with β the inverse thermal energy, $\beta = (k_{\rm B}T)^{-1}$, μ the chemical potential, N the number of electrons and H the total energy of the system. We consider a system of two-dimensional electrons in a random potential $V(\vec{x})$ and a static magnetic field B pointing along the positive z-axis. We work in units where all lengths are expressed in terms of the magnetic length $\ell = \sqrt{2\hbar/eB}$ and where $\hbar = 1$, $e = 1$. In these units, the electron mass m has the dimension of an inverse energy, while the static magnetic field and the vector potential are dimensionless,

$$m = m_{\rm SI}\cdot \ell^2/\hbar^2 \qquad ; \qquad \vec{A} = \vec{A}_{\rm SI}\cdot e\ell/\hbar. \tag{2.2.2}$$

We write the vector potential as $\vec{A}^{\rm st} + \vec{A}$, where the static part satisfies $\nabla\times\vec{A}^{\rm st} = B\vec{e}_z$ and $\vec{A}$ represents the fluctuations. In the units chosen above, the static magnetic field is normalised to $B = 2$. The fluctuations in the scalar potential are denoted by A_0. In the path integral formulation, the partition function for our system is written in the following way

$$Z = \int \mathcal{D}\bar{\psi}\psi\;\mathcal{D}A_\mu\; e^{S[\bar{\psi},\psi,A_\mu]} \tag{2.2.3}$$

$$\begin{aligned} S[\bar{\psi},\psi,A_\mu] &= \int_0^\beta {\rm d}\tau \int {\rm d}^2x\; \bar{\psi}(\vec{x},\tau)[-\partial_0 + iA_0(\vec{x},\tau) + \mu - \mathcal{H}(\vec{x}) - V(\vec{x})]\psi(\vec{x},\tau) \\ &\quad -\tfrac{1}{2}\int_0^\beta {\rm d}\tau \int {\rm d}^2x{\rm d}^2x'\; \bar{\psi}(\vec{x},\tau)\psi(\vec{x},\tau)U_0(\vec{x},\vec{x}')\bar{\psi}(\vec{x}',\tau)\psi(\vec{x}',\tau). \end{aligned} \tag{2.2.4}$$

Here, the ψ and $\bar{\psi}$ are Grassmann variables defined on the imaginary time interval $\tau\in[0,\beta]$, with the fermionic antiperiodicity condition $\psi(\vec{x},\beta) = -\psi(\vec{x},0)$. The A_μ are ordinary integration variables with the bosonic boundary condition $A_\mu(\vec{x},\beta) = A_\mu(\vec{x},0)$. The $U_0(\vec{x},\vec{x}')$ is the Coulomb interaction and $\mathcal{H}$ is a differential operator acting to the left and to the right

$$\mathcal{H} := \tfrac{1}{2m}\overleftarrow{\pi}\cdot\vec{\pi}, \tag{2.2.5}$$

where π is the covariant derivative,

$$\vec{\pi} := \tfrac{1}{i}\vec{\nabla} - \vec{A}^{\rm st} - \vec{A} \quad ; \quad \overleftarrow{\pi} := -\tfrac{1}{i}\overleftarrow{\nabla} - \vec{A}^{\rm st} - \vec{A}. \tag{2.2.6}$$

The Coulomb term is quartic in the fields ψ. We get rid of this quartic form by performing a Hubbard-Stratonovich transformation, introducing an extra path integration over a bosonic field $\lambda(\vec{x},\tau)$, the 'plasmon field',

$$\begin{aligned} &\exp -\tfrac{1}{2}\int {\rm d}\tau{\rm d}^2x{\rm d}^2x'\;\bar{\psi}\psi(\vec{x})U_0(\vec{x},\vec{x}')\bar{\psi}\psi(\vec{x}') \propto \\ &\int \mathcal{D}\lambda\;\exp\left[-\tfrac{1}{2}\int {\rm d}\tau{\rm d}^2x{\rm d}^2x'\;\lambda(\vec{x})U_0^{-1}(\vec{x},\vec{x}')\lambda(\vec{x}') + i\int {\rm d}\tau{\rm d}^2x\;\lambda\bar{\psi}\psi\right]. \end{aligned} \tag{2.2.7}$$

Here U_0^{-1} stands for the matrix inverse of U_0. In order to find the disorder average $\overline{\ln Z}$ we use the well known *replica trick* [58, 63]. In the path integral formalism this amounts to the introduction of replicated fields $\bar{\psi}^\alpha, \psi^\alpha, \lambda^\alpha, A_\mu^\alpha$ with $\alpha = 1,\cdots,N_r$. The quantities μ, ρ, V and $A_\mu^{\rm st}$ are identical in all replicas. The replicated partition function is given by

$$\begin{aligned} Z &= \int \prod_{\gamma=1}^{N_r} \mathcal{D}\bar{\psi}^\gamma\psi^\gamma\mathcal{D}\lambda^\gamma\mathcal{D}A_\mu^\gamma\;\exp\sum_{\alpha=1}^{N_r}\int_0^\beta {\rm d}\tau\left[\int {\rm d}^2x\;\bar{\psi}^\alpha[-\partial_0 + iA_0^\alpha + \mu - \mathcal{H}^\alpha - V]\psi^\alpha\right. \\ &\quad \left. -\tfrac{1}{2}\int {\rm d}^2x{\rm d}^2x'\;\lambda^\alpha(\vec{x})U_0^{-1}(\vec{x},\vec{x}')\lambda^\alpha(\vec{x}') + i\int {\rm d}^2x\;\lambda^\alpha\bar{\psi}^\alpha\psi^\alpha\right]. \end{aligned} \tag{2.2.8}$$

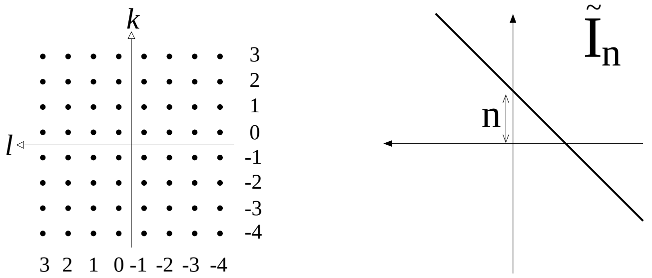

Figure 2.1: *Left: our way of drawing matrices* $[\cdots]_{kl}$ *in frequency space. Right: the structure of* $\tilde{I}_n$ $(n>0)$.

As a next step we perform a Fourier transform from imaginary time τ to Matsubara frequencies. Since fermionic fields are antiperiodic on the interval $[0,\beta]$, while bosonic fields are periodic, the allowed frequencies for $\psi,\bar{\psi}$ and A_0,λ are, respectively

$$\omega_n = \tfrac{2\pi}{\beta}(n+\tfrac{1}{2}) \quad \text{(fermionic)} \qquad ; \qquad \nu_n = \tfrac{2\pi}{\beta}n \quad \text{(bosonic)}. \tag{2.2.9}$$

with n integer. We define the Fourier transformed fields by

$$\begin{aligned} \psi^\alpha(\tau) &= \sum_{n=-\infty}^{\infty} \psi_n^\alpha e^{-i\omega_n\tau} \qquad ; \qquad \bar{\psi}^\alpha(\tau) = \sum_{n=-\infty}^{\infty} \bar{\psi}_n^\alpha e^{+i\omega_n\tau} \\ \lambda^\alpha(\tau) &= \sum_{n=-\infty}^{\infty} \lambda_n^\alpha e^{-i\nu_n\tau} \qquad ; \qquad A_\mu^\alpha(\tau) = \sum_{n=-\infty}^{\infty} (A_\mu)_n^\alpha e^{-i\nu_n\tau} \end{aligned} \tag{2.2.10}$$

which results in the following form of the action:

$$S = \beta\int \mathrm{d}^2x\ \psi^\dagger(i\omega + i\hat{A}_0 + \mu + i\hat{\lambda} - \hat{\mathcal{H}} - \hat{V})\psi - \tfrac{\beta}{2}\int \mathrm{d}^2x\mathrm{d}^2x'\ \lambda^\dagger(\vec{x})U_0^{-1}(\vec{x},\vec{x}')\lambda(\vec{x}'). \tag{2.2.11}$$

Here we have used matrix notation for combined replica and frequency indices,

$$\psi^\dagger(\cdots)\psi := \sum_{nm,\alpha\beta} \bar{\psi}_n^\alpha(\cdots)_{nm}^{\alpha\beta}\psi_m^\beta. \tag{2.2.12}$$

The ω is a unity matrix in replica space, while in Matsubara space it is a diagonal matrix containing the frequencies ω_n. We will often also use a diagonal matrix η containing the numbers n,

$$\omega = \tfrac{2\pi}{\beta}(\eta + \tfrac{1}{2}\mathbb{1}) \qquad (\omega)_{nm}^{\alpha\beta} = \omega_n\delta^{\alpha\beta}\delta_{nm} \qquad (\eta)_{nm}^{\alpha\beta} = n\delta^{\alpha\beta}\delta_{nm}. \tag{2.2.13}$$

The 'hatted' quantities are defined according to

$$\hat{z} = \sum_{n,\alpha} z_n^\alpha \tilde{I}_n^\alpha \qquad \text{with} \quad (\tilde{I}_n^\alpha)_{kl}^{\beta\gamma} := \delta^{\alpha\beta}\delta^{\alpha\gamma}\delta_{k-l,n}. \tag{2.2.14}$$

The matrix $\tilde{I}_n^\alpha$ is the unit matrix in the α'th replica space, while in Matsubara space it is zero everywhere except on the n'th diagonal, where it is 1. The $\tilde{I}_n^\alpha$-matrices are very important, because they will turn out to be the generators of the electromagnetic $U(1)$ transformations. But before we elaborate on this, let us first take the disorder average of the replicated partition sum, in analogy with what has been done in the free particle formalism. This is done using a Gaussian distribution for the random potential $V(\vec{x})$,

$$\overline{Z} = \int \mathcal{D}V P[V] Z \quad ; \quad P[V] \propto \exp\left(-\tfrac{1}{2g}\int \mathrm{d}^2x\, V^2\right). \tag{2.2.15}$$

This integration leads to a quartic term in the action of the form $(\psi^\dagger\psi)^2$, which can be decoupled by means of a Hubbard-Stratonovich transformation, introducing hermitian matrix field variables $\widetilde{Q}^{\alpha\beta}_{nm}(\vec{x})$. The partition function now becomes

$$\begin{aligned} Z &= \int \mathcal{D}\bar{\psi}\psi\, \mathcal{D}\widetilde{Q}\, \mathcal{D}\lambda\, \mathcal{D}A_\mu\, e^{S[\bar{\psi},\psi,\widetilde{Q},\lambda,A_\mu]} & (2.2.16)\\ S[\bar{\psi},\psi,\widetilde{Q},\lambda,A_\mu] &= -\tfrac{1}{2g}\mathrm{Tr}\,\widetilde{Q}^2 + \beta\int \mathrm{d}^2x\, \psi^\dagger[i\omega + i\hat{A}_0 + \mu - \hat{\mathcal{H}} + i\hat{\lambda} + i\widetilde{Q}]\psi \\ &\quad -\tfrac{\beta}{2}\int \mathrm{d}^2x\mathrm{d}^2x'\, \lambda^\dagger(\vec{x})\lambda(\vec{x}')U_0^{-1}(\vec{x},\vec{x}') & (2.2.17)\end{aligned}$$

where the notation Tr stands for a trace over combined replica and Matsubara indices as well as spatial integration $\int \mathrm{d}^2x$. Notice that the only difference with the previously studied free particle case [58] is that we work with a Matsubara frequency label, rather than with advanced and retarded components alone. In the partition function (2.2.16), the integration over fermion fields can be performed, yielding an effective action for the variables $\widetilde{Q}$, A_μ and λ,

$$\begin{aligned} S[\widetilde{Q},\lambda,A_\mu] &= -\tfrac{1}{2g}\mathrm{Tr}\,\widetilde{Q}^2 + \mathrm{Tr}\,\ln[i\omega + i\hat{A}_0 + \mu - \hat{\mathcal{H}} + i\hat{\lambda} + i\widetilde{Q}] \\ &\quad -\tfrac{\beta}{2}\int \mathrm{d}^2x\mathrm{d}^2x'\, \lambda^\dagger(\vec{x})\lambda(\vec{x}')U_0^{-1}(\vec{x}-\vec{x}'). & (2.2.18)\end{aligned}$$

2.2.2 Gauge invariance

A generic local $U(1)$ gauge transformation on the fermion fields and the electromagnetic potentials has the form

$$\begin{aligned} &\psi^\alpha(\vec{x},\tau) \to e^{i\chi^\alpha(\vec{x},\tau)}\psi^\alpha(\vec{x},\tau) \quad ; \quad \bar{\psi}^\alpha(\vec{x},\tau) \to e^{-i\chi^\alpha(\vec{x},\tau)}\bar{\psi}^\alpha(\vec{x},\tau) & (2.2.19)\\ &A^\alpha_\mu(\vec{x},\tau) \to A^\alpha_\mu(\vec{x},\tau) + \partial_\mu\chi^\alpha(\vec{x},\tau) \end{aligned}$$

with χ^α real-valued functions periodic in τ. In frequency notation this gauge transformation is written as a unitary matrix acting on the vector ψ

$$\begin{aligned} \psi \to e^{i\hat{\chi}}\psi \quad &; \quad \psi^\dagger \to \psi^\dagger e^{-i\hat{\chi}} & (2.2.20)\\ \hat{A}_i \to \hat{A}_i + \partial_i\hat{\chi} \quad &; \quad (A_0)^\alpha_n \to (A_0)^\alpha_n - i\nu_n\chi^\alpha_n. & (2.2.21)\end{aligned}$$

From (2.2.20) it is clear that the $\tilde{I}$-matrices are the generators of gauge transformations. From their multiplication

$$\tilde{I}^\alpha_n \tilde{I}^\beta_m = \delta^{\alpha\beta}\tilde{I}^\alpha_{n+m} \tag{2.2.22}$$

it is readily seen that they span an Abelian algebra, and that a gauge transformation indeed acts in every replica channel separately, as seen in (2.2.19). The $\widetilde{Q}$ transforms according to

$$\widetilde{Q} \to e^{i\hat{\chi}}\widetilde{Q}e^{-i\hat{\chi}}. \tag{2.2.23}$$

The gauge invariance of the action (2.2.17) is easily checked: First of all, the plasmon field λ and the combinations $\psi^\dagger \widetilde{Q}\psi$ and $\psi^\dagger\psi$ are invariant. Secondly, the fact that the $\tilde{I}$ commute leads to $e^{-i\hat{\chi}}\nabla e^{i\hat{\chi}} = i\nabla\hat{\chi}$, from which it follows that the term $\psi^\dagger\hat{\mathcal{H}}\psi$ is also invariant. Finally, using the following commutation relation,

$$[\tilde{I}_n^\alpha, \omega] = -\nu_n \tilde{I}_n^\alpha, \tag{2.2.24}$$

in combination with the transformation rule (2.2.21) for A_0, we find that the term $\psi^\dagger(i\omega + i\hat{A}_\tau)\psi$ is also invariant. The gauge invariance of (2.2.18) is evident as well. We only have to rewrite the gauge transformed second term Tr $\ln[\cdots + ie^{i\hat{\chi}}\widetilde{Q}e^{-i\hat{\chi}}]$ into Tr $\ln[e^{-i\hat{\chi}}(\cdots)e^{i\hat{\chi}} + i\widetilde{Q}]$ and repeat the arguments given above.

2.2.3 $\mathcal{F}$-invariance; truncation of frequency space

Ve propoze an unconventional *solutsion to zis probljem.*
– an evil scientist

Obtaining an effective theory for the low energy $\widetilde{Q}$ excitations is not as straightforward as in the free particle case. The expansion of the Tr ln in (2.2.18) for low temperature and small momenta involves divergent frequency sums due to the presence of the $\tilde{I}$-matrices. In order to handle these divergencies, it is necessary to introduce a frequency cutoff. This cutoff serves only as a regulator and we should be able to send it to infinity at the end of all calculations. We choose the cutoff in such a way that all frequencies ω_n are restricted to the range

$$n \in \{-N'_{\max}, \cdots, N'_{\max} - 1\}. \tag{2.2.25}$$

Unfortunately, the truncation destroys the simple $U(1)$ algebra of the $\tilde{I}$-matrices. If we denote the truncated matrices by I_n^α, we have

$$(\mathrm{I}_n^\alpha \mathrm{I}_m^\beta)_{kl}^{\mu\nu} = (\tilde{I}_n^\alpha \tilde{I}_m^\beta)_{kl}^{\mu\nu} g_{l+m} \quad ; \quad [\mathrm{I}_n^\alpha, \mathrm{I}_m^\beta]_{kl}^{\mu\nu} = \delta^{\alpha\beta\mu\nu}\delta_{k-l,m+n}(g_{l+m} - g_{l+n}) \tag{2.2.26}$$

where $\delta^{\alpha\beta\mu\nu}$ means that all replica indices have to be the same, and g_i is a step function equal to one if $i \in \{-N'_{\max}, \ldots, N'_{\max} - 1\}$ and zero otherwise.

The destruction of the $U(1)$ algebra is a serious complication. Our final aim is an effective action $S_{\text{eff}}[T, A]$ *in truncated frequency space* that is somehow capable of producing gauge invariant results. However, the lack of a simple $U(1)$ algebra in truncated form means that there is no criterion by which to check whether intermediate results are properly invariant. Even worse, the obvious procedure of first finding the theory for $A_\mu = 0$ (a very hard task in itself) and then including A_μ by some simple rule like minimal coupling fails. Assuming that the theory without gauge fields has been found, a transformation like (2.2.23) still leads to a mess that cannot be cleaned up by inserting counter terms involving A_μ.

A new symmetry comes to the rescue. Apart from the electromagnetic gauge invariance, the action (2.2.18) possesses a second $U(1)$ invariance. In the theory *without* A_μ, the action does not change under a spatially global $U(1)$ transformation,

$$\widetilde{Q} \to e^{i\hat{\chi}}\widetilde{Q}e^{-i\hat{\chi}} \quad , \quad \nabla\chi = 0. \tag{2.2.27}$$

This transformation amounts to a spatially independent shift of the plasmon field inside the Tr ln according to

$$\lambda(\vec{x}) \to \lambda'(\vec{x}) = \lambda(\vec{x}) - \partial_0\chi. \tag{2.2.28}$$

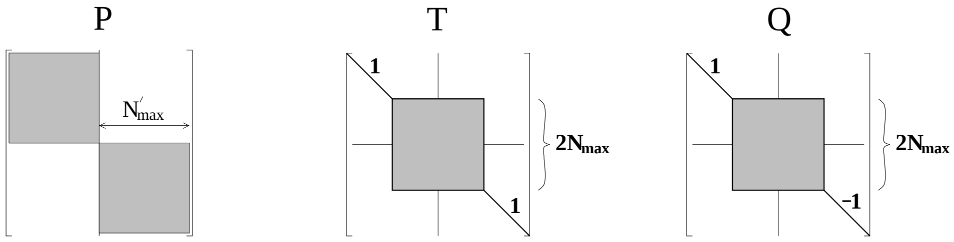

Figure 2.2: *The form of the matrices P, T and Q in truncated frequency space.*

If we now choose λ' as the integration variable instead of λ, then the shift occurs in the last term of (2.2.18). The effect of this shift is zero provided that the interaction U_0 is long-ranged, i.e. $U_0^{-1}(\vec{q}) \to 0$ as $|\vec{q}| \to 0$. In two spatial dimensions, the Coulomb interaction is infinitely ranged. The Fourier transform is given by

$$U_0(\vec{q}) \propto \int \mathrm{d}^2x \, |\vec{x}|^{-1} e^{-i\vec{q}\cdot\vec{x}} \propto |\vec{q}|^{-1}. \tag{2.2.29}$$

We have given the invariance under the transformation (2.2.27) the name of '$\mathcal{F}$-invariance' (after Finkelstein). The associated Noether charge is the fermion number or total electric charge.

We cannot overemphasise the importance of $\mathcal{F}$-invariance. First, it provides us with a very strong restriction on the possible form of the effective action in the absence of gauge fields, namely that it should be invariant under the truncated equivalent of (2.2.27),

$$S_{\text{eff}}[TW^{-1}] = S_{\text{eff}}[T] \qquad W = \exp i \sum_{n\alpha} \chi_n^\alpha \mathrm{I}_n^\alpha = e^{i\hat{\chi}} \qquad \nabla\chi = 0. \tag{2.2.30}$$

(From this point on, the 'hat' notation implies that truncated I-matrices are used). We denote the truncated form of a general $U(1)$ transformation by the name of 'W-rotation'. In this terminology the truncated version of $\mathcal{F}$-invariance is formulated as invariance under spatially independent W-rotations.

Second, after finding a way to handle the truncated form of $\mathcal{F}$-invariance, we are in a position to generalise to spatially dependent W-rotations and so solve the problem of truncated electromagnetic gauge invariance. All of this is of course purely hypothetical if we do not succeed in taming the mess which W-rotations create. We will now present a procedure that indeed satisfies the $\mathcal{F}$-invariance demanded in (2.2.30). It involves a completely new truncation prescription in frequency space. We proceed as in the free particle analysis and split the $\widetilde{Q}$ matrix variable into 'transverse' and 'longitudinal' components,

$$\widetilde{Q} = T^{-1}PT \qquad P = P^\dagger \qquad T \in U(2N'). \tag{2.2.31}$$

Here, P has only block-diagonal components in frequency space (i.e. $P_{nm}^{\alpha\beta}$ is nonzero only for $\omega_n\omega_m > 0$) and T is a unitary rotation. $2N'$ is the size of the Matsubara space times the number of replica channels, and represents the size of the $\widetilde{Q}$-matrix.

This change of variables (2.2.31) is motivated by the saddlepoint structure of the theory (2.2.18) in the absence of the fields A_μ, λ and at zero temperature (i.e. $\omega_n \to 0$). This saddlepoint can be written as

$$\widetilde{Q}_{\text{sp}} \propto T^{-1}\Lambda T \qquad ; \qquad \Lambda_{nm}^{\alpha\beta} = \delta^{\alpha\beta}\delta_{nm}\text{sgn}(\omega_n), \tag{2.2.32}$$

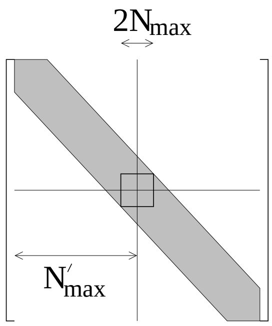

Figure 2.3: *The frequency interval* $n\in\{-2N_{\max}+1,\cdots,2N_{\max}-1\}$ *for which* λ_n^α *and* $\mathrm{tr}\ \mathrm{I}_n^\alpha Q$ *are nonzero.*

indicating that the longitudinal fluctuations P are the 'massive' components of the theory whereas the T-matrix fields are the lowest energy excitations (Goldstone modes) in the problem. In analogy with the free particle case we define a matrix field Q

$$Q := T^{-1}\Lambda T \qquad Q^2 = \mathbb{1} \qquad \mathrm{tr}\ Q = 0. \tag{2.2.33}$$

This is the only local variable on which the effective action $S_{\mathrm{eff}}[T]$ can depend that is consistent with the symmetry present in the zero-temperature saddlepoint (2.2.32).

The way in which (2.2.31) is going to be handled is illustrated in Fig. 2.2: we impose on the T-rotations a 'small' cutoff $N_{\max}$ in Matsubara frequency space, such that $1\ll N_{\max}\ll N'_{\max}$. The frequency indices on T range from $-N_{\max}$ to $N_{\max}-1$. As a result, we can write

$$Q = \Lambda + \delta Q \tag{2.2.34}$$

where δQ only contains small frequencies. At the same time we restrict the frequencies for which $\lambda_n \neq 0$ to the interval $\{-2N_{\max}+1,\ldots,2N_{\max}-1\}$, so that the matrix $\hat\lambda$ is nonzero only in a band that touches the 'small' δQ. In Fig. 2.3 this band is shown. (In the presence of gauge fields, $\hat A_\mu$ is restricted to the same band).

It is important to keep in mind that working with a finite $N_{\max}$ is just a calculational device which will enable us to derive an effective action for the lowest energy excitations T by formally integrating out the massive components P. (The latter can be done explicitly by employing saddlepoint methods). Once an effective action for the T-fields has been obtained, we have to find some procedure by which the cutoff $N_{\max}$ can be sent to infinity.

A formal justification of the 'smallness' of the T-rotation is postponed until after section 2.3.3 where we introduce vector and scalar potentials in the effective action. It will turn out that a 'large' T-rotation can always be written as a product of a 'large' block diagonal matrix times a 'small' rotation; the block diagonal part can be absorbed into P. A certain intuition for the truncation can be gained by noting, as Finkelstein did, that the frequency matrix η plays the role of the 'magnetic field' if the analogy with the Heisenberg ferromagnet is used. At large frequencies Q is forced to 'line itself up' and go to Λ.

2.2.4 $\mathcal{F}$-algebra

We use the word '$\mathcal{F}$-algebra' as a collective name for all the complicated W-rotations, commutation relations etc. in truncated frequency space that come in the place of the

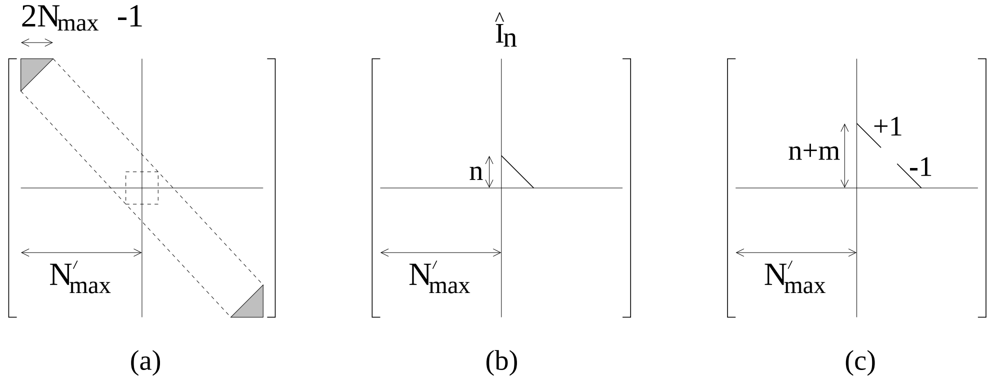

Figure 2.4: (a) *Nonzero regions of* $[\hat{x}_1, \hat{x}_2]$ *with* x_1 *and* x_2 *arbitrary functions.* (b) *The matrix* $\hat{\mathrm{I}}_n$ *for* $n>0$. (c) *The matrix* $[\mathrm{I}_m, \hat{\mathrm{I}}_n]$ *for positive* m *and* n. *The size of the nonzero segments is* $\min\{n, m\}$.

simple $U(1)$ results which one has in infinite frequency space. Here we give a list of the most important ingredients, which will be needed later on in order to determine whether certain quantities are $\mathcal{F}$-invariant or not. Commutation of I-matrices among themselves gives

$$\begin{aligned} [\mathrm{I}_n^\alpha, \mathrm{I}_m^\beta]_{kl}^{\mu\nu} &= \delta^{\alpha\beta\mu\nu}\delta_{k-l,m+n}(g_{l+m} - g_{l+n}) \\ \left[\mathrm{I}_p^\gamma, [\mathrm{I}_n^\alpha, \mathrm{I}_m^\beta]\right]_{kl}^{\mu\nu} &= \delta^{\alpha\beta\gamma\mu\nu}\delta_{k-l,m+n+p}\left[g_{l+m+n}(g_{l+m} - g_{l+n}) - g_{l+p}(g_{l+p+m} - g_{l+p+n})\right] \end{aligned} \tag{2.2.35}$$

An interesting thing to note is that $[\mathrm{I}_n^\alpha, \mathrm{I}_m^\beta]$ is nonzero only in small triangular regions in the upper left and lower right corner (see Fig. 2.4a). The reason is the following. The outcome of $[\mathrm{I}_n, \mathrm{I}_m]$ is only nonzero if n and m have opposite signs. The result lies on the $n+m$'th diagonal, with $n+m \in \{-2N_{\max}+2, \ldots, 2N_{\max}-2\}$. The number of nonzero components at one end of this diagonal is given by $\min(|n|, |m|)$. The maximum number of nonzero components (at one end) for a given diagonal d is $2N_{\max}-1-|d|$.

Higher order commutators are also restricted to the upper left and lower right corner. The size of the corner region grows with an amount $2N_{\max}$ at every extra commutation. Commutation of I-matrices with η and Λ yields

$$[\mathrm{I}_n^\alpha, \eta] = -n\mathrm{I}_n^\alpha \tag{2.2.36}$$

$$[\mathrm{I}_n^\alpha, \Lambda]_{kl}^{\beta\gamma} = \delta^{\alpha\beta\gamma}\delta_{k-l,n}(\sigma_l - \sigma_k) =: -2\,\mathrm{sgn}(n)(\hat{\mathrm{I}}_n^\alpha)_{kl}^{\beta\gamma} \tag{2.2.37}$$

$$\left[\mathrm{I}_m^\beta, [\mathrm{I}_n^\alpha, \Lambda]\right]_{kl}^{\mu\nu} = \delta^{\alpha\beta\mu\nu}\delta_{k-l,m+n}\left[g_{l+n}(\sigma_l - \sigma_{l+n}) - g_{l+m}(\sigma_{l+m} - \sigma_{l+m+n})\right] \tag{2.2.38}$$

Here σ_n denotes the sign of ω_n and we have defined the matrix $\hat{\mathrm{I}}_n^\alpha$ as the part of I_n^α that lives in the $[+-]$ or $[-+]$ part of frequency space (see Fig. 2.4b). Using (2.2.35–2.2.38) we are now able to determine the effect of a W-rotation on the quantities tr $\mathrm{I}_n^\alpha Q$ and tr ηQ, which will turn out to be important later on. We find

$$\mathrm{tr}\; \mathrm{I}_n^\alpha e^{i\hat{\chi}} Q e^{-i\hat{\chi}} = \mathrm{tr}\; \mathrm{I}_n^\alpha Q + \tfrac{\beta}{\pi}(\partial_0\chi)_{-n}^\alpha \tag{2.2.39}$$

$$\mathrm{tr}\; \eta e^{i\hat{\chi}} Q e^{-i\hat{\chi}} = \mathrm{tr}\; \eta Q - \tfrac{\beta}{2\pi}\mathrm{tr}\; Q\widehat{\partial_0\chi} - (\tfrac{\beta}{2\pi})^2(\partial_0\chi)^\dagger(\partial_0\chi). \tag{2.2.40}$$

The remarkable thing about this result is that it is exact. Higher order terms in a commutation power series expansion are all zero. This can be seen as follows. We write tr $\mathrm{I}_n^\alpha e^{i\hat{\chi}} Q e^{-i\hat{\chi}}$ in the form tr $\mathrm{I}_n^\alpha e^{i\hat{\chi}} \Lambda e^{-i\hat{\chi}} +$ tr $e^{-i\hat{\chi}} \mathrm{I}_n^\alpha e^{i\hat{\chi}} \delta Q$. It is the first commutator with Λ in the first term that gives rise to the $\partial_0\chi$ in (2.2.39). All higher order commutators with

Λ have a form where there are as many matrix elements $+1$ as -1 on the diagonal (see Fig. 2.4c). Traced with an I-matrix, this gives zero. In the second term all commutators with I_n^α live in the corner regions. Therefore, tracing with the 'small' δQ produces zero. One would need a number of commutations of the order of $N'_{\max}/N_{\max} \gg 1$ before the corner regions start to meet δQ.

Eq. (2.2.40) can be understood by writing the l.h.s. as tr $e^{-i\hat{\chi}}\eta e^{i\hat{\chi}}Q$. The first and second commutator produce the first and second terms respectively in the r.h.s. Third and higher orders give rise to a broken up diagonal structure in each corner, that contains an equal number of $+1$ and -1 entries. The trace with both Λ and δQ vanishes.

What is more, the arbitrary cutoff $N'_{\max}$ does not enter into (2.2.39, 2.2.40), which means that it can be safely taken to infinity. Eqs. (2.2.39) and (2.2.40) will play a key role in the $\mathcal{F}$-invariance of the effective action. Although the rules in this section seem complicated and counterintuitive, strategic application of W-rotations can lead to huge computational advantages.

2.3 Effective action

2.3.1 The case $A_\mu = 0$

The effective action in the absence of gauge fields is defined by

$$\begin{aligned} e^{S_{\rm eff}[T]} \;\propto\; & \int \mathcal{D}\lambda\mathcal{D}P\, I[P]\, \exp\Big(-\tfrac{1}{2g}\mathrm{Tr}\, P^2 - \tfrac{\beta}{2}\int \mathrm{d}^2x\mathrm{d}^2x'\, \lambda^\dagger(\vec{x})U_0^{-1}(\vec{x}-\vec{x}')\lambda(\vec{x}') \\ & +\mathrm{Tr}\, \ln[i\omega + \mu + i\hat{\lambda} - \mathcal{H}_0 + iT^{-1}PT]\Big) \end{aligned} \tag{2.3.1}$$

where $I[P]$ is the Jacobian of the change of variables $\widetilde{Q}\to T^{-1}PT$ and $\mathcal{H}_0$ is the differential operator (2.2.5) without the fluctuations $\vec{A}$. From general symmetry considerations one can impose two important conditions on the actual form of $S_{\rm eff}[T]$.

1. The only local variable on which $S_{\rm eff}[T]$ can depend and which is consistent with the symmetries of the problem is precisely Q (2.2.33).

2. The effective action must be $\mathcal{F}$-invariant:

$$S_{\rm eff}[e^{i\hat{\chi}}Qe^{-i\hat{\chi}}] = S_{\rm eff}[Q] \quad ; \quad \nabla\chi = 0 \tag{2.3.2}$$

We refer to appendix 2B for details of the derivation and here just present the result,

$$S_{\rm eff}[Q] = S_\sigma[Q] + S_{\rm F}[Q] + S_{\rm Y}[Q] \tag{2.3.3}$$

$$S_\sigma[Q] = -\tfrac{1}{8}\sigma^0_{\rm xx}\mathrm{Tr}\,(\nabla Q)^2 + \tfrac{1}{8}\sigma^0_{\rm xy}\mathrm{Tr}\,\varepsilon_{ij}Q\partial_i Q\partial_j Q \tag{2.3.4}$$

$$S_{\rm F}[Q] = z\tfrac{\pi}{2\beta}\sum_{\alpha n}{}' \mathrm{Tr}\,[\mathrm{I}_n^\alpha, Q][\mathrm{I}_{-n}^\alpha, Q] \tag{2.3.5}$$

$$S_{\rm Y}[Q] = -\tfrac{1}{2\beta}\sum_{\alpha n}\int \mathrm{d}^2x\mathrm{d}^2x'\,[\mathrm{tr}\,\mathrm{I}_n^\alpha Q(\vec{x})]Y(\vec{x}-\vec{x}')[\mathrm{tr}\,\mathrm{I}_{-n}^\alpha Q(\vec{x}')] \tag{2.3.6}$$

where z is the singlet interaction amplitude and Y is given by (in momentum space)

$$Y(\vec{p}) = \frac{\pi\rho/2}{1+2\pi\rho U_0(\vec{p})} \tag{2.3.7}$$

with ρ the density of states $\mathrm{d}n/\mathrm{d}\mu$. The prime on the summation in (2.3.5) indicates that the frequency n is restricted to the values $n \in \{-N'_{\max}, \cdots, N'_{\max}-1\}$. For a general understanding of the result (2.3.3) let us discuss the various pieces separately. First, by putting the temperature β^{-1} equal to zero we obtain the same result S_σ as in the free electron theory. The 'bare' parameters σ^0_{xx} and σ^0_{xy} are generally modified by the electron-electron interactions. The modifications are of a Fermi-liquid type and in the limit of strong magnetic fields the results depend on the ratio of disorder energy Γ_0 (the width of the Landau band) and the typical Coulomb energy E_0 ($= U_0(\ell)$, where ℓ is the magnetic length). Notice that

$$S_\sigma[Q] = S_\sigma[\delta Q] \qquad ; \qquad \delta Q := Q - \Lambda \tag{2.3.8}$$

and the same statement holds for S_{Y}. The most important part of the action is the Finkelstein action S_{F} (2.3.5). The diagonal parts of Q in this expression give rise to nontrivial contributions and we have the following identity:

$$S_{\mathrm{F}}[Q] = z\tfrac{\pi}{\beta}\int \mathrm{d}^2x \left[\sum_{\alpha n} \mathrm{tr}\ \mathrm{I}^\alpha_n Q\ \mathrm{tr}\ \mathrm{I}^\alpha_{-n} Q + 4\mathrm{tr}\ \eta Q - 6\mathrm{tr}\ \eta\Lambda\right]. \tag{2.3.9}$$

Discarding the (infinite) constants $\propto$ tr $\eta\Lambda$ for the moment, we see that replacing Q by δQ yields identical results. Hence, the difference between the quadratic terms in (2.3.5) and (2.3.9) is the so-called frequency term tr ηQ. Eq. (2.3.9) can also be written more explicitly as

$$S_{\mathrm{F}}[Q] = z\tfrac{\pi}{\beta}\int \mathrm{d}^2x \left[\sum_{\alpha}\sum_{klmn} \delta Q^{\alpha\alpha}_{kl}\delta Q^{\alpha\alpha}_{mn}\delta_{k-l,n-m} + 4\mathrm{tr}\ \eta\delta Q\right] + \mathrm{const.} \tag{2.3.10}$$

which is the result originally obtained by Finkelstein [22]. In Finkelstein's formulation of the problem, the very special relative coefficient '4' between the first (singlet-interaction) and second (frequency) terms in (2.3.10) is related to the macroscopic conservation laws in a very indirect way. The advantage of the present formalism is (among other things) the simple algebraic interpretation of the result which can be directly obtained from the symmetries of the problem.

Next we elaborate on the significance of the 'Coulomb term' S_{Y} (2.3.6). It is usually ignored in renormalisation group analyses, since it really stands for a higher dimensional operator. (Notice that $Y(\vec{p}) \propto |\vec{p}|$ in the small momentum limit). However, it cannot be discarded. In the large momentum limit z can be replaced by $\frac{1}{2}\pi\rho$ [6]. Consequently the sum of S_{F} and S_{Y} does not contain the term quadratic in tr IQ in this limit and (2.3.3) reduces to the free particle action. This means that the full theory of (2.3.3–2.3.6) is appropriately interpreted in terms of a cross-over phenomenon between free electron behaviour at short distances (or high temperature β^{-1}) and an interaction dominated behaviour which appears at larger distances (or lower temperatures β^{-1}) only.

Secondly, the complete form of (2.3.6) and (2.3.7) unravels important information on the nature of quantum transport even for ordinary metals. This will be the main subject of section 2.4, where we compute at a tree level the complete momentum and frequency dependent response to electromagnetic perturbations.

Finally, we mention that although (2.3.6) is naively irrelevant from the weak coupling renormalisation point of view, it nevertheless dominates the quantum transport problem in the strong coupling (insulating) phase which is characterised by strong interaction effects such as the appearance of the Coulomb gap in the quasiparticle density of states. This will be the main subject of chapter 3, where we embark on the renormalisation group behaviour of the theory.

2.3.2 $\mathcal{F}$-invariance of the effective action

It is easily checked that the effective action (2.3.3) is indeed $\mathcal{F}$-invariant. For completeness and for later use we will now show how the action transforms under general, spatially *dependent* W-rotations. Using the $\mathcal{F}$-algebra presented in section 2.2.4 we obtain

$$\begin{aligned} S_\sigma[WQW^{-1}] &= -\tfrac{1}{8}\sigma^0_{\rm xx}\mathrm{Tr}\,[\nabla + i\nabla\hat{\chi}, Q]^2 + \tfrac{1}{8}\sigma^0_{\rm xy}\mathrm{Tr}\;\varepsilon_{ij}Q\partial_i Q\partial_j Q \\ &\quad + \tfrac{i}{2}\oint \mathrm{d}x\;\mathrm{tr}\;Q\widehat{\partial_x\chi} + \tfrac{i\beta}{4\pi}\oint \mathrm{d}x\;(\partial_0\chi)^\dagger\partial_x\chi \qquad (2.3.11) \\ S_{\rm F}[WQW^{-1}] &= S_{\rm F}[Q] \qquad (2.3.12) \\ S_{\rm Y}[WQW^{-1}] &= -\tfrac{\pi}{\beta}\sum_{\alpha n}\int \mathrm{d}^2p\;\left[\mathrm{tr}\;\mathrm{I}^\alpha_n Q(\vec{p}) + \tfrac{\beta}{\pi}(\partial_0\chi)^\alpha_{-n}(\vec{p})\right] Y(\vec{p}) \\ &\quad \times\left[\mathrm{tr}\;\mathrm{I}^\alpha_{-n}Q(-\vec{p}) + \tfrac{\beta}{\pi}(\partial_0\chi)^\alpha_n(-\vec{p})\right]. \qquad (2.3.13) \end{aligned}$$

Here $\oint\mathrm{d}x$ stands for integration along the boundary of the sample, with x a coordinate running counterclockwise along the edge. Putting $\nabla\chi = 0$, the $\mathcal{F}$-invariance of all three pieces of the action is evident. For the moment we will discard the boundary and only consider the infinite plane. We will come back to this issue in chapter 4 where we discuss chiral edge states. There, the edge terms in (2.3.11) will play a major role.

The invariance of the Finkelstein part of the action deserves some extra attention. It follows from the transformations (2.2.39, 2.2.40) applied to (2.3.9). In the $\mathcal{F}$-algebra the frequency matrix ω does not transform simply according to the linear rule $\omega \to \omega - \widehat{\partial_0\chi}$ as one would naively expect. Terms quadratic in the gauge field χ are generated such that $S_{\rm F}$ as a whole remains invariant. The results of the $\mathcal{F}$-algebra are therefore somewhat counterintuitive. Another way to check the $\mathcal{F}$-invariance of the Finkelstein term is to consider a gauge transformation of (2.3.5). Under the trace, the W-rotation can be moved from the Q onto the I-matrices. We write $\mathrm{I} \to \mathrm{I}+\delta\mathrm{I}$ with $\delta\mathrm{I} = W^{-1}\mathrm{I}W - \mathrm{I}$. From section 2.2.4 we know that $\delta\mathrm{I}$ exists only in the upper left and lower right corners. Therefore $\delta\mathrm{I}$ commutes with Q, which proves the invariance of the expression $\mathrm{tr}\;[\mathrm{I}^\alpha_n, Q][\mathrm{I}^\alpha_{-n}, Q]$.

In summary we can say that $\mathcal{F}$-invariance of the action can be retained after truncation of Matsubara frequency space. For this purpose we introduced the 'smallness' concept, which ensures that δQ has no contact with the complex non-$U(1)$ character of the I,I commutators. The coupling between the 'large' W-rotations and the 'small' δQ matrix as discussed in this section seems to be the *only consistent way* of carrying through $U(1)$ gauge transformations in the effective action formalism. In a theory which includes external fields, electromagnetic gauge transformations therefore also have to take the form of W-rotations, and the frequencies for which the fields are nonzero have to be restricted as in Fig. 2.3.

The stringent requirements on the cutoff procedure do not, however, provide an answer to the fundamental question of $U(1)$ gauge invariance of the whole theory. More specifically, since the Q and WQW^{-1} do not (by construction) belong to the same manifold, we generally cannot absorb the W-rotation into the measure of the Q-integration. The general idea behind our approach is, however, that full gauge invariance is only obtained after the cutoff $N_{\max}$ is sent to infinity. This way of handling the $U(1)$ gauge invariance is completely new and special care should therefore be taken. The proof of gauge invariance of the Finkelstein theory ultimately relies on the results of explicit, laborious calculations, both perturbative and non-perturbative [64]. This, then, puts extra weight on statements of renormalisability. We will embark on this problem in chapter 3.

From now onward we are going to treat the $\mathcal{F}$-invariance as a good symmetry of the problem, keeping in mind that the limit $N_{\max}\to\infty$ is always taken in the end.

2.3.3 Including external fields

In the previous section we have established the correct procedure for treating $U(1)$ gauge transformations. The task at hand now is to include gauge fields in the effective action (2.3.3). In principle, the straightforward way of doing this would be to start from (2.2.18), truncate, and redo the whole procedure of integrating out P and λ. However, this is a highly nontrivial task! Instead, we are going to use our knowledge of the $\mathcal{F}$-algebra and demand that the extension of (2.3.3) is both gauge and $\mathcal{F}$-invariant,

$$\begin{aligned} S_{\text{eff}}[Q, A_\mu] &= S_{\text{eff}}[e^{i\hat{\chi}} Q e^{-i\hat{\chi}}, A_\mu + \partial_\mu \chi], && \text{arbitrary } \chi \\ S_{\text{eff}}[Q, A_\mu] &= S_{\text{eff}}[e^{i\hat{\chi}} Q e^{-i\hat{\chi}}, A_\mu], && \nabla\chi = 0. \end{aligned} \tag{2.3.14}$$

The χ here does not have a zero frequency component. The zero frequency part of the gauge fields is contained in the P-field theory as a static background. The Q-theory only deals with the fluctuations A_μ.

An important point to notice is that the most naive guess for an extension of (2.3.3), namely the minimal coupling $\partial_\mu \to D_\mu = \partial_\mu - iA_\mu$, does not satisfy (2.3.14). The S_{Y} (2.3.6) in itself is not gauge invariant; the tr $\mathrm{I}_n^\alpha Q$ needs an A_0 term to compensate for the $\partial_0\chi$ that gets generated by the W-rotation.

Another important observation is that the topological term in S_σ *on the infinite plane* is gauge invariant all by itself and does not give us a clue how to couple it to A_μ. When a boundary is included, however, (2.3.11) prescribes a unique coupling. The only expression constructed from the topological term that is both gauge *and* $\mathcal{F}$-invariant is given by

$$\begin{aligned} \text{Invariant} &= \tfrac{1}{8}\varepsilon_{ij} \text{Tr}\, Q[D_i, Q][D_j, Q] \\ &= S_{\text{top}}[Q] - \tfrac{i}{2}\oint \mathrm{d}x \ \text{tr}\ Q\hat{A}_x + \tfrac{i}{4\pi}\left[\sum_\alpha \int_M A^\alpha \wedge dA^\alpha + \beta \oint \mathrm{d}x\ A_0^\dagger A_x\right] \\ &\quad + \tfrac{i}{2}\sum_{n\alpha}\int \mathrm{d}^2x\ B_n^\alpha [\text{tr}\ \mathrm{I}_n^\alpha Q - \tfrac{\beta}{\pi}(A_0)_{-n}^\alpha] \end{aligned} \tag{2.3.15}$$

Here we have used the notation[1]

$$S_{\text{top}}[Q] = \tfrac{1}{8}\text{Tr}\ \varepsilon_{ij} Q\partial_i Q \partial_j Q = \tfrac{1}{2}\oint \mathrm{d}x\ \text{tr}\ (\Lambda T \partial_x T^{-1}) \tag{2.3.16}$$

$$\sum_\alpha \int_M A^\alpha \wedge dA^\alpha = \sum_\alpha \int \mathrm{d}^3x\ \varepsilon^{\mu\nu\lambda} A_\mu^\alpha \partial_\nu A_\lambda^\alpha = \beta\int \mathrm{d}^2x\ [2A_0^\dagger B - \vec{A}^\dagger \times \partial_0 \vec{A}] - \beta\oint \mathrm{d}x\ A_0^\dagger A_x$$

[1]Notice that the topological term gives rise to a Chern-Simons action which is also of topological nature.

with M the manifold consisting of the plane of the sample times the imaginary time interval $[0,\beta]$. With our knowledge of the $\mathcal{F}$-algebra we are now in a position to state that the effective action is restricted to the following form

$$\begin{aligned} S_{\text{eff}}[Q,A_\mu] &= -\tfrac{1}{8}\sigma^0_{\text{xx}}\text{Tr }[\vec{D},Q]^2 + \tfrac{1}{8}\sigma^0_{\text{xy}}\varepsilon_{ij}\text{Tr }Q[D_i,Q][D_j,Q] + S_{\text{F}}[Q] \\ &\quad -\tfrac{\pi}{\beta}\sum_{n\alpha}\int \text{d}^2q\; Y(\vec{q})\left|\text{tr I}^\alpha_n Q(\vec{q}) - \tfrac{\beta}{\pi}(A_0)^\alpha_{-n}(\vec{q})\right|^2 \\ &\quad +\mathcal{F}\text{-invariant terms involving } B \text{ and } \left[\text{tr I}^\alpha_n Q - \tfrac{\beta}{\pi}(A_0)^\alpha_{-n}\right]. \end{aligned} \tag{2.3.17}$$

The details of the additional $\mathcal{F}$-invariant terms follow from a closer study of the edge, which we discuss in chapter 4. There we present a combined effective action for the bulk and the edge. In the limit of zero density of edge states, the following bulk action is obtained.

$$\begin{aligned} &S_{\text{eff}}[Q,A] = \\ &-\tfrac{1}{8}\sigma^0_{\text{xx}}\text{Tr }[\vec{D},Q]^2 + \tfrac{1}{8}\sigma^0_{\text{xy}}\varepsilon_{ij}\text{Tr }Q[D_i,Q][D_j,Q] + S_{\text{F}}[Q] - \tfrac{\beta}{8\pi^2\rho}(\sigma^{\text{II}}_{\text{xy}})^2\int \text{d}^2x\; B^\dagger B \\ &-\tfrac{\pi}{\beta}\sum_{n\alpha}\int \text{d}^2q\; Y\left[\text{tr I}^\alpha_n Q - \tfrac{\beta}{\pi}(A_0)^\alpha_{-n} + \tfrac{i\beta}{2\pi^2\rho}\sigma^{\text{II}}_{\text{xy}}B^\alpha_{-n}\right]\left[\text{tr I}^\alpha_{-n}Q - \tfrac{\beta}{\pi}(A_0)^\alpha_n + \tfrac{i\beta}{2\pi^2\rho}\sigma^{\text{II}}_{\text{xy}}B^\alpha_n\right]. \end{aligned} \tag{2.3.18}$$

The terms containing $\sigma^{\text{II}}_{\text{xy}} \propto \partial n/\partial B$ [61, 63] are the result of the diamagnetic edge currents. We stress that the complete result (2.3.18) clearly demonstrates the theoretical subtleties of the effective action procedure which can not be taken for granted. Without the boundary and the concept of $\mathcal{F}$-invariance we wouldn't have had a clue about the A_μ coupling.

Eq. (2.3.18) is a nontrivial statement on the low energy, large wavelength excitations of the theory. In order to see this, we consider the theory at a classical level, i.e. we put $Q=\Lambda$. The results now represent an effective action for the external fields A_μ which contains the *same* microscopic parameters σ^0_{ij} etc. as those appearing in $S_{\text{eff}}[Q]$ (2.3.3–2.3.6) without external fields. This result is truly remarkable if one realizes that the effective actions $S_{\text{eff}}[A_\mu, Q=\Lambda]$ and $S_{\text{eff}}[A_\mu=0, Q]$ follow from fundamentally different expansion procedures applied to the original theory (2.2.18). In appendix 2A we elaborate further on this point and show that the different expansion procedures are in fact related by Ward identities. These Ward identities are not only crucially important in the microscopic derivation of the general result (2.3.18), they also provide a formal justification of the 'smallness' concept. In appendix 2B we give a simple example and show how the theory (2.3.3–2.3.6) can be obtained in this way.

2.4 Response at tree level

2.4.1 Perturbative expansion

It is straightforward to check that the Q-field theory at a classical level (putting $Q=\Lambda$) does not provide a gauge invariant response to the external fields A_μ. In order to obtain a $U(1)$ invariant result, one has to work with the propagators of the Q-field fluctuations. A $U(1)$ invariant result at a tree level is obtained by taking the Q-field fluctuations to lowest order into account.

The most effective way to proceed is to first make use of a W- or gauge transformation such that the A_0 is absorbed into the vector potential $\vec{A}$. It is easy to check that under such a W-rotation the fields transform according to

$$\begin{aligned} \vec{A}_n^\alpha &\to \vec{z}_n^{\,\alpha} := \vec{A}_n^\alpha + \tfrac{\nabla(A_0)_n^\alpha}{i\nu_n} \\ A_0 &\to 0. \end{aligned} \tag{2.4.1}$$

It is obviously advantageous to deal directly with the gauge invariant quantity $\vec{z}_n^{\,\alpha} = i\vec{E}_n^\alpha/\nu_n$, where $\vec{E}$ is the electric field $(\partial_0\vec{A} - \nabla A_0)$. In order to define a perturbative expansion in the Q-field we write

$$Q = \begin{pmatrix} 1-2VV^\dagger & 2\sqrt{1-VV^\dagger}\cdot V \\ 2V^\dagger\sqrt{1-VV^\dagger} & -1+2V^\dagger V \end{pmatrix}, \tag{2.4.2}$$

where the matrices $V, V^\dagger$ are taken as independent field variables. We use the following convention for the Matsubara indices: the quantities $n_1, n_3, \cdots$ with odd subscripts run over non-negative integers, such that the corresponding fermionic frequencies ω_{n_i} are positive. By the same token, the $n_2, n_4, \cdots$ run over negative integers and the corresponding ω_{n_i} are all negative. The action can be written as a series in powers of the fluctuation fields $V, V^\dagger$. The propagators of the Gaussian theory are given by

$$\begin{aligned} \left\langle V^{\alpha\beta}_{n_1n_2}(p)\ \ [V^\dagger]^{\gamma\delta}_{n_4n_3}(-p')\right\rangle &= \tfrac{4}{\sigma^0_{\mathrm{xx}}}\delta^{\alpha\delta}\delta^{\beta\gamma}\delta(p-p')\delta_{n_{12},n_{34}}f_{n_{12}}(p)\times \\ &\quad\times\left\{\delta_{n_1n_3} + \delta^{\alpha\beta}\kappa[z-Y(p)]g_{n_{12}}(p)\right\} \end{aligned} \tag{2.4.3}$$

$$\begin{aligned} f_m(p) &= \left[p^2+\kappa z m\right]^{-1} \quad ; \quad g_m(p) = \left[p^2+\kappa m Y(p)\right]^{-1} \\ \kappa &= \tfrac{8\pi}{\beta\sigma^0_{\mathrm{xx}}} \quad ; \quad n_{12} = n_1 - n_2. \end{aligned}$$

2.4.2 Zero magnetic field

Putting $\sigma^0_{\mathrm{xy}} = 0$ for simplicity for the moment we obtain the following result for the response at tree level

$$S[A_\mu] = -\sigma^0_{\mathrm{xx}}(\tfrac{\beta}{2\pi})^2\sum_{\alpha,n>0}\int d^2p\ \tfrac{1}{n}\ (E_i)^\alpha_{-n}(-p)\left[\delta_{ij} - \frac{p_ip_j}{p^2+\kappa nY(p)}\right](E_j)^\alpha_n(p) \tag{2.4.4}$$

This expression provides important physical information on the process of quantum transport. In order to show this we write for the charge density N (using $\tau = x^0 = it$)

$$-\beta N^\alpha_m(p) = \frac{\delta S[A_\mu]}{i\delta(A_0)^\alpha_{-m}(-p)} = -\sigma^0_{\mathrm{xx}}(\tfrac{\beta}{2\pi})^2\tfrac{1}{m}p_i\left[\delta_{ij} - \frac{p_ip_j}{p^2+\kappa mY(p)}\right](E_j)^\alpha_m(p) \tag{2.4.5}$$

which can be written as (omitting replica indices)

$$[\nu_m + \tfrac{1}{4}\sigma^0_{\mathrm{xx}}p^2\tfrac{1}{Y(p)}]N_m(p) = -i\vec{p}\cdot(\vec{\jmath}_{\mathrm{ext}})_m(p). \tag{2.4.6}$$

Here we have written $\vec{\jmath}_{\mathrm{ext}} = \frac{i}{2\pi}\sigma^0_{\mathrm{xx}}\vec{E}$ for the current density generated by the external field. The quantity containing Y in (2.4.6) has a different meaning depending on the momentum scale one is looking at. For large momenta we have

$$\begin{aligned} &1/Y(p) \to 2/\pi\rho \quad \text{as } |p|\to\infty \\ &\tfrac{1}{4}\sigma^0_{\mathrm{xx}}p^2\tfrac{1}{Y(p)}N(p) \approx \vec{p}\cdot D^0_{\mathrm{xx}}\vec{p}N(p) = i\vec{p}\cdot\vec{\jmath}_{\mathrm{diff}}(p) \end{aligned} \tag{2.4.7}$$

where $D^0_{\rm xx}$ is the diffusion constant, equal to $\sigma^0_{\rm xx}/2\pi\rho$ by the Einstein relation, and $\vec{\jmath}_{\rm diff}$ stands for the diffusive current density, which is driven by the density gradient. In space-time notation we then have, instead of (2.4.6),

$$\partial_t N = -\nabla\cdot(\vec{\jmath}_{\rm ext} + \vec{\jmath}_{\rm diff}) \qquad (|p|\to\infty) \tag{2.4.8}$$

which is the usual statement of particle number conservation in a metallic system driven by density gradients and external electric fields. On the other hand, in the limit of small momenta the $1/Y$ reduces to the bare Coulomb potential

$$\begin{aligned} &1/Y(p) \to 2\pi\rho U_0(p) \quad \text{as} \quad |p|\to 0 \\ &\tfrac{1}{4}\sigma^0_{\rm xx}p^2\tfrac{1}{Y(p)}N(p) \approx \vec{p}\cdot\tfrac{1}{2\pi}\sigma^0_{\rm xx}\vec{E}_{\rm c}(p) = i\vec{p}\cdot\vec{\jmath}_{\rm c}(p). \end{aligned} \tag{2.4.9}$$

Here, $\vec{E}_{\rm c}$ is the internally generated electric field of the interacting electron gas and $\vec{\jmath}_{\rm c}$ the current induced by this field. Instead of (2.4.8) we now get

$$\partial_t N = -\nabla\cdot(\vec{\jmath}_{\rm ext} + \vec{\jmath}_{\rm c}) \qquad (|p|\to 0) \tag{2.4.10}$$

expressing the fact that the system only responds to the sum of external and internally generated electric fields. The instantaneous Coulomb potential apparently wins over the much slower diffusive processes. The continuity equation (2.4.6) is a well known result from the theory of metals [56].

2.4.3 Including magnetic fields

Now we are going to include the $\sigma^0_{\rm xy}$ term. Let us first consider weak magnetic fields, in which case the $\sigma^{\rm II}_{\rm xy}$ can be neglected [63] and edge effects become immaterial. This leads to the following gauge invariant response

$$\begin{aligned} S[A_\mu] &= -(\tfrac{\beta}{2\pi})^2\sum_{\alpha,n>0}\int \mathrm{d}^2p\,\tfrac{1}{n}\,(E_i)^\alpha_{-n}(-p)\left[\sigma^0_{\rm xx}\delta_{ij} + \sigma^0_{\rm xy}\varepsilon_{ij}\right](E_j)^\alpha_n(p) \\ &+ \tfrac{1}{\sigma^0_{\rm xx}}(\tfrac{\beta}{2\pi})^2\sum_{\alpha,n>0}\int \mathrm{d}^2p\,\tfrac{g_n(p)}{n}\,\left[\sigma^0_{\rm xx}\vec{p}\cdot\vec{E}^\alpha_{-n} - \sigma^0_{\rm xy}\nu_n B^\alpha_{-n}\right](-p)\left[\sigma^0_{\rm xx}\vec{p}\cdot\vec{E}^\alpha_n + \sigma^0_{\rm xy}\nu_n B^\alpha_n\right](p). \end{aligned} \tag{2.4.11}$$

If we now repeat the calculation of the electron density in section 2.4 using the action (2.4.11), we find that the results (2.4.6)–(2.4.10) still hold, with one modification: The 'external' current $\vec{\jmath}_{\rm ext}$ and the internally generated current $\vec{\jmath}_c$ now also include a Hall current,

$$j^i_{\rm c/ext} = \tfrac{i}{2\pi}[\sigma^0_{\rm xx}\delta^{ij} + \sigma^0_{\rm xy}\varepsilon^{ij}]E^j_{\rm c/ext}. \tag{2.4.12}$$

The modification of $\vec{\jmath}_c$ is not apparent in the equation for current conservation, however, since

$$\nabla\cdot\vec{\jmath}^{\,\rm Hall}_c(p) \propto \varepsilon_{ij}p_ip_j = 0. \tag{2.4.13}$$

Let us now consider strong magnetic fields. In this case we can no longer neglect the $\sigma^{\rm II}_{\rm xy}$ contributions in (2.3.18). The action (2.4.11) is modified as follows. First, the $(\sigma^{\rm II}_{\rm xy})^2$ term is carried over from (2.3.18). Second, there is a shift $A_0 \to A_0 + i\frac{\sigma^{\rm II}_{\rm xy}}{2\pi\rho}B$, since the A_0 is contained only in the Y-term. The result is

$$S_{\sigma^{\rm II}_{\rm xy}}[B,\vec{E}] = S_{\sigma^{\rm II}_{\rm xy}=0}[B,\vec{E} - i\tfrac{\sigma^{\rm II}_{\rm xy}}{2\pi\rho}\nabla B] - \tfrac{\beta}{8\pi^2\rho}(\sigma^{\rm II}_{\rm xy})^2\int \mathrm{d}^2p\; B^\dagger B. \tag{2.4.14}$$

The ensuing externally generated current is given by

$$j^i_{\rm ext} = \tfrac{1}{2\pi}[\sigma^0_{\rm xx}\delta^{ij} + \sigma^0_{\rm xy}\varepsilon^{ij}](iE^j_{\rm ext} + \tfrac{\sigma^{\rm II}_{\rm xy}}{2\pi\rho}\partial_j B) \tag{2.4.15}$$

For convenience later on, we write the result for weak B (2.4.11) in terms of new variables Φ, Ψ

$$z_i = \partial_i\Phi + \varepsilon_{ij}\partial_j\Psi \tag{2.4.16}$$

$$S[\Phi,\Psi] = -\sigma^0_{\rm xx}\sum_{\alpha,n>0}\int \mathrm{d}^2p\; np^2(\Phi^*,\Psi^*)M\begin{pmatrix}\Phi\\ \Psi\end{pmatrix}. \tag{2.4.17}$$

The 2×2 matrix M is given by

$$M = \begin{bmatrix}\mathcal{G} & -\omega_c\tau\mathcal{G}\\ \omega_c\tau\mathcal{G} & 1+(\omega_c\tau)^2(1-\mathcal{G})\end{bmatrix} \qquad \text{with} \qquad \mathcal{G} = 1 - p^2 g_n(p) \tag{2.4.18}$$

where we have made use of the semiclassical notation

$$\sigma^0_{\rm xx} = \frac{\sigma_0}{1+(\omega_c\tau)^2} \qquad ; \qquad \sigma^0_{\rm xy} = \omega_c\tau\;\sigma^0_{\rm xx}. \tag{2.4.19}$$

We will use the form (2.4.17) in chapter 5 where we perform the Chern-Simons mapping.

2A Justification of 'smallness'

Why is it possible to work with a matrix T that contains only small Matsubara frequencies? One can of course start out from an ungauged action (i.e. without λ and A_μ) in which T has the 'small' form. However, in our theory with λ and A_μ we have to allow $U(1)$ gauge transformations and one would expect that a W-rotation, which is 'large', changes T into a 'large' matrix. In this appendix we show that that does not happen. Thanks to the special structure of the $\mathcal{F}$-algebra the 'small' form is retained.

Consider first how $T = \mathbb{1}$ changes under a gauge transformation. We have $T \to W^{-1} = \exp -i\hat{\chi}$. We write

$$\hat{\chi} = \hat{\chi}_B + \hat{\chi}_S \tag{2A.1}$$

where $\hat{\chi}_B$ is block diagonal (nonzero only in the $[+,+]$ and $[-,-]$ Matsubara blocks) and $\hat{\chi}_S$ is block off-diagonal (nonzero in the $[+,-]$ and $[-,+]$ blocks). Now write

$$W = e^{-i\hat{\chi}} = e^{-i\hat{\chi}_B}\; t. \tag{2A.2}$$

Using the Campbell-Baker-Hausdorff formula, t can be written as the exponent of a power series as follows:

$$t = \exp i\sum_{n=1}^{\infty}\tfrac{1}{n!}X_n \tag{2A.3}$$

where

$$X_1 = \hat{\chi}_S \qquad X_2 = [\hat{\chi}_S, \hat{\chi}_B] \qquad X_3 = \tfrac{1}{2}[[\hat{\chi}_S, \hat{\chi}_B], \hat{\chi}_B] \qquad \ldots \tag{2A.4}$$

The important point is that the X_n are all block off-diagonal matrices and that their size in frequency space is of the order of $n\cdot N_{\rm max}$. The series can be truncated beyond small orders in n, resulting in a matrix t that satisfies the condition of smallness. Notice that the restriction of χ_n to a narrow band of frequencies is an essential ingredient here. In the theory with variables P, T the large component $e^{-i\hat{\chi}_B}$ can be absorbed into a redefinition of the P-field which leads to the statement

$$T^{-1}PT = WPW^{-1} = t^{-1}P't. \tag{2A.5}$$

In the theory where P has been integrated out, we get

$$W\Lambda W^{-1} = t^{-1}e^{i\hat{\chi}_B}\Lambda e^{-i\hat{\chi}_B}t = t^{-1}\Lambda t. \tag{2A.6}$$

This shows that a W-rotation acting on $T = \mathbb{1}$ effectively results in a 'small' T. Let us next consider the effect of a W-rotation on a general 'small' T. Here we have the transformation $T \to Te^{-i\hat{\chi}_B}t$. It is easily checked that 'small' $\times$ 'large' block diagonal $\times$ 'small' = 'small'. This then shows that a W-rotation acting on a general 'small' T again results in a 'small' matrix, which provides a justification for the whole truncation procedure.

2B Matching procedure

Let us consider the action (2.3.1) with A_μ included and reflect on the possibility of constructing an effective action $S_{\text{eff}}[A_\mu, T]$ which contains the two distinctly different sets of field variables A_μ and T simultaneously. For this purpose we consider the effective actions $S_{\text{eff}}[A_\mu, T{=}\mathbb{1}]$ and $S_{\text{eff}}[A_\mu{=}0, T]$ separately.

$$\exp S_{\text{eff}}[A_\mu, \mathbb{1}] \propto \tag{2B.1}$$
$$\int \mathcal{D}\lambda\mathcal{D}P\; I[P]\; \exp\left(-\tfrac{1}{2g}\text{Tr}\; P^2 - \tfrac{\beta}{2}\int \lambda^\dagger U_0^{-1}\lambda + \text{Tr}\; \ln[i\omega + \mu + i\hat{A}_0 + i\hat{\lambda} - \hat{\mathcal{H}} + iP]\right)$$
$$\exp S_{\text{eff}}[0, T] \propto \tag{2B.2}$$
$$\int \mathcal{D}\lambda\mathcal{D}P\; I[P]\; \exp\left(-\tfrac{1}{2g}\text{Tr}\; P^2 - \tfrac{\beta}{2}\int \lambda^\dagger U_0^{-1}\lambda + \text{Tr}\; \ln[i\omega + \mu + i\hat{\lambda} - \hat{\mathcal{H}}_0 + iT^{-1}PT]\right).$$

The $S_{\text{eff}}[A_\mu, T{=}\mathbb{1}]$ is formally obtained by expanding the Tr ln to lowest order in powers of the large matrices $\hat{A}_\mu$ and this means that complicated infinite sums over frequencies will have to be performed. The situation for $S_{\text{eff}}[A_\mu{=}0, T]$ is quite different and one has to follow the procedure of [58] in order to formally express this action in terms of the 'small' variable δQ to lowest orders in a derivative and temperature expansion.

However, since one is usually interested in the limit of small momenta, frequencies and temperatures, only a finite number of terms in S_{eff} needs to be considered in both cases. The coefficients are microscopic parameters which are generally given as complicated correlations defined by the underlying theory with plasmon (λ) and disorder (P) variables alone. These coefficients (coupling constants) of $S_{\text{eff}}[A_\mu, \mathbb{1}]$ and $S_{\text{eff}}[0, T]$ are related to one another by gauge invariance, as will be shown next, and this then provides the starting point for constructing a complete action $S_{\text{eff}}[A_\mu, T]$, which is done by 'matching' the known results for both pieces.

To establish a relation between (2B.1) and (2B.2) we start out by taking a pure gauge $A_\mu{=}\partial_\mu\chi$, $T{=}e^{-i\hat{\chi}}$. With this choice, the actions (2B.1) and (2B.2) certainly stand for one and the same thing,

$$S_{\text{eff}}[\partial_\mu\chi, \mathbb{1}] = S_{\text{eff}}[0, T]. \tag{2B.3}$$

This procedure can be extended as follows: Suppose we have found $S_{\text{eff}}[A_\mu, T]$ from a 'matching' procedure as mentioned above. A useful check upon this result is obtained by a generalisation of (2B.3),

$$S_{\text{eff}}[A_\mu + \partial_\mu \chi, T] = S_{\text{eff}}[A_\mu, \mathbb{1}]. \tag{2B.4}$$

In order to give an example of the matching procedure, we derive an effective action for the plasmon field λ in the absence of A_μ. We consider first the Tr ln term,

$$\begin{aligned}
\exp S_{\ln}[\lambda, \mathbb{1}] &= \int \mathcal{D}P\; I[P] \exp \text{Tr } \ln[i\omega + \mu - \hat{\mathcal{H}}_0 + i\hat{\lambda} + iP] && \text{(2B.5)}\\
\exp S_{\ln}[0, T] &= \int \mathcal{D}P\; I[P] \exp \text{Tr } \ln[i\omega + \mu - \hat{\mathcal{H}}_0 + iT^{-1}PT]. && \text{(2B.6)}
\end{aligned}$$

Eq. (2B.6) is precisely the free particle problem and we can write

$$S_{\ln}[0, T] = S_\sigma[Q] + \tfrac{2\pi}{\beta}\pi\rho_0 \text{Tr } \eta Q \tag{2B.7}$$

where ρ_0 is the free particle density of states, which can be written as

$$\rho_0 = \tfrac{i}{2\pi}\left\langle G^{\alpha\alpha}_{n_1 n_1}(x,x) - G^{\alpha\alpha}_{n_2 n_2}(x,x)\right\rangle_{\text{av}} \quad ; \quad G(x,x') = \langle x|(i\omega + \mu - \hat{\mathcal{H}}_0 + iP)^{-1}|x'\rangle \tag{2B.8}$$

where the average is with respect to P. In (2B.8) the indices are kept fixed with $n_1 > 0$ and $n_2 < 0$ as usual. Eq. (2B.8) is identical to the more familiar expression for ρ_0 from the standard replica field theory. More specifically, for quantities like (2B.8) which involve unmixed averages over the positive and negative blocks of P, one can transform the problem back and trade in the P-integral for the average over the original random potential $V(\vec{x})$. On the other hand, (2B.5) can be written, to lowest order in λ,

$$S_{\ln}[\lambda, \mathbb{1}] = -\tfrac{1}{2}\sum_{\alpha\beta}\sum_{nm}\int \lambda^\alpha_n M^{\alpha\beta}_{nm}\lambda^\beta_m \tag{2B.9}$$

where

$$\begin{aligned}
M^{\alpha\beta}_{nm}(x,x') &= -\text{tr } \left\langle \hat{G}(x,x')\, \mathrm{I}^\beta_m \hat{G}(x',x)\, \mathrm{I}^\alpha_n \right\rangle_{\text{av}} + \left\langle \text{tr } [\hat{G}(x,x)\, \mathrm{I}^\beta_m]\text{tr } [\hat{G}(x',x')\, \mathrm{I}^\alpha_n] \right\rangle_{\text{cum}}\\
\hat{G}(x,x') &= \langle x|(i\omega + \mu - \hat{\mathcal{H}}_0 + iP)^{-1}|x'\rangle && \text{(2B.10)}
\end{aligned}$$

where the subscript 'cum' stands for the cumulant average. This expression can be worked out to give

$$M^{\alpha\gamma}_{nm}(x,x') = \delta^{\alpha\gamma}\delta_{n+m,0}\delta(x-x')\beta\rho_0 + \cdots \tag{2B.11}$$

where the dots stand for higher order terms in derivatives and frequencies. Only the $[+,+]$ and $[-,-]$ terms in (2B.9) contribute to (2B.11). The mixed components $[+,-]$ and $[-,+]$ are proportional to the frequency (n) and contribute to higher orders only. Eq. (2B.9) can now be written as

$$S_{\ln}[\lambda, \mathbb{1}] = -\tfrac{\beta}{2}\rho_0 \int \lambda^\dagger \lambda \tag{2B.12}$$

The 'match' between (2B.12) and (2B.7) is obtained as follows

$$S_{\ln}[\lambda, T] = S_\sigma[Q] + \tfrac{\pi}{\beta}2\pi\rho_0 \int \left[\text{tr } \eta Q + \tfrac{\beta}{2\pi}\text{tr } \hat{\lambda} Q - (\tfrac{\beta}{2\pi})^2 \lambda^\dagger \lambda\right] \tag{2B.13}$$

It can be shown that this expression satisfies (2B.4) with A_0 replaced by λ. Eq. (2B.13) therefore is the desired result. Moreover, comparison with (2.2.40) shows that (2B.13) is $\mathcal{F}$-invariant. Next we can add the Coulomb part $\propto \lambda^\dagger U_0 \lambda$ to $S_{\rm ln}$. By making the appropriate shift $\lambda \to \lambda'$, the final result decouples such that

$$\begin{aligned} S_{\rm eff}[\lambda', T] &= -\tfrac{\beta}{4\pi} \int (\lambda')^\dagger (\tfrac{1}{U_0} + 2\pi\rho_0)\lambda' + S_\sigma[Q] + \tfrac{2\pi}{\beta}\pi\rho_0 \mathrm{Tr}\; \eta Q \\ & \quad + \tfrac{\pi}{4\beta} 2\pi\rho_0 \int (1 - \tfrac{1}{1+2\pi\rho_0 U_0}) \sum_{n\alpha} |\mathrm{tr}\; \mathrm{I}_n^\alpha Q|^2 \,. \end{aligned} \tag{2B.14}$$

This is precisely of the form written in (2.3.3–2.3.6). We mention that the procedure can be extended to higher orders [6], which means that higher orders in λ can be considered in (2B.9) along with higher order operators in Q (proportional to β^{-2}) in (2B.7). These higher order contributions renormalise the coefficients in (2B.14) and they generate higher order operators as well.

Chapter 3

Renormalisation of the bulk theory

The fact that something is infinite
does not necessarily mean it is zero.
– S. Weinberg

3.1 Introduction

In this chapter we present the results of a renormalisation group study of the perturbative weak coupling regime. Our main objective is to further investigate the $\mathcal{F}$-invariance and to establish the renormalisability of the theory to two loop order. In section 3.2 we present a different version of Finkelstein's momentum shell computations [22], namely the background field method in dimensional regularisation, which leads to major computational advantages. One of the most important aspects of the RG procedure is that the infrared behaviour of the theory can be extracted from $\mathcal{F}$-invariant quantities or correlation functions only. For arbitrary correlation functions or RG procedures which are not $\mathcal{F}$-symmetric, the perturbative expansions are plagued by infrared problems which cannot be resolved by the RG. This means, for example, that the expansion procedure cannot be used to show that the diffusion propagator (which is *not* an $\mathcal{F}$-invariant quantity) contains an infrared cutoff of the type $1/\tau_{\text{in}}$, i.e. the inelastic scattering length as naively obtained from the Golden Rule. For the same reason it is also fundamentally incorrect to interpret the theory in terms of a 'Fermi-liquid with length scale dependent parameters' [12].

Examples of $\mathcal{F}$-invariant quantities are the linear response formulae which involve current and density correlation functions and, of course, the free energy or grand canonical potential itself. The most important results of this chapter are presented in sections 3.3 and 3.4 where we compute these quantities and in section 3.5 where we derive explicit scaling functions for them.

Based on a two-loop expansion for the free energy (section 3.4.4 and appendix 3A) we are able to identify a new quantity in the problem, namely a bosonic quasiparticle density of states which enters the expression for the specific heat (section 3.6). This quantity develops a Coulomb gap as one approaches the metal-insulator transition in $2+2\varepsilon$ dimensions from the metallic side (section 3.6.2). We identify this quasiparticle density of states, rather than the tunnelling density of states, as the physical observable to which the heuristic arguments of Efros and Shklovskiĭ [18] are applicable.

As a logical follow-up we then discuss the problem of the strong coupling, insulating phase (section 3.6.4). We show that important progress can be made by relying on the analogy with the more familiar theory of the Heisenberg ferromagnet. Our results indicate that the insulating phase is dominated by additional terms in the action which are usually dismissed as 'irrelevant'. This shows the necessity of an extended RG program.

3.2 Background field renormalisation

3.2.1 The background field method

Let us start with the action $S_{\text{eff}}[Q]$ (2.3.3). We drop the topological term, since it is not going to contribute to perturbation theory. At the moment we are not interested in the full significance of the S_{Y} term, but only in the low-momentum limit. Since S_{Y} stands for a higher dimensional operator, it is preferable to work with a simpler theory in which the Y is replaced by a δ-function. For reasons to be explained shortly, we add a $U(N)\times U(N)$ invariant regulator ($N = N_{\max}\cdot N_r$) proportional to h_0^2. This gives

$$S[Q] = -\tfrac{\sigma_0}{8}\text{Tr}(\nabla Q)^2 + \tfrac{\pi}{2\beta}z_0{\sum_{\alpha n}}' \text{Tr}\ [\text{I}_n^\alpha, Q][\text{I}_{-n}^\alpha, Q] - \tfrac{\pi}{\beta}z_0\alpha_0\int_x\sum_{\alpha n}|\text{tr}\ \text{I}_n^\alpha Q|^2 + \tfrac{\sigma_0 h_0^2}{4}\text{Tr}\ \Lambda Q. \tag{3.2.1}$$

In the place of Y we have introduced a parameter α_0 times z_0 such that the theory interpolates between the Coulomb case ($\alpha_0 = 0$) and the free particle case ($\alpha_0 = 1$). Let us now define

$$\widehat{Q} = T_0^{-1}QT_0, \tag{3.2.2}$$

where T_0 is a 'small' but fixed and slowly varying background field of size, say, $2n_{\max}\times 2n_{\max}$, where $n_{\max}\ll N_{\max}$. Eq. (3.2.2) is loosely interpreted as a change of variables where T_0 stands for the 'slow' modes which should be kept and Q represents the 'fast' modes which should be eliminated. This 'change of variables' idea is clearly somewhat cavalier and complications arise in pursuing the theory beyond one-loop order. It is the purpose of this section however to show that the basic idea can be put to work. If we keep working with a fixed but 'small' background field, then we can employ the more powerful method of dimensional regularisation and compute the effective action for the T_0-field. Everywhere in (3.2.1) except in the regulating Tr ΛQ term we make the replacement $Q\to\widehat{Q}$. The effective action for T_0 is given by

$$\exp S_{\text{eff}}[T_0] = \int\mathcal{D}Q\ e^{S[Q,T_0]}. \tag{3.2.3}$$

As was shown in detail in the context of the ordinary σ model [66], this results in a very effective way of extracting the pole terms in ε and the RG coefficients of the theory in $2+2\varepsilon$ dimensions. In addition to this we can give precise meaning to the idea of 'small' background field T_0. After the replacement $Q\to\widehat{Q}$ has been made in (3.2.1), it is readily seen that the theory is invariant under a local transformation $T_0\to UT_0$, where $U\in U(N)\times U(N)$. By switching to a new integration variable $Q' = U^{-1}QU$ the action takes its old form, except for the regulating term which becomes $\propto$ Tr $\Lambda UQ'U^{-1}$. However, we have $U^{-1}\Lambda U = \Lambda$ since $U\in U(N)\times U(N)$ and therefore the regulating term is also invariant. We conclude that $S_{\text{eff}}[T_0]$ can be expressed in terms of the local quantity

$$Q_0 = T_0^{-1}\Lambda T_0. \tag{3.2.4}$$

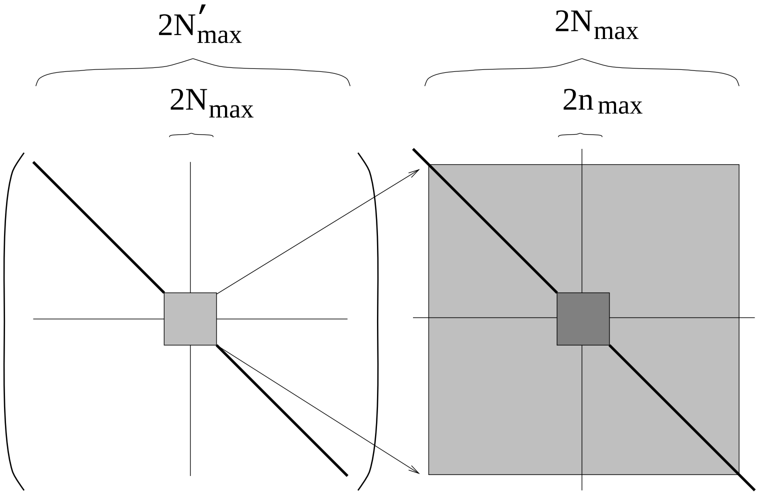

Figure 3.1: *The relative sizes of Q and the background field matrix Q_0.*

This quantity is of the same form as the original variable Q (2.2.33) except that the sizes $N'_{\max}$, $N_{\max}$ are now replaced by $N_{\max}$ and $n_{\max}$ respectively (see Fig. 3.1). In other words, the requirement that the form of the original action and the effective action in Q_0 are the same provides additional information on the renormalisation. Although the basic idea is straightforward, the scheme turns out to be quite rich and subtle. One remarkable thing is that the singlet interaction term (proportional to $1-\alpha_0$) cannot be treated as an operator insertion as one would naively expect. It affects the ultraviolet singularity structure of the problem (i.e. poles in ε) and, hence, changes the beta functions of the theory. This eventually happens in the limit where $N_{\max}$ is sent to infinity. This limit is taken in such a way that the quantity $W = T \cdot N_{\max}$ remains finite, and this corresponds physically to having a finite 'effective' bandwidth W.

Let us now write the Finkelstein term in the form of an interaction term plus a frequency term as in (2.3.9), neglecting the constant $\propto \mathrm{tr}\, \eta\Lambda$, and insert the background field T_0 everywhere except in the h_0^2 term. The result can be written as

$$S[Q,T_0] = S_{\text{kin}}[T_0^{-1}QT_0] + S_{\text{int}}[T_0^{-1}QT_0] + S_{\text{freq}}[T_0^{-1}QT_0] + \tfrac{\sigma_0 h_0^2}{4}\mathrm{Tr}(\Lambda Q), \qquad (3.2.5)$$

where 'kin', 'int' and 'freq' denote the kinetic, interaction and frequency part of the action. We have

$$\begin{aligned} S_{\text{kin}}[T_0^{-1}QT_0] &= -\tfrac{\sigma_0}{8}\mathrm{Tr}\left[\nabla + \vec{C}, Q\right]^2, && (3.2.6)\\ S_{\text{int}}[T_0^{-1}QT_0] &= \tfrac{\pi}{\beta} z_0(1-\alpha_0)\int \mathrm{d}^D x \sum_{\alpha n} \mathrm{tr}\,(\mathrm{I}_n^\alpha + J_n^\alpha)Q\ \mathrm{tr}\,(\mathrm{I}_{-n}^\alpha + J_{-n}^\alpha)Q, && (3.2.7)\\ S_{\text{freq}}[T_0^{-1}QT_0] &= \tfrac{4\pi}{\beta} z_0\ \mathrm{Tr}\,(\eta + H)Q. && (3.2.8) \end{aligned}$$

Here all T_0 dependence is collected into the 'potentials' $\vec{C}$, J_n^α and H

$$\vec{C} = T_0 \nabla T_0^{-1} \qquad J_m^\alpha = T_0\left[\mathrm{I}_m^\alpha, T_0^{-1}\right] \qquad H = T_0\left[\eta, T_0^{-1}\right]. \qquad (3.2.9)$$

The effective action $S_{\text{eff}}[T_0]$ is going to be obtained by treating the 'potentials' as a perturbation about the original action (3.2.1). It is easy to see, however, that this procedure breaks the $U(N)\times U(N)$ gauge invariance, which means that the final result cannot be expressed in terms of the local variable Q_0 (3.2.4). In order to retain $U(N)\times U(N)$ gauge invariance it will be necessary to drop all temperature and frequency dependence as infrared regulators such that the parameters in the effective theory only depend on the regulating field h_0^2. We will come back to this problem at a later stage (section 3.2.4).

3.2.2 Perturbation expansion

The theory (3.2.5) is going to be worked out perturbatively. Writing

$$Q = \begin{pmatrix} 1-2VV^\dagger & 2\sqrt{1-VV^\dagger}\cdot V \\ 2V^\dagger\sqrt{1-VV^\dagger} & -1+2V^\dagger V \end{pmatrix}, \tag{3.2.10}$$

the theory can be written as an infinite power series in the matrix fields V, $V^\dagger$ which contain $N\times N$ independent complex field variables. Notice that classically ($Q=\Lambda$) the effective action $S[\Lambda,T_0]$ is of the same form as the original one (3.2.1) except that the $\mathrm{Tr}(\Lambda Q)$ term is lacking.

In order to discuss the theory on a quantum level we regroup the various quantities in the background field action (3.2.5) according to

$$\begin{aligned} S[Q,T_0] &= S[Q]+S[T_0^{-1}\Lambda T_0]+\int d^Dx\left\{-\tfrac{\sigma_0}{4}\cdot O_{\mathrm{kin}}(Q,T_0)+\tfrac{4\pi}{\beta}z_0\cdot O_{\mathrm{freq}}(Q,T_0)\right. \\ &\quad \left. +\tfrac{\pi}{\beta}z_0[1-\alpha_0]\cdot O_{\mathrm{int}}(Q,T_0)\right\}. \end{aligned} \tag{3.2.11}$$

We split O_{kin} and O_{int} into contributions containing different powers of $\delta T_0 = T_0 - \mathbb{1}$ and δQ,

$$O_{\mathrm{kin}} = O_{\mathrm{kin}}^{(1)} + O_{\mathrm{kin}}^{(2),1} + O_{\mathrm{kin}}^{(2),2} \quad ; \quad O_{\mathrm{int}} = O_{\mathrm{int}}^{(1),1} + O_{\mathrm{int}}^{(1),2} + O_{\mathrm{int}}^{(2),1} + O_{\mathrm{int}}^{(2),2}. \tag{3.2.12}$$

The superscript between parentheses denotes the lowest occurring power of δT_0. The second superscript gives the power of δQ. The eight different operators O are listed in Table 3.1.

$O_{\mathrm{kin}}^{(1)} = 2\mathrm{tr}(C_i\delta Q\partial_i\delta Q)$	$O_{\mathrm{int}}^{(1),1} = 2\sum_{m,\alpha}\mathrm{tr}(\mathrm{I}_{-m}^{\alpha}Q_0)\mathrm{tr}(\mathrm{I}_{m}^{\alpha}\delta Q)$
$O_{\mathrm{kin}}^{(2),1} = 2\mathrm{tr}(\delta Q C_i\Lambda C_i)$	$O_{\mathrm{int}}^{(1),2} = 2\sum_{m,\alpha}\mathrm{tr}(\mathrm{I}_{-m}^{\alpha}\delta Q)\mathrm{tr}(J_{m}^{\alpha}\delta Q)$
$O_{\mathrm{kin}}^{(2),2} = \mathrm{tr}(C_i\delta Q C_i\delta Q)$	$O_{\mathrm{int}}^{(2),1} = 2\sum_{m,\alpha}\mathrm{tr}(\mathrm{I}_{-m}^{\alpha}Q_0)\mathrm{tr}(J_{m}^{\alpha}\delta Q)$
$O_{\mathrm{freq}} = \mathrm{tr}(H\delta Q)$	$O_{\mathrm{int}}^{(2),2} = \sum_{m,\alpha}\mathrm{tr}(J_{-m}^{\alpha}\delta Q)\mathrm{tr}(J_{m}^{\alpha}\delta Q)$

Table 3.1: *Contributions to the background field action.*

The propagator for the V-field can be read from the quadratic part of the action (3.2.1),

$$S^{(2)}[Q] = -\sigma_0\sum_{n>0}\sum_{\alpha\beta}\int \mathrm{d}^2p\sum_{a,b=0}^{n-1}[V^\dagger]^{\beta\alpha}_{b-n,b}(-p)V^{\alpha\beta}_{a,a-n}(p)\left\{\delta_{ab}[p^2+h_0^2+\kappa z_0 n]-\delta^{\alpha\beta}\kappa z_0 c_0\right\}. \tag{3.2.13}$$

Here we have defined

$$\kappa = \tfrac{8\pi}{\beta\sigma_0} \quad ; \quad c_0 = 1-\alpha_0. \tag{3.2.14}$$

Notice that (3.2.13) separates into independent actions. At fixed (n,α,β) we can define an n-component vector field $(M_n^{\alpha\beta})_a = V_{a,a-n}^{\alpha\beta}$ for which the expression between brackets $\{\cdots\}$ can be written as the sum of a unit matrix and a projection matrix ('quasi-unit-matrix'). This leads to the following propagator,

$$\left\langle V_{n_1 n_2}^{\alpha\beta}(p)\;[V^\dagger]_{n_4 n_3}^{\delta\gamma}(-p)\right\rangle = \tfrac{1}{\sigma_0}\delta^{\alpha\gamma}\delta^{\beta\delta}\delta_{n_1-n_2,n_3-n_4} D_p(n_{12})\left\{\delta_{n_1 n_3} + \delta^{\alpha\beta}\kappa z_0 c_0 D_p^c(n_{12})\right\}, \tag{3.2.15}$$

with

$$D_p(n_{12}) = \frac{1}{p^2+h_0^2+\kappa z_0 n_{12}} \qquad ; \qquad D_p^c(n_{12}) = \frac{1}{p^2+h_0^2+\kappa z_0\alpha_0 n_{12}}$$

$$n_{12} = n_1 - n_2 > 0.$$

Here and in what follows we use the following convention: Matsubara indices with odd subscripts run only over nonnegative values, $n_{\mathrm{odd}} \in \{0,\cdots,N_{\max}-1\}$, while those with even subscripts run only over negative values, $n_{\mathrm{even}} \in \{-N_{\max},\cdots,-1\}$. This choice is made in order to incorporate the structure (3.2.10), where it is clear that the first and second index of V have to be nonnegative and negative respectively. (And vice versa for $V^\dagger$).

In the following sections we present the results of a computation that shows that the final results for $S_{\mathrm{eff}}[T_0]$ are of the same form as the original action (3.2.1) with, however, modified 'effective' parameters σ', z' and α' instead of σ_0, z_0 and α_0.

3.2.3 Renormalisation of σ_0

The contributions to the conductivity σ' have four different sources. Schematically, they come from the contractions in the following terms: $\langle O_{\mathrm{kin}}^{(2),1}\rangle$, $\langle [O_{\mathrm{kin}}^{(1)}]^2\rangle$, $\langle [O_{\mathrm{int}}^{(1),2}]^2\rangle$ and $\langle O_{\mathrm{kin}}^{(1)} O_{\mathrm{int}}^{(1),2}\rangle$. The last two of these contributions involve a small momentum expansion in the background field T_0. In Table 3.2 we present the various contributions to the $1/\varepsilon$ pole terms obtained by working in $D=2+2\varepsilon$ dimensions and using dimensional regularisation. The most important terms generated by the background field procedure are $(C_i)_{n_1 n_2}^{\alpha\beta}(C_i)_{n_2 n_1}^{\beta\alpha}$, which are indicated by tr $C_i^{+-}C_i^{-+}$ in Table 3.2, the $(C_i)_{n_1 n_3}^{\alpha\beta}(C_i)_{n_3 n_1}^{\beta\alpha}$, which are denoted as tr $C_i^{++}C_i^{++}$ and finally those obtained by interchanging the positive and negative Matsubara frequency indices. Notice that the first terms can be written in σ-model form

$$\mathrm{tr}\;\left(C_i^{+-}C_i^{-+} + C_i^{-+}C_i^{+-}\right) = -\tfrac{1}{4}\mathrm{tr}\;(\nabla Q_0)^2. \tag{3.2.16}$$

	$\langle O_{\mathrm{kin}}^{(2),1}\rangle$	$-\frac{1}{2}\langle [O_{\mathrm{kin}}^{(1)}]^2\rangle$	$-\frac{1}{2}\langle [O_{\mathrm{int}}^{(1),2}]^2\rangle$	$-\langle O_{\mathrm{kin}}^{(1)} O_{\mathrm{int}}^{(1),2}\rangle$
$\mathrm{tr}(C_i^{+-}C_i^{-+} + C_i^{-+}C_i^{+-})$	$\ln\alpha_0$	0	$-(2+\frac{1+\alpha_0}{1-\alpha_0}\ln\alpha_0)$	0
$\mathrm{tr}(C_i^{++}C_i^{++} + C_i^{--}C_i^{--})$	$-\ln\alpha_0$	$-2(1+\frac{\alpha_0}{1-\alpha_0}\ln\alpha_0)$	$-(2+\frac{1+\alpha_0}{1-\alpha_0}\ln\alpha_0)$	$2(2+\frac{1+\alpha_0}{1-\alpha_0}\ln\alpha_0)$

Table 3.2: $1/\varepsilon$ *pole terms. Shown are the quadratic expressions in* $\vec{C}$ *that are obtained from the various expectation values.*

The terms on the second line of Table 3.2 can not be written in terms of Q_0 as they break the local $U(N_{\max})\times U(N_{\max})$ gauge invariance. They sum up to zero, however, such that the final result retains the $U(N_{\max})\times U(N_{\max})$ symmetry. The result for the effective parameter σ' becomes

$$\sigma' = \sigma_0 \left(1 + \tfrac{4h_0^{2\varepsilon}}{\sigma_0 \varepsilon}\Omega_D[1 + \tfrac{\alpha_0}{1-\alpha_0}\ln\alpha_0]\right) \tag{3.2.17}$$

where $\Omega_D = \frac{1}{2}S_D/(2\pi)^D$ with S_D the surface of the unit sphere in D dimensions. Notice that (3.2.17) behaves smoothly as α_0 approaches zero (Coulomb case) whereas the individual contributions listed in Table 3.2 do not. On the other hand, (3.2.17) reduces to the well known free particle case as $\alpha_0 \to 1$. In what follows, it will be convenient to introduce the inverse conductivity t_0,

$$t_0 = 4\Omega_D/\sigma_0. \tag{3.2.18}$$

We can remove the pole term in ε from (3.2.17) by defining a renormalised theory $t_0 = \mu^{-2\varepsilon} t Z_1$ as usual, leading to a finite expression in ε. From (3.2.17) we extract

$$Z_1 = 1 + \tfrac{t}{\varepsilon}[1 + \tfrac{\alpha_0}{1-\alpha_0}\ln\alpha_0] \tag{3.2.19}$$

yielding

$$\tfrac{1}{4\Omega_D}\sigma' = t^{-1}\left(1 + t\ln(\mu^2 h_0^2)[1 + \tfrac{\alpha_0}{1-\alpha_0}\ln\alpha_0]\right). \tag{3.2.20}$$

Notice that the parameter h_0 appears as an infrared cutoff rather than the frequency or temperature, and here it plays the role of the (inverse) sample size. In the next section we give an explicit example indicating the kind of complications one is running into by working with finite temperature or frequency rather than h_0.

We now have all the necessary ingredients to calculate the beta function for the conductance,

$$\beta = \frac{\mathrm{d}t}{\mathrm{d}\ln\mu} = 2\varepsilon t - 2t^2[1 + \frac{\alpha_0}{1-\alpha_0}\ln\alpha_0]. \tag{3.2.21}$$

3.2.4 Renormalisation of z_0 and α_0

The singular contributions corresponding to the terms quadratic in J and H (3.2.9) in S_{eff} originate from many sources. They are contained in the following contractions: $\langle O_{\text{freq}}\rangle$, $\langle O_{\text{int}}^{(1),2}\rangle$, $\langle O_{\text{int}}^{(2),1}\rangle$, $\langle O_{\text{int}}^{(2),2}\rangle$, $-\frac{1}{2}\langle [O_{\text{int}}^{(1),2}]^2\rangle$ and $-\langle O_{\text{int}}^{(1),1} O_{\text{int}}^{(1),2}\rangle$, while all other possible sources give answers which are finite in ε. Each of the individual contractions gives a divergent answer as α_0 approaches zero, just as in the computation of σ'. However, numerous cancellations of these divergencies take place in the total sum. The resulting expression for $S_{\text{eff}}[T_0]$ is given by

$$\begin{aligned} &-\tfrac{\pi}{\beta} z_0 \left\langle 4O_{\text{freq}} + 2c_0 O_{\text{int}}^{(1),2} + 2c_0 O_{\text{int}}^{(2),1} + c_0 O_{\text{int}}^{(2),2} - \tfrac{\pi}{\beta} z_0 c_0^2 [O_{\text{int}}^{(1),2}]^2 - \tfrac{2\pi}{\beta} z_0 c_0^2\, O_{\text{int}}^{(1),1} O_{\text{int}}^{(1),2}\right\rangle \\ &= -\tfrac{4\pi}{\beta}\langle O_{\text{freq}}\rangle + \Gamma_0 + \sum_{\alpha\beta\gamma}\sum_{mn_1n_2} \Gamma_1(m,n_1,n_2)(T_0 \mathrm{I}_m^\alpha T_0^{-1})_{n_2n_1}^{\beta\gamma} (T_0 \mathrm{I}_{-m}^\alpha T_0^{-1})_{n_1n_2}^{\gamma\beta} \end{aligned} \tag{3.2.22}$$

where

$$\Gamma_1(m,n_1,n_2) = -\kappa z_0 c_0 \int_p D_p(n_{12})\left[1 + \kappa z_0 c_0 |m| D_p^c(|m|)\right]. \tag{3.2.23}$$

In (3.2.22) we have included a constant Γ_0 which is such that it cancels the constant that appears in the Γ_1 term. The singular contribution is obtained by replacing the square bracket in (3.2.23) by unity, and we get simply

$$\int_p D_p(n_{12}) = -\tfrac{1}{\varepsilon}\Omega_D(h_0^2+\kappa z_0 n_{12})^\varepsilon \approx -\tfrac{1}{\varepsilon}\Omega_D - \Omega_D \ln(h_0^2+\kappa z_0 n_{12}). \tag{3.2.24}$$

We mention two important aspects of this result. First, the pole term in ε is independent of the frequency indices n_1, n_2 and this implies that the result (3.2.22) becomes $U(N_{\max})\times U(N_{\max})$ invariant, i.e. it can be expressed in the quantity Q_0 (see below). Notice that $U(N_{\max})\times U(N_{\max})$ gauge invariance is automatically obtained by putting the ratio $\kappa n_{12}/h_0^2$ equal to zero, i.e. by dropping the temperature and frequency as infrared regulator in the final answer. This, however, is just a different way of saying that the background field procedure, along with any other arbitrary RG program, can never provide information on the dynamics of the problem, such as the AC conductivity.

Secondly, we evaluate the result (3.2.22) as follows. When one of the frequencies n_1, n_2 is large, the Γ_1 term yields a contribution that exactly cancels the $\langle O_{\text{freq}}\rangle$ term. From the remaining small frequencies we get

$$\begin{aligned} S_{\text{eff}}[T_0] &= \Gamma_0 - \kappa z_0 c_0 \tfrac{h_0^{2\varepsilon}}{\varepsilon}\Omega_D \sum_{\alpha\beta\gamma}\sum_{mn_1n_2}{}^{'} (T_0 \mathrm{I}_m^\alpha T_0^{-1})^{\beta\gamma}_{n_2n_1}(T_0\mathrm{I}_{-m}^\alpha T_0^{-1})^{\gamma\beta}_{n_1n_2} \\ &= \Gamma_0 + \kappa z_0 c_0 \tfrac{h_0^{2\varepsilon}}{8\varepsilon}\Omega_D \sum_{m\alpha}{}^{'} \text{tr}\,[T_0\mathrm{I}_m^\alpha T_0^{-1},\Lambda][T_0\mathrm{I}_{-m}^\alpha T_0^{-1},\Lambda] \\ &= \Gamma_0 + \kappa z_0 c_0 \tfrac{h_0^{2\varepsilon}}{8\varepsilon}\Omega_D \sum_{m\alpha}{}^{'} \text{tr}\,[\mathrm{I}_m^\alpha, Q_0][\mathrm{I}_{-m}^\alpha, Q_0] \end{aligned} \tag{3.2.25}$$

which has precisely the $\mathcal{F}$-invariant form in Q_0 given in (2.3.5). Hence, the significance of working with 'small' background fields T_0 (relative to the field variables Q) is now well recognised and the renormalisation procedure for the interacting electron gas involves not only the usual rescaling of momenta, it necessarily involves also a rescaling of the time or frequency variable or, rather, of the 'effective' bandwidth $W = TN_{\max}$.

Next, from (3.2.25) we can read off the numerical value for the constant Γ_0 which is given by

$$\Gamma_0 = \tfrac{3}{2}\kappa z_0 c_0 \tfrac{h_0^{2\varepsilon}}{\varepsilon}\Omega_D N_r \sum_m |m|. \tag{3.2.26}$$

It can be shown (section 3.4.3) that Γ_0 is precisely cancelled by the one-loop result for the free energy ($\mathcal{F}_1$) which still appears in the definition of S_{eff}. Hence the expression in Q_0 (3.2.25) is an eigenoperator of the renormalisation procedure, and this is in sharp contrast to what one is used to in the conventional σ-model theory applied to the free electron problem [60]. Operators bilinear in Q as well as higher order operators in that case describe density fluctuations which become anomalous (multifractal) as one approaches the critical point (mobility edge) in $2+2\varepsilon$ dimensions. The physics of the interacting electron gas appears to be quite different in this respect and, as will be shown in later sections, the theory is much closer to the Heisenberg ferromagnet. In particular, the interacting system is characterised by a conventional order parameter, as well as a (Coulomb) gap in the quasiparticle density of states.

We summarise the results of this section by giving the complete form for the effective background field action (including constants)

$$S_{\rm eff}[Q_0] = -\tfrac{1}{8}\sigma'{\rm Tr}\,(\nabla Q_0)^2 + \tfrac{\pi}{2\beta}z'\sum_{\alpha n}{}' {\rm Tr}\,[{\rm I}_n^\alpha, Q_0][{\rm I}_{-n}^\alpha, Q_0] - \tfrac{\pi}{\beta}z'\alpha'\sum_{n\alpha}\int_x {\rm tr}\,{\rm I}_n^\alpha Q_0\,{\rm tr}\,{\rm I}_{-n}^\alpha Q_0 \tag{3.2.27}$$

where the effective parameters σ', $c'=1-\alpha'$ and z' are given by (3.2.17) and

$$\begin{aligned} z' &= z_0(1+\tfrac{2}{\sigma_0}c_0\Omega_D\tfrac{h_0^{2\varepsilon}}{\varepsilon}) = z_0(1+\tfrac{h_0^{2\varepsilon}}{2\varepsilon}c_0 t_0) \\ \alpha' z' &= (1-c')z' = (1-c_0)z_0 = \alpha_0 z_0. \end{aligned} \tag{3.2.28}$$

Introducing renormalisation constants for the amplitudes, $z_0=\mu^{2\varepsilon}zZ_2$ and $\alpha_0=\alpha Z_\alpha$, we then extract from (3.2.28)

$$Z_\alpha = Z_2^{-1} \qquad ; \qquad Z_2 = 1-\tfrac{t}{2\varepsilon}c \tag{3.2.29}$$

such that the renormalisation of the α is obtained as

$$\frac{{\rm d}\alpha}{{\rm d}\ln\mu} = -t\alpha(1-\alpha) \tag{3.2.30}$$

and the anomalous dimension of the amplitude z is expressed in terms of the γ-function

$$\gamma_z = -\frac{{\rm d}(\mu^{2\varepsilon}z)}{{\rm d}\ln\mu} = \frac{{\rm d}\ln Z_2}{{\rm d}\ln\mu} = -t(1-\alpha). \tag{3.2.31}$$

Eq. (3.2.30) implies that $\mathcal{F}$-invariance of the theory at $\alpha_0=0$ is retained by the RG. The theory is unstable with respect to the symmetry breaking α_0 which plays a role analogous to the 'range' of the electron-electron interactions [64]. From (3.2.31) and the beta-function (3.2.21) we conclude that the Coulomb interaction problem ($\alpha_0=0$) has a fixed point $t_c \propto \varepsilon$ in $2+2\varepsilon$ dimensions which separates a metallic phase ($t<t_c$) from an insulating phase ($t>t_c$). In two spatial dimensions, the metallic phase disappears altogether and this, then, leads to the familiar complications as far as the quantum Hall effect is concerned [64].

3.3 Linear response

In the previous section we have seen that within the background field approach, the singlet interaction term contributes in a peculiar way to the conductivity renormalisation. In this section we present the results of a computation of σ' by considering the linear response to a vector potential insertion. If one views the insertion of a vector potential as the result of a ('large') background field rotation, then it is clear that this special type of background field leaves the Finkelstein action invariant, the symmetry being broken only by $\alpha_0\neq 0$ for $h_0^2=0$. In contrast to the background field procedure, we can now expect that the limit $h_0\to 0$ can be taken in the end such that the frequency and T dependence of the σ' can be computed. We stress that it is a priori completely un-obvious that the two different ways of computing σ' should yield identical RG results in the end. In fact, the different methods correspond to two completely different ways of handling the small frequency, large wavelength excitations of the system. In chapter 2 we have shown that these different methods are formally related by the $U(N_{\max})\times U(N_{\max})$ gauge invariance of the theory. Hence, the results of this section should be considered as a demonstration of the significance of the $\mathcal{F}$-invariance.

We start out by replacing the derivatives in (3.2.1) by covariant derivatives

$$\partial_j Q \to [\partial_j - i\hat{A}_j, Q]. \tag{3.3.1}$$

The effective action for the vector potential $\vec{A}$ is defined by

$$\exp S_{\text{eff}}[\vec{A}] = \int \mathcal{D}Q \ \exp S[Q, \vec{A}] \tag{3.3.2}$$

which is generally given as

$$S_{\text{eff}}[\vec{A}] = \tfrac{2\pi}{\beta} \int_x \sum_{n\alpha} \sigma'(n) \nu_n \vec{A}^\alpha_{-n} \cdot \vec{A}^\alpha_n. \tag{3.3.3}$$

The unknown quantity $\sigma'(n)$ can be expressed in terms of correlations of the Q-fields as follows

$$\begin{aligned} \sigma'(n) &= \sigma_0 - \tfrac{\sigma_0}{4n} \text{tr} \left\langle [\mathrm{I}^\alpha_n, Q][\mathrm{I}^\alpha_{-n}, Q] \right\rangle \\ &\quad + \tfrac{\sigma_0^2}{16nD} \int_{x'} \left\langle \text{tr} \ [\mathrm{I}^\alpha_n, Q] \nabla Q(x) \cdot \text{tr} \ [\mathrm{I}^\alpha_{-n}, Q] \nabla Q(x') \right\rangle \end{aligned} \tag{3.3.4}$$

where the expectation is with respect to the theory (3.2.1). Here we take a single fixed replica channel. At a classical level ($Q = \Lambda$), (3.3.4) gives us $\sigma' = \sigma_0$ as it should. The quantum fluctuations of the Q-field give rise to various contributions which are individually singular as α_0 approaches zero. The result is given by

$$\begin{aligned} \sigma'(n) &= \sigma_0 - \tfrac{16\kappa^2 c_0}{D} \tfrac{2\pi}{\beta} z_0^2 \sum_{m=1}^{\infty} m \int_p p^2 D_p(m+n) D_p(m+2n) D^c_p(m) \times \\ &\quad \times [D_p(m) + \tfrac{1}{2}\alpha_0 D^c_p(m+n)]. \end{aligned} \tag{3.3.5}$$

The sum and integral can be performed by keeping one of the regulators h_0^2, temperature or frequency ν_n fixed and putting the other two to zero. The general result involves complex expressions containing dilogs and ψ-functions. Here we limit the discussion to the most interesting case of the Coulomb interaction ($\alpha_0 = 0$). We find to lowest order in ε

$$\sigma'(T) = \sigma_0 - 4\Omega_D(-\tfrac{1}{\varepsilon} - \ln(\kappa z_0) + 2 + \gamma) \tag{3.3.6}$$

$$\sigma'(\nu_n) = \sigma_0 - 4\Omega_D(-\tfrac{1}{\varepsilon} - \ln \tfrac{z_0 \nu_n}{\sigma_0} + \tfrac{3}{2} - 2\ln 2) \tag{3.3.7}$$

$$\sigma'(h_0^2) = \sigma_0 - 4\Omega_D(-\tfrac{1}{\varepsilon} - \ln h_0^2). \tag{3.3.8}$$

Here, $\gamma \approx 0.5772\ldots$ denotes Euler's constant. As before, the numerical factor Ω_D can be absorbed in a redefinition of the parameters σ', σ_0. Notice that (3.3.8) is precisely the result we previously obtained by employing the background field procedure. A good approximation to the general case is obtained by replacing $\ln h_0^2$ in (3.3.8) by the following expression with appropriately chosen constants a and b

$$\ln h_0^2 \to \ln[h_0^2 + a\tfrac{z_0 \omega}{\sigma_0} + b\tfrac{z_0 T}{\sigma_0}]. \tag{3.3.9}$$

This expression indicates that the squares of the phase breaking lengths due to the different mechanisms of sample size, frequency and temperature add up in parallel. We have convinced ourselves of this result by performing numerical analyses.

3.4 Free Energy

3.4.1 The method

In this section we build upon the results obtained from the background field procedure. In particular the construction of the eigenoperators (which includes the constants in the quantum fluctuations $V, V^\dagger$) can be exploited in a very effective way by extracting higher loop renormalisation results from a computation of the free energy,[1] in complete analogy with what has been done for ordinary σ models [9]. However, the computation of the free energy in our case not only provides additional information on the theory in strong coupling, it also identifies a new physical quantity in the problem, namely the quasiparticle density of states entering the specific heat.

Since the apparatus of the RG can only be put to work once the singularity structure of the free theory is known, we proceed by first presenting the results of a computation of the free energy to two-loop order. These results indicate that the quantum theory retains the tree level structure of the free energy, a result which is obtained by putting $Q_0 = \Lambda$ in our background field functional (3.2.27). More specifically, write

$$S_{\text{eff}}[Q_0 = \Lambda] = -V_D N_r {\sum_{n>0}}' z' \nu_n \tag{3.4.1}$$

where the prime on the summation indicates that only small frequencies are of interest. The V_D stands for the volume of the D-dimensional system.

For reasons which have been well explained in sections 3.2 and 3.3, the background field procedure only provides information on the quantity z' in the case where the h_0^2 provides the dominant infrared regularisation. Since $\mathcal{F}$-invariant quantities like the free energy are, in principle, free of such a constraint, it is important to know whether and how the regulator h_0^2 can be put equal to zero. Most of the answer to the question is already provided by the theory in one-loop approximation which will be discussed in section 3.4.3. In section 3.4.4 we give the results of the much more complicated theory at two-loop level. The reader who is not interested in technical details might skip this section and proceed immediately to section 3.5 where the scaling implications are analysed.

3.4.2 Loop expansion; diagrams

Let us start from the renormalisable action (3.2.1). The free energy $\mathcal{F}$ is defined by

$$e^{-V_D N_r \mathcal{F}} = Z = \int \mathcal{D}Q \; e^{S[Q]}. \tag{3.4.2}$$

The nonquadratic parts (in $V, V^\dagger$) of the action are treated as a perturbation of the Gaussian theory (3.2.13). Denoting expectation values with respect to the Gaussian theory by $\langle \cdots \rangle$, the free energy is given by

$$-V_D N_r \mathcal{F} = S^{(0)} + \ln Z^{(2)} + \left\langle S^{(4)} + \tfrac{1}{2} \left[S^{(3)} \right]^2 \right\rangle \tag{3.4.3}$$

where a superscript between parentheses gives the order in $V, V^\dagger$. We write a loop expansion as follows

$$\mathcal{F} = \mathcal{F}_0 + \mathcal{F}_1 + \mathcal{F}_2 + \cdots \tag{3.4.4}$$

[1] Well, actually the grand canonical potential. We use the term 'free energy' to stress the analogy with the Q-field theory for free electrons.

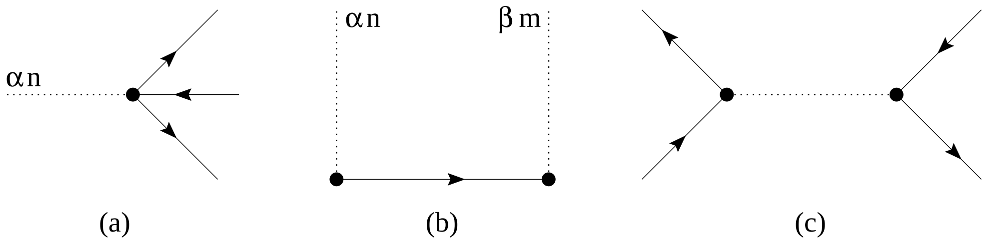

Figure 3.2: *Some examples of how diagrams are drawn.*
(a) $\mathrm{tr}\ \mathrm{I}_n^\alpha V V^\dagger V$; (b) $\langle \mathrm{tr}\ \mathrm{I}_n^\alpha V\ \mathrm{tr}\ \mathrm{I}_m^\beta V^\dagger \rangle$; (c) $\sum_{n\alpha} \mathrm{tr}\ \mathrm{I}_n^\alpha V V^\dagger\ \mathrm{tr}\ \mathrm{I}_{-n}^\alpha V V^\dagger$.

where the subscript i on $\mathcal{F}_i$ indicates the number of loops. Diagrams will be drawn in the following way (see Fig. 3.2). Every matrix trace is represented as a point. The matrices whose trace is taken are represented by lines emerging from this point. The ordering of the lines is such that the counterclockwise direction in the diagram corresponds to reading the argument of the trace from left to right. An I-matrix is drawn as a dotted line, a V-matrix as a solid line with an outgoing arrow and $V^\dagger$ as a solid line with an incoming arrow. Points can be connected by lines in two ways. A dotted line between two points means that a sum of the form $\sum_{n\alpha} \mathrm{tr}\ (\mathrm{I}_n^\alpha X_1)\mathrm{tr}\ (\mathrm{I}_{-n}^\alpha X_2)$ is taken. A solid line between two points represents a contraction of a V with a $V^\dagger$, i.e. $\mathrm{tr}\ X_1 V\ \mathrm{tr}\ V^\dagger X_2$. Points can be connected by more than one solid line. A solid line can also connect a point to itself.

At zero and one loop level, diagrams do not give much additional insight and we will not draw them. The zero loop result is simply given by $S^{(0)}$, i.e. the $Q = \Lambda$ part of (3.2.1),

$$\mathcal{F}_0 = 2z_0 {\sum_{n>0}}' \nu_n. \tag{3.4.5}$$

3.4.3 One-loop theory

The one loop result is given by the term $\ln Z^{(2)}$ in (3.4.2). Using (3.2.15) we find

$$\mathcal{F}_1 = \sum_{n>0} \int_p \ln \frac{p^2 + h_0^2 + \alpha_0 \kappa z_0 n}{p^2 + h_0^2 + \kappa z_0 n} = -\Omega_D \sum_{n>0} \frac{(h_0^2 + \kappa z_0 n)^{1+\varepsilon} - (h_0^2 + \alpha_0 \kappa z_0 n)^{1+\varepsilon}}{\varepsilon(1+\varepsilon)}. \tag{3.4.6}$$

By taking the $\ln T$-derivative of $\mathcal{F}$ one reproduces the result of the background field procedure (3.4.1),

$$\frac{\partial \mathcal{F}}{\partial \ln T} = 2 {\sum_{n>0}}' z_0 \nu_n \left\{ 1 + \frac{2\Omega_D}{\sigma_0} \cdot \frac{(h_0^2 + \kappa z_0 n)^\varepsilon - \alpha_0 (h_0^2 + \alpha_0 \kappa z_0 n)^\varepsilon}{\varepsilon} \right\}. \tag{3.4.7}$$

By neglecting the ν_n terms relative to the infrared regulator h_0^2, one obtains

$$\frac{\partial \mathcal{F}}{\partial \ln T} = 2z' {\sum_{n>0}}' \nu_n \tag{3.4.8}$$

with z' precisely given by the background field result of (3.2.28). Next we consider (3.4.7) in the limit $h_0^2 \to 0$. In this case (3.4.7) defines a frequency dependent effective parameter $z'(\nu_n)$,

$$\frac{\partial \mathcal{F}}{\partial \ln T} = 2 {\sum_{n>0}}' z'(\nu_n) \nu_n \quad ; \quad z'(\nu_n) = z_0 \left\{ 1 + \frac{2\Omega_D}{\sigma_0} \left(\frac{4 z_0 \nu_n}{\sigma_0} \right)^\varepsilon \frac{1 - \alpha_0^{1+\varepsilon}}{\varepsilon} \right\}. \tag{3.4.9}$$

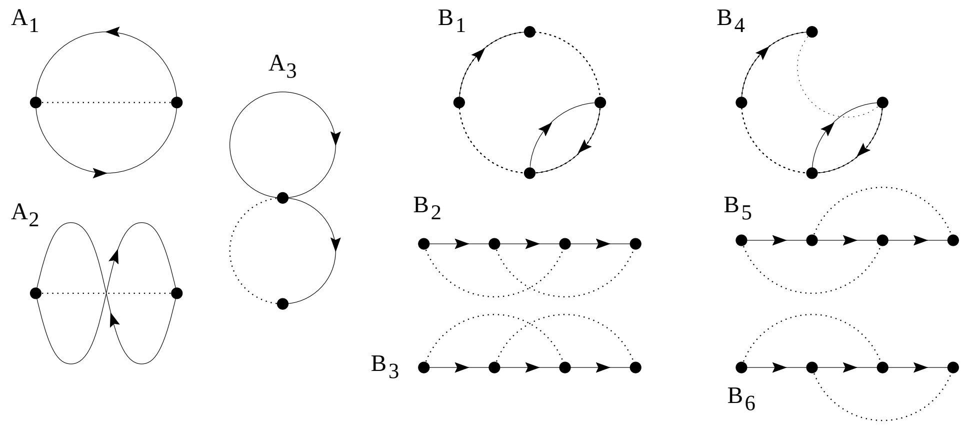

Figure 3.3: *Two-loop diagrams contributing to $\mathcal{F}_2$.*

3.4.4 Two-loop theory

The two-loop contribution to the free energy is given by the expectation value in (3.4.3). Since the kinetic part of the action does not contain third order contributions in $V, V^\dagger$ and the frequency term is only of second order, we can write

$$-V_D N_r \mathcal{F}_2 = \left\langle S_{\text{kin}}^{(4)} + S_{\text{int}}^{(4)} + \tfrac{1}{2}\left(S_{\text{int}}^{(3)}\right)^2 \right\rangle . \tag{3.4.10}$$

The fourth order part of the kinetic term is given by

$$S_{\text{kin}}^{(4)} = -\tfrac{1}{2}\sigma_0 \text{Tr}\ \left\{(V^\dagger \nabla V)^2 + (V \nabla V^\dagger)^2\right\} . \tag{3.4.11}$$

The fourth and third order part of the interaction term are given by

$$\begin{aligned} S_{\text{int}}^{(4)} &= \tfrac{2\pi}{\beta} z_0 c_0 \int_x \left\{ 2\sum_{n\alpha} \text{tr}\ \mathrm{I}_n^\alpha [V^\dagger, V]\ \text{tr}\ \mathrm{I}_{-n}^\alpha [V^\dagger, V] + \left(\sum_{\alpha, n>0} \text{tr}\ \mathrm{I}_n^\alpha V^\dagger\ \text{tr}\ \mathrm{I}_{-n}^\alpha V V^\dagger V + \text{c.c.} \right) \right\} \\ S_{\text{int}}^{(3)} &= \tfrac{8\pi}{\beta} z_0 c_0 \sum_{\alpha, n>0} \int_x \left\{ \text{tr}\ \mathrm{I}_n^\alpha V^\dagger\ \text{tr}\ \mathrm{I}_{-n}^\alpha [V^\dagger, V] + \text{tr}\ \mathrm{I}_{-n}^\alpha V\ \text{tr}\ \mathrm{I}_n^\alpha [V^\dagger, V] \right\} . \end{aligned} \tag{3.4.12}$$

The calculation of (3.4.10) is simplified by the fact that $\langle S_{\text{kin}}^{(4)} \rangle$ vanishes. The propagators only depend on the absolute values of the momenta, but the derivatives in (3.4.11) give rise to a cosine of the relative angle between two momenta; integration over the angle yields zero.

Performing the rest of the contractions in (3.4.10) is straightforward but laborious. In terms of diagrams, they are written as

$$\begin{aligned} \langle S_{\text{int}}^{(4)} \rangle &= 2\kappa\sigma_0 z_0 c_0 \{A_1 - A_2 - \tfrac{1}{2} A_3\} \\ \tfrac{1}{2}\langle [S_{\text{int}}^{(3)}]^2 \rangle &= (\kappa\sigma_0 z_0 c_0)^2 \left\{2B_1 + B_2 + B_3 - 2B_4 - B_5 - B_6\right\} , \end{aligned} \tag{3.4.13}$$

where the A_i, B_i are drawn in Fig. 3.3. The factor of 2 in front of B_1 and B_4 originates from the fact that the arrow on the single solid line in the upper left corner can be reversed without changing the value of the diagram. The following identities are very useful in doing the contractions,

$$
\begin{aligned}
\mathrm{tr}\ \mathrm{I}_n^\alpha V^\dagger V &= 0 \qquad \text{for } n \neq 0 && (3.4.14)\\
\mathrm{tr}\ \mathrm{I}_{-m}^\alpha V(q)\ \mathrm{tr}\ \mathrm{I}_n^\beta V^\dagger(-q') &= \tfrac{1}{\sigma_0}\delta(q-q')\delta_{mn}^{\alpha\beta} n D_q^c(n) && (3.4.15)\\
\mathrm{tr}\ \mathrm{I}_n^\alpha V^\dagger(-q')\ \mathrm{tr}\ V(q) M &= \tfrac{1}{\sigma_0}\delta(q-q') D_q^c(n)\ \mathrm{tr}\ \hat{\mathrm{I}}_n^\alpha M. && (3.4.16)
\end{aligned}
$$

Here the $\hat{\mathrm{I}}_n^\alpha$ is defined as in (2.2.38). In the end we get an expression for $\mathcal{F}_2$ in terms of the propagators D and D^c,

$$
\begin{aligned}
\mathcal{F}_2 \;=\; & \tfrac{1}{\sigma_0}\kappa z_0 c_0 \sum_{s>0} s \cdot \int_{p,q} D_q(s) D_p(s) + \tfrac{2}{\sigma_0}(\kappa z_0 c_0)^2 \sum_{m,n>0} \min(m,n) \times && (3.4.17)\\
& \times \int_{p,q} \left\{ -D_p(n) D_p^c(n) D_q(m+n) + D_q^c(m) D_p(n) D_p^c(n) \right.\\
& \left. -D_q^c(m) D_p^c(n) D_{p+q}(m+n) - \kappa z_0 c_0 \cdot m \cdot D_q^c(m) D_p(n) D_p^c(n) D_{p+q}(m+n) \right\}.
\end{aligned}
$$

The overall factor $\min(m,n)$ in front of all the terms with more than two propagators can be roughly understood as follows. The summation $m>0$ always enters the stage when a redefinition occurs of the summation variable n_2 on $V_{n_1 n_2}$ to a new summation variable $m = n_1 - n_2 > 0$. This shift leaves behind a constraint on the range of n_1, namely $n_1 < m$. The tracing with I_n in (3.4.12) induces a shift and a constraint of the type $n_1 < n$ (or $n_1 + n \geq m$ or something similar). In every contribution a free sum over n_1 finally occurs, yielding $\min(n,m)$ due to the constraints imposed on the summation interval.

We wish to present the result of (3.4.17) in the form (3.4.5, 3.4.6), i.e. as a frequency sum of corrections to z_0. This means we have to perform two momentum integrals and one frequency sum, while keeping one frequency fixed. This can be done in the following way:

$$
\begin{aligned}
\sum_{m,n>0} \min(m,n) \cdot f(m,n) &= \sum_{m=1}^{\infty} m \sum_{n=m+1}^{\infty} f(m,n) + \sum_{n=1}^{\infty} n \sum_{m=n+1}^{\infty} f(m,n)\\
&= \sum_{s=1}^{\infty} s \sum_{l=s+1}^{\infty} \Big\{ f(s,l) + f(l,s) \Big\}.
\end{aligned}
$$

Thus $\mathcal{F}_2$ (3.4.17) can be rewritten in the form

$$
\begin{aligned}
& \tfrac{1}{\sigma_0}\kappa z_0 c_0 \sum_{s>0} s \cdot \int_{p,q} D_q(s) D_p(s) + \tfrac{2}{\sigma_0}(\kappa z_0 c_0)^2 \sum_{s>0}\sum_{l>s} \int_{p,q} \Big\{ - D_q(s) D_q^c(s) D_p(s+l)\\
& -D_q(s+l) D_p(l) D_p^c(l) + D_q^c(s) D_p(l) D_p^c(l) + D_q(s) D_q^c(s) D_p^c(l) - 2 D_q^c(s) D_p^c(l) D_{p+q}(s+l)\\
& - \kappa z_0 c_0 l D_q(s) D_q^c(s) D_p^c(l) D_{p+q}(s+l) - \kappa z_0 c_0 s D_q^c(s) D_p(l) D_p^c(l) D_{p+q}(s+l) \Big\}. \qquad (3.4.18)
\end{aligned}
$$

Calculations are now straightforward but cumbersome. Some details are presented in appendix 3A. The final result for the poles in ε in two loop contribution is given by

$$
\begin{aligned}
\mathcal{F}_2 \;=\; & \tfrac{2}{\sigma_0}\kappa z_0 \sum_{s>0} s \cdot h_0^{4\varepsilon} \Omega_d^2 \left\{ -\frac{(1+h_s^2/h_0^2)^\varepsilon - 1}{\varepsilon \cdot h_s^2/h_0^2} \cdot \frac{2+\ln\alpha_0}{\varepsilon} - \frac{\ln(1+h_s^2/h_0^2)}{\varepsilon} \right.\\
& \left. + \frac{1}{2\varepsilon^2} + \frac{\pi^2}{2\varepsilon} + \mathcal{O}(\varepsilon^0) \right\}, \qquad (3.4.19)
\end{aligned}
$$

where we have introduced $h_s^2 = \kappa z_0 s$ and put $\alpha_0 = 0$ whenever possible. It should be mentioned that in obtaining (3.4.19) drastic cancellations of many singular terms in α_0 take place (see appendix 3A). Combining (3.4.6) and (3.4.19), one gets

$$\mathcal{F}_1 + \mathcal{F}_2 = \Omega_d \cdot \frac{1}{\varepsilon} \frac{(h'^2 + {h'_s}^2)^{1+\varepsilon} - (h'^2)^{1+\varepsilon}}{1+\varepsilon} + \frac{1}{\sigma'} \sum_{s>0} {h'_s}^2 \cdot h_0^{4\varepsilon} \Omega_d^2 \left\{ \frac{1}{\varepsilon^2} + \frac{\pi^2}{\varepsilon} \right\}, \tag{3.4.20}$$

where we make use of the one-loop renormalisation of h_0, σ_0 and z_0:

$$t_0 \quad \to \quad t' = t_0 \cdot \left\{ 1 - \frac{h_0^{2\varepsilon}}{\varepsilon} t_0 \right\} \tag{3.4.21}$$

$$z_0 \quad \to \quad z' = z_0 \cdot \left\{ 1 + \frac{h_0^{2\varepsilon}}{2\varepsilon} t_0 \right\} \tag{3.4.22}$$

$$h_0^2 \quad \to \quad h'^2 = h_0^2 \cdot \left\{ 1 - [2 + \ln \alpha_0] \frac{h_0^{2\varepsilon}}{\varepsilon} t_0 \right\}. \tag{3.4.23}$$

Eqs. (3.4.21)–(3.4.23) imply the following renormalisation of the quantity $h_s^2 \propto t_0 z_0$

$$h_s^2 \to {h'_s}^2 = h_s^2 \cdot \left\{ 1 - \frac{h_0^{2\varepsilon}}{2\varepsilon} t_0 \right\}. \tag{3.4.24}$$

Let us also present the result for the case when $h_0^2 = 0$, so that the frequency serves as an infrared regulator:

$$\mathcal{F}_0 = \frac{2\Omega_D}{t_0} {\sum_{s>0}}' h_s^2 \tag{3.4.25}$$

$$\mathcal{F}_1 = \frac{2\Omega_D}{t_0} {\sum_{s>0}}' h_s^2 \frac{h_s^{2\varepsilon}}{\varepsilon(1+\varepsilon)} \frac{t_0}{2} \tag{3.4.26}$$

$$\mathcal{F}_2 = -\frac{2\Omega_D}{t_0} {\sum_{s>0}}' h_s^2 \cdot \frac{h_s^{2\varepsilon}}{\varepsilon^2} \frac{t_0^2}{8} \left[1 - \varepsilon(\frac{\pi^2}{3} + 4) + \mathcal{O}(\varepsilon^2) \right]. \tag{3.4.27}$$

3.5 Scaling results

Our RG program of the Finkelstein theory consisted of two separate parts. First we studied the background field procedure in dimensional regularisation and we focused primarily on how the interaction term affects the ultraviolet behaviour of the theory. This procedure provides two renormalisation constants, Z_1 and Z_2. (From now on we consider the Coulomb case $\alpha_0 = 0$).

The background field procedure and other, arbitrary, RG programs [7] do not provide us, however, with the full temperature and frequency dependence of physical observables such as the conductivity. This was the subject of the second part in which we recognised that observables should be constructed out of only those correlations or 'background' fields that leave the interaction term S_{F} invariant. There are two classes of such correlations: the free energy itself and the one denoted as 'linear response', which provide us with equilibrium statistical mechanics and transport theory, respectively. For these quantities we can derive general scaling results from the RG.

First we summarise the results for the conductivity σ_0 and the free energy $\mathcal{F}_0+\mathcal{F}_1+\mathcal{F}_2$ as follows

$$\sigma'(s) = 4\Omega_D \tfrac{1}{t_0} R_\sigma(t_0,h_s) \quad ; \quad R_\sigma = 1 + \tfrac{h_s^{2\varepsilon}}{\varepsilon} t_0 \left(1 - \varepsilon[\tfrac{5}{2} - 2\ln 2]\right). \tag{3.5.1}$$

The $\ln T$-derivative of the free energy can be written as

$$\tfrac{\partial \mathcal{F}}{\partial \ln T} = 2\Omega_D \sum_{s>0} (h_s')^2/t' = 2\Omega_D \sum_{s>0} \tfrac{h_s^2}{t_0} M(t_0,h_s) \tag{3.5.2}$$

where

$$M(t_0,h_s) = 1 + \tfrac{h_s^{2\varepsilon}}{2\varepsilon} t_0 + \tfrac{h_s^{4\varepsilon}}{\varepsilon^2} t_0^2 \left(-\tfrac{1}{8} + \varepsilon \tfrac{1}{4}[1 + \tfrac{\pi^2}{6}]\right). \tag{3.5.3}$$

These results are completely analogous to what has been obtained for the Heisenberg ferromagnet [66]. The quantity M is the analogue of the magnetisation in ferromagnetic language and we will next follow up on the analysis of [66]. Equation (3.5.3) can be understood in terms of the effective parameters h_s' and t' with the latter defined by

$$\tfrac{1}{t'} = \tfrac{1}{t_0} R(t_0,h_s) \qquad ; \qquad R(t_0,h_s) = 1 + \tfrac{h_s^{2\varepsilon}}{\varepsilon} t_0 + \mathcal{O}(t_0^2). \tag{3.5.4}$$

Notice that the expressions for σ' and $1/t'$ in (3.5.1) and (3.5.4) are not necessarily the same, since they describe different physics (linear response and equilibrium statistical mechanics respectively). Nevertheless, both (3.5.1,3.5.3) and (3.5.4) can be used to extract the renormalisation constants Z_1 and Z_2 obtained by putting

$$t_0 = \mu^{-2\varepsilon} t Z_1 \qquad ; \qquad \Omega_D \tfrac{h_s^2}{t_0} = \nu_s z_0 = \nu_s \mu^{2\varepsilon} z Z_2 \tag{3.5.5}$$

where t, z now stand for the parameters of the renormalised theory. Following the scheme of minimal subtraction we get

$$\begin{aligned} Z_1 &= 1 + t/\varepsilon & (3.5.6)\\ Z_2 &= 1 - \tfrac{t}{2\varepsilon} - \tfrac{t^2}{\varepsilon^2}\left(\tfrac{1}{8} + \varepsilon[\tfrac{\pi^2}{24} + \tfrac{3}{4}]\right). & (3.5.7) \end{aligned}$$

The renormalisation β- and γ-functions are obtained as usual

$$\beta = \frac{\mathrm{d}t}{\mathrm{d}\ln\mu} = \frac{2\varepsilon t}{1 + t\,\mathrm{d}\ln Z_1/\mathrm{d}t} \qquad ; \qquad \gamma = \beta \frac{\mathrm{d}\ln Z_2}{\mathrm{d}t}, \tag{3.5.8}$$

yielding

$$\beta = 2\varepsilon t - 2t^2 \qquad ; \qquad \gamma = -t - t^2(3 + \pi^2/6). \tag{3.5.9}$$

Next we express the free energy and conductivity in terms of the renormalised parameters

$$\frac{\partial \mathcal{F}}{\partial \ln T} = 2\mu^{2\varepsilon} \sum_{s>0} z\nu_s M \qquad \tfrac{1}{t'} = \mu^{2\varepsilon} \tfrac{1}{t} R \qquad \sigma' = 4\Omega_D \mu^{2\varepsilon} \tfrac{1}{t} R_\sigma \tag{3.5.10}$$

where M, R and R_σ can be written in scaling form following the method of characteristics

$$\begin{aligned} M &= M_0(t) g(\nu_s z \xi^D M_0) \\ R &= R_0(t) h(\nu_s z \xi^D M_0) \\ R_\sigma &= R_0(t) f(\nu_s z \xi^D M_0) \end{aligned} \tag{3.5.11}$$

with M_0, R_0 and ξ determined from the RG functions β and γ according to

$$\begin{aligned}&[\mu\partial_\mu + \beta\partial_t]\xi(t) = 0\\ &[\beta\partial_t - \gamma]M_0(t) = 0\\ &[\beta\partial_t + 2\varepsilon - t^{-1}\beta]R_0(t) = 0.\end{aligned} \tag{3.5.12}$$

More explicitly, in $2+2\varepsilon$ dimensions we have for the metallic phase ($t<t_c$)

$$\begin{aligned}\xi &= \mu^{-1}t^{1/2\varepsilon}(1-t/t_c)^{-\nu}\\ R_0 &= (1-t/t_c)^{2\varepsilon\nu}\\ M_0 &= (1-t/t_c)^{\beta_0}\end{aligned} \tag{3.5.13}$$

where $t_c=\varepsilon+\mathcal{O}(\varepsilon^2)$ is the critical point $\beta(t_c)=0$, and

$$\nu = -1/\beta'(t_c) \qquad ; \qquad \beta_0 = -\nu\gamma(t_c). \tag{3.5.14}$$

It is interesting to conclude that the interacting electron gas is much closer to the physics of the Heisenberg ferromagnet than to the free electron problem which has $\beta_0=0$, indicating that the density of states is nonsingular. The metal/insulator transition for interacting electrons is characterised by the appearance of a real order parameter. We will show below (section 3.6) that $\beta_0\neq 0$ in this case means that the specific heat becomes singular as one approaches from the metallic side.

Important conceptual quantities of the theory are the various length scales (L_φ) induced by the phase breaking parameters such as temperature T, frequency ν_n and the parameter denoted as h_0. From the RG we have

$$\begin{aligned}L_\varphi^{-D}(h_0) &= (h')^2\\ L_\varphi^{-D}(T) &= TzM_0 g_T(Tz\xi^D M_0)\\ L_\varphi^{-D}(\nu_n) &= \nu_n z M_0 g_\nu(\nu_n z\xi^D M_0).\end{aligned} \tag{3.5.15}$$

Finally, we can obtain 'equations of state' for the quantities M and R. As shown in [66] to lowest order in ε, these quantities obey

$$\begin{aligned}\frac{\nu_s zt}{M^\delta} &= (t_c/t)^{1/\varepsilon}\left(1-2\varepsilon\nu\frac{1-t/t_c}{M^{1/\beta_0}}\right)^{1/\varepsilon}\\ \frac{\nu_s zt}{R^\kappa} &= (t_c/t)^{1/\varepsilon}\left(1-\frac{1-t/t_c}{R^{1/2\varepsilon\nu}}\right)^{1/\varepsilon}\end{aligned} \tag{3.5.16}$$

with the following relations between the exponents

$$D\nu = \beta_0(\delta+1) \qquad ; \qquad \kappa = \beta_0\delta/(2\varepsilon\nu). \tag{3.5.17}$$

The equation of state for the conductivity is obtained by replacing $R\to R_\sigma$ and $\nu_s\to i\omega$ in (3.5.16).

3.6 Specific heat

In order to make contact with an important quantity like the specific heat, we will first evaluate the derivative $\partial\mathcal{F}/\partial\ln T$ from the fermionic path integral for the case of free particles, and compare the results to those obtained from Q-field theory. By the rules of statistical mechanics we have

$$-\frac{\partial \ln Z}{\partial \ln \beta} = N_r \beta(\bar{E} - \mu \bar{N}) \tag{3.6.1}$$

where $\bar{E}$ is the average energy and $\bar{N}$ the number of electrons. Since the specific heat is defined as the change of $\bar{E}$ with varying T for a fixed number of particles, we still need some procedure that eliminates the explicit μ-dependence in (3.6.1).

3.6.1 Free particles

From the original fermionic path integral we obtain directly

$$V_D \tfrac{\partial \mathcal{F}}{\partial \ln \beta} = \sum_n i\omega_n \;\mathrm{tr}\; \tfrac{1}{i\omega_n + \mu - H_0} \tag{3.6.2}$$

with $\ln Z = -V_D N_r \mathcal{F}$ and 'tr' here denoting a trace over position space. By writing the frequency sum as a contour integral in the usual way,

$$V_D \tfrac{\partial \mathcal{F}}{\partial \ln \beta} = -\tfrac{\beta}{2\pi i} \oint \mathrm{d}z \; \tfrac{z}{e^{\beta z}+1} \;\mathrm{tr}\; \tfrac{1}{z+\mu-H_0}, \tag{3.6.3}$$

we obtain the more transparent expression

$$\beta V_D \int_{-\infty}^{\infty} \mathrm{d}\epsilon \; \frac{\epsilon}{e^{\beta\epsilon}+1} \rho(\epsilon) \tag{3.6.4}$$

where $\rho(\epsilon)$ is the density of states at energy $\mu+\epsilon$ and V_D the volume of the system. The equivalence between (3.6.1) and (3.6.4) is now easily established. By splitting (3.6.4) into zero-T and finite-T parts we finally obtain

$$\begin{aligned} V_D \tfrac{\partial \mathcal{F}}{\partial \ln \beta} &= \beta(f_0 + f_T) \\ &= \beta V_D \int_{-\infty}^{0} \mathrm{d}\epsilon \; \epsilon\rho(\epsilon) + \beta V_D \int_0^{\infty} \mathrm{d}\epsilon \; \frac{\epsilon}{e^{\beta\epsilon}+1} [\rho(\epsilon) + \rho(-\epsilon)] \\ &= \beta \left(-\int_{-\infty}^{\mu} \mathrm{d}E \; N(E) + V_D \int_0^{\infty} \mathrm{d}\epsilon \; \frac{\epsilon}{e^{\beta\epsilon}+1} [\rho(\epsilon) + \rho(-\epsilon)] \right). \end{aligned} \tag{3.6.5}$$

with $N(E)$ the total number of states below energy E, $V_D \rho(E) = \mathrm{d}N(E)/\mathrm{d}E$. The f_T determines the well known specific heat of the free electron gas for low temperatures ($c_v = \partial f_T / \partial T = \gamma T$). Next, we wish to redo the various steps which take us from (3.6.2) to (3.6.5), but now starting from the effective theory in Q. We obtain, instead of (3.6.2)

$$V_D \tfrac{\partial \mathcal{F}}{\partial \ln \beta} = \pi \rho_0 N_r^{-1} \;\mathrm{Tr}\; \langle \omega Q \rangle = 2\pi V_D \rho_0 \sideset{}{'}\sum_{n \geq 0} \omega_n. \tag{3.6.6}$$

By taking a simple exponential form, i.e. replacing (3.6.6) by

$$2\pi V_D \rho_0 \sum_{n \geq 0} \omega_n e^{-\omega_n \tau_0}, \tag{3.6.7}$$

and by redoing the various steps which take us from (3.6.2) to (3.6.5), it is easily seen that the low T behaviour of the specific heat is the same as was obtained before. (All the dependence on the cutoff τ_0 is absorbed in the zero-T part of (3.6.5), i.e. f_0, and this quantity is of secondary interest). Hence, the low-T specific heat of the electron gas is correctly retained by the effective Q-field formalism.

3.6.2 Coulomb interactions; quasiparticles

Next we embark on the Coulomb interaction problem. We obtain from the Finkelstein action, instead of (3.6.6),

$$V_D \frac{\partial \mathcal{F}}{\partial \ln \beta} = \frac{\pi}{2\beta} z_0 \left\langle {\sum_n}' \mathrm{Tr}\, [\mathrm{I}_n^\alpha, Q][\mathrm{I}_{-n}^\alpha, Q] \right\rangle . \tag{3.6.8}$$

The expectation on the r.h.s. has been the main subject of section 3.5. Expression (3.6.8) will be denoted by

$$-2V_D z_0 \sum_{n>0} \nu_n M(\nu_n) e^{-\nu_n \tau_0} \tag{3.6.9}$$

where, as before, we introduced an arbitrary τ_0 for convergence purposes. Notice that the quantity $M(\nu_n)$, in contrast to the density of states ρ_0 in the expression for the free electron problem (3.6.7), now acquires nontrivial ν_n-dependence. We proceed by evaluating the discrete sum in (3.6.9) as a contour integral. Repeating the same steps which lead to (3.6.5) we obtain

$$V_D \tfrac{\partial \mathcal{F}}{\partial \ln \beta} = \beta (f_0 + f_T) \tag{3.6.10}$$

where we write

$$f_0 = -V_D \frac{z_0}{\pi} \int_0^\infty \mathrm{d}\epsilon\; \epsilon M(\epsilon) e^{-\epsilon \tau_0} \quad ; \quad f_T = V_D \frac{z_0}{\pi} \int_0^\infty \mathrm{d}\epsilon\; \frac{\epsilon}{e^{\beta\epsilon} - 1} \rho_{\mathrm{qp}}(\epsilon) \tag{3.6.11}$$

with

$$\rho_{\mathrm{qp}}(\epsilon) = e^{i\epsilon\tau_0} M(-i\epsilon) + e^{-i\epsilon\tau_0} M(i\epsilon). \tag{3.6.12}$$

As before (3.6.5), the specific heat is obtained as $c_V = \partial f_T / \partial T$. The quantity $\rho_{\mathrm{qp}}(\epsilon)$ is identified as the quasiparticle density of states of the interacting system. Notice that these quasiparticles are a direct consequence of the particle-hole pair instability and, hence, they are bosonic.

The expression for f_T, (3.6.11), is the most important result of this section and we must make use of the results for $M(\nu_n)$, section 3.5, in order to allow some general conclusions about the specific heat. First, in the metallic phase in $2+2\varepsilon$ dimensions, the arbitrary cutoff factor containing τ_0 in f_T does not contribute to the leading behaviour of c_V at low T, and τ_0 can be safely put to zero. This can be seen most simply from the behaviour near the fixed points $t=0$ and $t=t_c$, in which case we have

$$\begin{array}{rlll} t \to 0 : & & & \\ M(\nu_n) \approx 1 & \rho_{\mathrm{qp}} \to 2\cos(\epsilon\tau_0) \approx 2 & c_v \to \gamma_0 T & \\ t \lesssim t_c : & & & (3.6.13) \\ M(\nu_n) \approx |\nu_n|^{1/\delta} & \rho_{\mathrm{qp}} \approx |\epsilon|^{1/\delta} 2\cos(\frac{\pi}{2\delta} + \epsilon\tau_0) & c_v \to \gamma^* T^{1+1/\delta}. & \end{array}$$

Hence, approaching the MI transition from the metallic side, the quasiparticle density of states develops a (Coulomb) gap and the exponent $1/\delta$ is of order $\varepsilon = (D-2)/2$. More generally, we have the following scaling result for the quasiparticle density of states,

$$\rho_{\mathrm{qp}}(\epsilon) = M_0 G(\epsilon z \xi^D M_0) = M(-i\epsilon) + M(i\epsilon) \tag{3.6.14}$$

which may be obtained in closed form using the results for M of section 3.5.

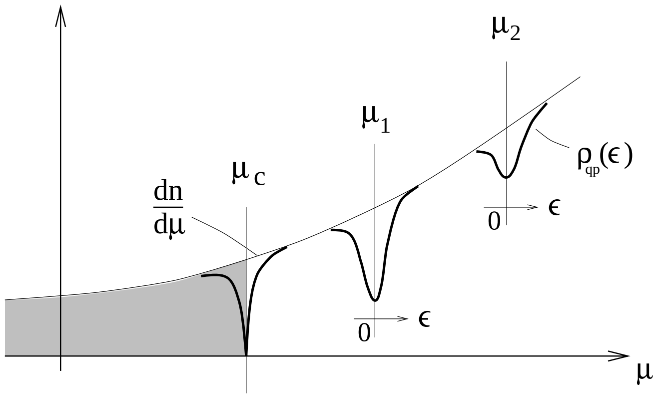

Figure 3.4: *The quasiparticle density of states. Shown is $\rho_{\rm qp}$ for three values of μ and the curve* $\mathrm{d}n/\mathrm{d}\mu$.

3.6.3 Crossing over between free electrons and quasiparticles

For completeness we give a sketch of $\rho_{\rm qp}(\epsilon)$ as compared to the thermodynamic density of states $\mathrm{d}n/\mathrm{d}\mu$, Fig. 3.4. The $\mathrm{d}n/\mathrm{d}\mu$ does not enter the effective Q-theory and is taken as a regular function of the chemical potential, since it is solely determined by the underlying theory with the P matrix field variables. Fig. 3.4 indicates that $\rho_{\rm qp}(\epsilon)$ as a whole moves along as one varies the chemical potential μ, with ϵ really standing for the thermal energy fluctuations for a given, fixed μ. The effective band width $(\Delta\epsilon)$ for quasiparticle excitation is on the order of $\hbar/\tau_0$, i.e. the characteristic energy scale below which the Coulomb effects become noticeable. It is, of course, understood that for free electrons there is no distinction between $\rho(\epsilon)$ (3.6.5) and $\mathrm{d}n/\mathrm{d}\mu$, and the quasiparticles are the electrons themselves.

Let us next come back to our earlier discussion (section 3.2.1), where we replaced the original problem with Coulomb interactions (2.3.3) by a simpler one (3.2.1) introducing the parameter c_0. We argued that since the Coulomb term $S_{\rm Y}$ is irrelevant we obtain the same results by putting $c_0 = 1$ $(\alpha_0 = 0)$ in the simpler theory. We are now in a position to appreciate the fact that the two theories really stand for entirely different physical scenarios by which the free electron and Coulomb theories are related.

As already mentioned before, the parameter α_0 in the simple theory can loosely be interpreted in terms of the 'range' of the electron-electron interactions, and the results indicate that infinite range interactions $(\alpha_0 = 0)$ and finite range interactions $(\alpha_0 > 0)$ belong to different universality classes. On the other hand, the Coulomb term $S_{\rm Y}$ in the original theory really interpolates between the action for free particles at high momenta and the Finkelstein theory which only appears in the limit of large distances (relative to the screening length). This means that the true, physical theory predicts *free electron* behaviour with Fermi-Dirac statistics at *high* T (3.6.5) and *quasiparticle* behaviour with scaling and Bose-Einstein statistics at *low* T (3.6.13). This crossover mechanism for which the Coulomb term $S_{\rm Y}$ is responsible obviously applies to other physical quantities such as $\sigma_{\rm xx}$. This mechanism is important, since it has been experimentally shown [75] to apply to the more complicated metallic phases of the quantum Hall regime as well. More specifically, the transport data taken from low mobility heterostructures follow the well known free electron behaviour at high T (4K$\leq T <$20K),

$$\sigma_{ij}(T) = \int \mathrm{d}E\, \frac{\partial f(T)}{\partial E} \sigma^0_{ij}(E). \tag{3.6.15}$$

Here $f(T)$ stands for the Fermi-Dirac distribution (see 3.6.2) and σ^0_{ij} are the mean field conductances with varying energy E. This behaviour crosses over smoothly but rapidly into a scaling behaviour in T, which was observed at lower temperatures (20mK$\leq T <$4K) only. This, then, indicates that the complete theory as given by (2.3.3) has all the ingredients necessary to describe plateau transitions.

3.6.4 Strong coupling aspects

The analogy with the Heisenberg ferromagnet naturally supposes that one can take the theory one step further by extending the ferromagnetic language to include the insulating phase as well. For example, for ordinary ferromagnets we know that, since there is no phase transition in the symmetric phase, the magnetisation $M(\nu_n)$ should become a regular and odd function of the 'external field' ν_n. From the RG it can be shown that such a condition on $M(\nu_n)$ implies that the $R(\nu_n)$ (but not R_σ!) is regular as well but it must be an even function of ν_n. This analyticity statement (Griffith analyticity) would imply in our case, however, that the quasiparticle density of states collapses, i.e. (3.6.14) vanishes to all orders in ϵ!

This oddly looking conclusion can be understood as follows. First we remark that a vanishing $\rho_{\mathrm{qp}}(\epsilon)$ is a direct consequence of working with an action where the Coulomb part S_{Y} has been replaced by a symmetry breaking term S_α and where the 'external' symmetry breaking α_0 has been put equal to zero. Notice that for Fermi level quantities like $\rho_{\mathrm{qp}}(\epsilon)$ and $\sigma(\omega)$ the role of α_0 itself is analogous to the role of the magnetic field in the Heisenberg ferromagnet. This is evident from the linear response procedure, where the perturbing vector potential can be considered as a 'generator' of $\mathcal{F}$-symmetry which is broken by the external α_0-field. Hence we should expect that the quantities ρ_{qp} and σ, which are a direct measure of the low energy excitations of the electron gas, become regular in α_0 as one enters deeply into the insulating phase. In other words, they become zero by putting α_0 equal to zero. The important conclusion that can be drawn from this discussion is that the Coulomb term S_{Y} can no longer be regarded as 'irrelevant' but, instead, it is going to completely determine the physics of the insulating phase.

Let us next come back to the subject of linear response. It is clear that for the computation of conductivities one generally has to take the limit $q \to 0$ first and then $\omega \to 0$. Notice that the abovementioned complications with with the limit $\alpha_0 \to 0$ are, in fact, foreshadowed by the response at tree level (2.4.4). For instance, working in a theory where $Y(q)$ is replaced by α_0 (as was done for RG purposes), linear response leads to the correct result provided $\alpha_0 \to 0$ in the end. As one approaches the insulating phase, however, it seems from (2.4.4) that the RG result ($\sigma \approx \alpha_0$ as $\alpha_0 \to 0$) is indeed the only way of obtaining a smooth $\alpha_0 \to 0$ limit. However, in a theory with $Y(q)$ it is more likely that $\sigma \approx \xi\omega$ in the insulating phase, which means that the response becomes local in time for system sizes much larger than the localisation length ξ.

3A Details of the two-loop calculation

Be obscure clearly.
– E. B. White

In this appendix we present the final expressions for the various terms in (3.4.18) together with some calculational details. The calculation is straightforward because the internal momenta decouple, and we just present the results (only pole terms in ε), taking the limit $\alpha_0 \to 0$ wherever possible:

$$\begin{aligned}
\int_{p,q} D_q(s) D_p(s) &\to \Omega_D^2 z_0 t_0^2 \tfrac{h_0^{4\varepsilon}}{2\varepsilon^2}(1+\widehat{\omega}_s)^{2\varepsilon}\left[\tfrac{\pi\varepsilon}{\sin\pi\varepsilon}\right]^2 \\
\kappa z_0 \sum_{l>s}\int_{p,q} D_q(s) D_q^c(s) D_p(s+l) &\to \Omega_D^2 h_0^{4\varepsilon}\cdot \tfrac{1}{\varepsilon^2(1+\varepsilon)}\left\{(1+2\widehat{\omega}_s)^{1+\varepsilon}\tfrac{(1+\widehat{\omega}_s)^\varepsilon-1}{\widehat{\omega}_s}\right\} \\
\kappa z_0 \sum_{l>s}\int_{p,q} D_q(s+l) D_p(l) D_p^c(l) &\to \Omega_D^2 h_0^{4\varepsilon}\left\{\tfrac{\mathrm{I}_3(\varepsilon)}{2\varepsilon^2}\tfrac{\alpha_0^{-\varepsilon}-1}{\varepsilon} + \tfrac{1}{\varepsilon}\int_0^1 \tfrac{\mathrm{d}x}{x}\ln(1+x\widehat{\omega}_s)\right\} \\
\kappa z_0 \sum_{l>s}\int_{p,q} D_q^c(s) D_p(l) D_p^c(l) &\to \Omega_D^2 h_0^{4\varepsilon}\left\{-\tfrac{\mathrm{I}_3(\varepsilon)}{\varepsilon^2}\ln\alpha_0\cdot\tfrac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)} + \tfrac{1}{\varepsilon}\int_0^1 \tfrac{\mathrm{d}x}{x}\ln(1+x\widehat{\omega}_s)\right\} \\
\kappa z_0 \sum_{l>s}\int_{p,q} D_q(s) D_q^c(s) D_p^c(l) &\to \Omega_D^2 h_0^{4\varepsilon}\left\{\tfrac{1}{\alpha_0\varepsilon^2}\tfrac{(1+z)^\varepsilon-(1+\alpha_0 z)^\varepsilon}{z(1-\alpha_0)(1+\varepsilon)} + \tfrac{1}{\varepsilon^2}\left[(1+\widehat{\omega}_s)^\varepsilon-1\right]\right\},
\end{aligned}$$

where we have defined $\widehat{\omega}_s = h_s^2/h_0^2$ and $\mathrm{I}_3(\varepsilon)$ stands for the combination

$$\mathrm{I}_3(\varepsilon) = \tfrac{\Gamma(1-2\varepsilon)}{[\Gamma(1-\varepsilon)]^2}\left[\tfrac{\pi\varepsilon}{\sin\pi\varepsilon}\right]^2 \approx 1 + \tfrac{\pi^2}{2}\varepsilon^2 + \mathcal{O}(\varepsilon^3).$$

The calculation of the last term on the second line in (3.4.18) is more difficult but standard. As far as we are interested only in poles in ε, we can replace the sum over l by an integral and put the lower limit to zero rather than ω_s:

$$\tfrac{2\pi}{\beta}\sum_{l>s} \to \int_0^\infty \mathrm{d}\omega_l + \mathcal{O}(\varepsilon^0).$$

Using the Feynman trick one can write

$$\begin{aligned}
\kappa z_0 \sum_{l>s}\int_{p,q} D_q^c(s) D_p^c(l) D_{p+q}(s+l) = 4\tfrac{z_0}{\sigma_0}\int_0^\infty \mathrm{d}\omega_l \int_{p,q}\int_0^1 \prod_{i=1}^3 \mathrm{d}x_i \times \qquad (3A.1)\\
\times 2\delta(1-\sum_i x_i)\left[h_0^2 + q^2 x_{12} + p^2 x_{23} + 2pq x_3 + \kappa z_0 s(\alpha_0 x_1 + x_3) + \kappa z_0 l(\alpha_0 x_2 + x_3)\right]^{-3},
\end{aligned}$$

where $x_{13} = x_1 + x_2$ and $x_{23} = x_2 + x_3$. By shifting $p \to p - q x_3/x_{23}$ we can decouple p and q in the denominator, after which we are able to perform the integrals over p, q and ω_l, with the result

$$-\tfrac{\mathrm{I}_3(\varepsilon)}{2\varepsilon}\Omega_D^2 h_0^{4\varepsilon}\int_0^1 \prod_{i=1}^3 \mathrm{d}x_i \tfrac{\delta(1-\sum_i x_i)}{[x_1x_2+x_1x_3+x_2x_3]^{1+\varepsilon}}\cdot\tfrac{[1+\widehat{\omega}_s(\alpha_0 x_1 + x_3)]^{2\varepsilon}}{(\alpha_0 x_2 + x_3)}. \qquad (3A.2)$$

Introducing new variables $x \in [0,\infty]$ and $u \in [0,1]$ as follows

$$x_1 = \tfrac{x}{x+1} \quad ; \quad x_2 = \tfrac{u}{x+1} \quad ; \quad x_3 = \tfrac{1-u}{x+1},$$

(3A.2) can be written in the form

$$-\tfrac{\mathrm{I}_3(\varepsilon)}{2\varepsilon}\Omega_D^2 h_0^{4\varepsilon}\int_0^1 \tfrac{\mathrm{d}u}{\alpha_0 u+1-u}\int_0^\infty \mathrm{d}x\,\tfrac{[1+\widehat{\omega}_s(1-u)+x(1+\alpha_0\widehat{\omega}_s)]^{2\varepsilon}}{[x+u(1-u)]^{1+\varepsilon}}. \tag{3A.3}$$

Using a well-known integral representation for the hypergeometric function, the x-integral can be performed and we get the following expression for (3A.1)

$$-\tfrac{\mathrm{I}_3(\varepsilon)}{2\varepsilon}\Omega_D^2 h_0^{4\varepsilon}\left[\mathrm{I}_1-\tfrac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)}\cdot\mathrm{I}_2\right], \tag{3A.4}$$

where

$$\begin{aligned}
\mathrm{I}_1 &= \int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{\alpha_0 u+1-u}\,[1+\widehat{\omega}_s(1-u)]^{2\varepsilon}\;\;{}_2\mathrm{F}_1(1,-2\varepsilon;1-\varepsilon;\tfrac{u(1-u)(1+\alpha_0\widehat{\omega}_s)}{1+\widehat{\omega}_s(1-u)}) \qquad (3A.5)\\
\mathrm{I}_2 &= 2\int_0^1 \tfrac{\mathrm{d}u}{\alpha_0 u+1-u}\left\{1+(1-u)\,[\widehat{\omega}_s-u(1+\alpha_0\widehat{\omega}_s)]\,\right\}^{\varepsilon}. \qquad (3A.6)
\end{aligned}$$

We are interested in the limit $\alpha_0\to 0$, which corresponds to the case of real electrons. But if one puts $\alpha_0=0$ in the above equations then (3A.6) becomes divergent on the upper limit. The integral (3A.5) will be proportional to ε^{-1} making the whole contribution (3A.1) proportional to ε^{-3} and causing the RG equation to become singular. For this reason we first rewrite (3A.5) in the following form

$$\begin{aligned}
\mathrm{I}_1 &= \int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{\alpha_0 u+1-u}+\int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{\alpha_0 u+1-u}\left\{[1+\widehat{\omega}_s(1-u)]^{2\varepsilon}-1\right\} \qquad (3A.7)\\
&\quad+\int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{\alpha_0 u+1-u}\,[1+\widehat{\omega}_s(1-u)]^{2\varepsilon}\left\{{}_2\mathrm{F}_1(1,-2\varepsilon;1-\varepsilon;\tfrac{u(1-u)(1+\alpha_0\widehat{\omega}_s)}{1+\widehat{\omega}_s(1-u)})-1\right\}.
\end{aligned}$$

The first integral in (3A.7) can be done exactly while in the other two we can safely put $\alpha_0=0$, because their integrands behave smoothly when $u\to 1$ even for $\alpha_0=0$. Thus I_1 can be written in the form

$$\begin{aligned}
\mathrm{I}_1 &= \tfrac{1}{\varepsilon}\left\{\alpha_0^{-\varepsilon}\tfrac{\pi\varepsilon}{\sin\pi\varepsilon}-\tfrac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)}\right\}+\mathcal{O}(\alpha_0^{1-\varepsilon})+\int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{1-u}\left\{[1+\widehat{\omega}_s(1-u)]^{2\varepsilon}-1\right\}\\
&\quad+\int_0^1 \mathrm{d}u\tfrac{[u(1-u)]^{-\varepsilon}}{1-u}\,[1+\widehat{\omega}_s(1-u)]^{2\varepsilon}\;\left\{{}_2\mathrm{F}_1(1,-2\varepsilon;1-\varepsilon;\tfrac{u(1-u)}{1+\widehat{\omega}_s(1-u)})-1\right\}.
\end{aligned}$$

Making an ε-expansion and then using the following asymptotic formula for the hypergeometric ${}_2\mathrm{F}_1$ function, which follows directly from its definition,

$${}_2\mathrm{F}_1(1,-2\varepsilon;1-\varepsilon;z)=1+2\varepsilon\ln(1-z)+\mathcal{O}(\varepsilon^2)$$

we get the final answer for I_1

$$\mathrm{I}_1=\tfrac{1}{\varepsilon}\left\{\alpha_0^{-\varepsilon}\tfrac{\pi\varepsilon}{\sin\pi\varepsilon}-\tfrac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)}\right\}+2\varepsilon\int_0^1 \tfrac{\mathrm{d}u}{1-u}\ln\left[1+(1-u)(\widehat{\omega}_s-u)\right]+\mathcal{O}(\varepsilon^2). \tag{3A.8}$$

The integral I_2 (3A.6) can be treated in the same way

$$\begin{aligned}
\mathrm{I}_2 &= 2\int_0^1 \tfrac{\mathrm{d}u}{\alpha_0 u+1-u}+2\int_0^1 \tfrac{\mathrm{d}u}{1-u}\left\{[1+(1-u)\,(\widehat{\omega}_s-u)]^{\varepsilon}-1\right\} \qquad (3A.9)\\
&= -2\ln\alpha_0+\mathcal{O}(\alpha_0\ln\alpha_0)+2\varepsilon\int_0^1 \tfrac{\mathrm{d}u}{1-u}\ln\left[1+(1-u)(\widehat{\omega}_s-u)\right]+\mathcal{O}(\varepsilon^2).
\end{aligned}$$

Substitution of (3A.8) and (3A.9) into (3A.3) gives the final answer for the pole terms in (3A.1)

$$\kappa z_0 \sum_{l>s} \int_{p,q} D_q^c(s) D_p^c(l) D_{p+q}(s+l) = \Omega_D^2 h_0^{4\varepsilon} \cdot \frac{\mathrm{I}_3(\varepsilon)}{2\varepsilon} \left\{ -\frac{2\ln\alpha_0}{\varepsilon} \frac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)} - \frac{1}{\varepsilon} \frac{\alpha_0^{-\varepsilon}-1}{\varepsilon} \cdot \frac{\pi\varepsilon}{\sin\pi\varepsilon} + \frac{\pi^2}{3} \right\}.$$

The calculation of the first term on the last line of (3.4.18) is ideologically the same but more lengthy. We will not present it here but give only some comments. In this case the second power of the combination $\alpha_0 u + 1 - u$ appears in the denominator at the stage (3A.4). It is useful then to perform integration by parts, reducing the power by one. The rest is similar to the calculations of (3A.1). The result is

$$-\kappa^2 z_0^2 c_0 \sum_{l>s} l \cdot \int_{p,q} D_p(s) D_p^c(s) D_q^c(l) D_{p+q}(s+l) \rightarrow$$
$$\Omega_D^2 h_0^{4\varepsilon} \frac{\mathrm{I}_3(\varepsilon)}{2\varepsilon^2} \left\{ -\frac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)} \cdot \frac{(1+z)^\varepsilon - (1+\alpha_0 z)^\varepsilon}{z(1-\alpha_0)(1+\varepsilon)} \frac{\alpha_0^{-\varepsilon}-1}{\varepsilon\alpha_0} \cdot \frac{2\pi\varepsilon}{\sin\pi\varepsilon} - 2\ln\alpha_0 \frac{[\Gamma(1-\varepsilon)]^2}{\Gamma(1-2\varepsilon)} \left[1 + \frac{(1+z)^\varepsilon - 1}{z} \right] \right\}.$$

The last contribution in (3.4.18) is finite and of no interest.

Chapter 4

The integer edge

4.1 Introduction

In this chapter we investigate edge excitations in the iqHe regime. We begin in section 4.2 by reviewing and extending the instanton vacuum approach, following the ordinary replica formalism for free electrons. We show that *massless* edge excitations are obtained from the *fluctuations about integer quantised topological charge*. We find a direct relationship between the instanton parameter σ^0_{xy} and the phenomenon of inter-channel scattering at the edge. Here the number of edge channels equals the number of fully occupied Landau levels, and the inter-channel scattering is the result of a random short-ranged potential. In spite of the free boundary conditions on Q, it is still possible to formulate a topological principle for the quantisation of σ_{xy}. We use a background field calculation to show that spherical boundary conditions are dynamically generated.

In section 4.3 we embark on the full interaction problem. We first consider the case of a bulk d.o.s. gap and derive a complete chiral edge theory for the Q-field, the plasmon field and A_μ. Then, using results from chapter 2, we extend the result to a theory for both bulk and edge and integrate out the plasmon field. As two limiting cases for zero edge d.o.s. and bulk d.o.s. we obtain effective actions $S_{\mathrm{eff}}[Q, A_\mu]$ in the bulk and at the edge, respectively.

In section 4.4 we show that the effective action $S_{\mathrm{eff}}[Q, A_\mu]$ at the edge is equivalent to a theory of *chiral edge bosons*. This demonstrates the connection between the instanton vacuum on the one hand and the pure Chern-Simons theory and chiral edge bosons on the other. Our chiral boson action represents the *complete* Luttinger liquid theory, in the sense that it includes Coulomb interactions and external electromagnetic fields. We use this theory to define the Hall conductance in a general, unambiguous way by expressing the appearance of a 'chiral anomaly' [5] in terms of Laughlin's gauge argument [57].

In the derivations presented in sections 4.3 and 4.4 we make use of $\mathcal{F}$-invariance and the $\mathcal{F}$-algebra in an essential way. This shows that $\mathcal{F}$-invariance, which we put to the test in the weak coupling regime in chapter 3, retains its significance all the way to strong coupling.

In all the discussions so far, only white noise random potentials have been considered. This was always done for technical reasons alone. However, it is well known that in real quantum Hall devices slowly varying potentials are often present [57, 67]. These have always been difficult to handle. By defining our edge theory on spatially separated, independent channels which represent the equipotential contours, we obtain a description that is able to treat the combination of interactions and smooth disorder.

In sections 4.5 and 4.6 we solve two problems where this combination gives rise to unexpected results. The first problem we address is that of electron tunnelling into the quantum Hall edge. We show that the Coulomb interactions between the edge and the 'localised' bulk orbits dramatically change the predictions of theories which are based on isolated edges alone. Tunnelling processes into the quantum Hall edge have, in fact, nothing to do with the quantisation of the Hall conductance or the 'incompressibility' statement which characterise the non-equilibrium properties of the electron gas. We find that the tunnelling d.o.s. near the edge can be understood in terms of an effective edge theory which describes the equilibrium properties of the combined edge and bulk degrees of freedom. The Luttinger liquid parameter g is related to the filling fraction ν of the bulk. This leads to a tunnelling exponent which varies like $1/\nu$, in agreement with recent experimental data on the tunnelling current, taken from samples in the fractional quantum Hall regime [27]. This situation is dramatically different from what is expected by assuming an isolated edge, or in the case of short-ranged disorder which gives rise to scattering between different edge states [36, 37].

In section 4.6 we embark on the problem of plateau transitions in the presence of smooth disorder. This we model as a percolating network of 'edge states' (equipotential contours) and widely separated 'saddlepoints'. The system is then 'mapped' onto the nonlinear σ model and the main problem is to identify the length and energy scales of the 'bare' parameters or the mean field conductances which together determine the renormalisation starting point, i.e. the point where scaling occurs first. This starting point can involve, in principle, arbitrarily large distances and arbitrarily small energies, which obviously complicates the observability of the critical behaviour of the plateau transitions. We argue that Coulomb interaction effects lead to a modified mean field theory of transport. This, then, might conceivably be the explanation for the empirical fits of recent transport data [71, 4]. In section 4.6.5 the chiral boson theory is used to actually compute the inelastic relaxation rate of the conducting electrons in the saddlepoint network.

4.2 Edge excitations

4.2.1 Sigma model

Let us recall the instanton vacuum theory [58, 59, 63] for the iqHe, which is expressed in terms of the local field variables $Q^{\alpha\beta}_{pp'}$, where $\alpha, \beta = 1, \ldots, N_r$ are the replica indices and $p, p' = \pm 1$ are the indices denoting advanced/retarded components. They can be represented as

$$Q = T^{-1}\Lambda T \qquad \text{with} \qquad \Lambda^{\alpha\beta}_{pp'} = \delta^{\alpha\beta}\delta_{pp'}\mathrm{sgn}(p) \tag{4.2.1}$$

and T a unitary matrix of size $2N_r \times 2N_r$. The action is given by

$$S[Q] = -\tfrac{1}{8}\sigma^0_{\mathrm{xx}} \int \mathrm{d}^2x \ \mathrm{tr}\ (\nabla Q)^2 + \tfrac{1}{8}\sigma^0_{\mathrm{xy}} \int \mathrm{d}^2x \ \mathrm{tr}\ \varepsilon_{ij} Q\partial_i Q \partial_j Q + \pi\rho\eta \int \mathrm{d}^2x \ \mathrm{tr}\ \Lambda Q. \tag{4.2.2}$$

Here σ^0_{ij} stands for the mean field conductances in units of e^2/h (see Fig. 1.2), ρ is the (exact) density of states at the Fermi energy and η is the frequency. The second term in (4.2.2), proportional to σ^0_{xy}, has remained one of the most difficult chapters in the theory of Anderson localisation in low dimensions. Most of the insight into the problem has come from weak coupling renormalisation theory, both perturbative and non-perturbative [58, 59]. In particular we mention the global scaling diagram of the conductances (Fig. 1.2) as well as the appearance of a critical fixed point in strong coupling regime [57]. This fixed point theory predicts a massless (metallic) phase at the Landau band center as well as the following scaling result for the conductances [62]

$$\sigma_{ij}(L,B) = g_{ij}([L/\xi]^{1/\nu}) \qquad ; \qquad \xi = |B - B^*|^{-\nu}. \tag{4.2.3}$$

Here, the function g_{ij} is a regular function of its argument, B^* is the critical magnetic field strength and ν stand for the critical index for the localisation length ξ. Following the experimental tests of (4.2.3) by H.P. Wei *et al.* [76], extensive numerical work on the free electron gas has been performed and the quoted best value for the critical index is $\nu = 2.3$ [33].

To date, no exact (conformal) scheme for the critical indices exists. All that one can say at this time is that the field of exactly solvable models is not sufficiently developed to be able to handle the subtleties of topology and replica field theory. These subtleties are all well understood in the framework of weak coupling expansion techniques [58, 59] and the results were used to predict the entire singularity structure of the theory, notably (4.2.3).

It was shown in [64] that the theories of free and interacting electrons share the same basic features such as asymptotic freedom, instantons etc. The same scaling diagram for the conductances was obtained, which means that (4.2.3) remains valid also when the Coulomb interactions are taken into account. This result was conjectured but not well understood at the time of the original experiments on criticality.

Strong coupling

In this section we address the subtleties of replica field theory in the exactly solvable limit where ρ and $\sigma^0_{\rm xx}$ vanish while the Hall conductance is integer quantised ($\sigma_{\rm xy} = m$). Physically this happens when the Fermi energy is located in a d.o.s. gap between adjacent Landau bands. In this *strong coupling* limit, massless excitations exist at the edges of the system. Since several, basic aspects of the problem have previously gone unnoticed we will proceed first within the free electron formalism (4.2.2). We come back to the fermionic path integral in section 4.3.

For m completely filled Landau levels the action becomes simply

$$S[Q] = \tfrac{m}{8}\int \mathrm{d}^2x \ \mathrm{tr}\ \varepsilon_{ij} Q\partial_i Q\partial_j Q = \tfrac{m}{2}\oint \mathrm{d}\vec{x}\cdot \mathrm{tr}\ (\Lambda T\nabla T^{-1}) \tag{4.2.4}$$

where the surface integral is taken over the sample's edge. Recall that (4.2.4) is quantised according to

$$S[Q] = 2\pi i m \cdot q[Q] \tag{4.2.5}$$

with q the integer topological charge, provided that the T-matrix reduces to a $U(N_r)\times U(N_r)$ gauge at the edge [63].

Under these circumstances the sample edge has been contracted to a single point (spherical boundary conditions) and (4.2.5) is a realisation of the homotopy theory result $\pi_2[U(2N_r)/U(N_r)\times U(N_r)] = \mathbb{Z}$ which states that the mapping of Q onto the 2D plane is described by a set of integers q. It is natural to take the theory one step further and propose the quantisation of the charge $q[Q]$ as the topological principle which forces the

Hall conductance itself to be integer quantised. The idea has indeed led to a consistent quantum theory of conductances that unifies a fundamental aspect of asymptotically free field theory (i.e. dynamic mass generation) with the quantum Hall effect [58, 59]. It says that the conductances in (4.2.3) always scale toward $\sigma_{\mathrm{xx}}=0, \sigma_{\mathrm{xy}}=m$ for L large enough.

One can show [63] that the $U(N_r)\times U(N_r)$ gauge condition at the edge is the replica field theory version of a static $U(1)$ gauge acting on the physical edge states. Such a $U(1)$ gauge implies that an integer number of edge levels has crossed the Fermi level. This level-crossing is necessarily induced by the averaging procedure over random potentials, because different disorder configurations at fixed μ lead to different numbers of electrons in the sample.

Nevertheless, the topological invariant in (4.2.2), as it was discovered originally in a microscopic derivation, is truly defined with *free* boundary conditions and without any separation between edge and bulk degrees of freedom [58, 59].

4.2.2 Inter-channel edge scattering

In what follows, we show that the fluctuations about precisely quantised values for the topological charge represent the dynamics of massless edge excitations. In order to see this, we write T as the product of a $U(N_r)\times U(N_r)$ gauge U and a small fluctuation t,

$$T = Ut. \tag{4.2.6}$$

The action now becomes

$$S[Q] = 2\pi i m \cdot q[U] + \tfrac{m}{2} \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\ (\Lambda t \nabla t^{-1}) + \pi \rho_{\mathrm{e}} \eta \oint \mathrm{d}x\ \mathrm{tr}\ \Lambda Q \tag{4.2.7}$$

where we have written ρ_{e} for the density of edge states. One way of identifying (4.2.7) as the effective theory of disordered chiral edge states is to redo the derivation of the Q-field theory, but now for the 1D system with Hamiltonian

$$\mathcal{H}_{\mathrm{edge}} = -iv_{\mathrm{d}}\partial_x + V(x) \tag{4.2.8}$$

where v_{d} is the drift velocity of the edge electrons and $V(x)$ the random potential. It turns out that our initial guess (4.2.8) is correct only in the case $m=1$. This problem is easily resolved once one realizes that m stands for the number of filled Landau levels and that (4.2.8) should be replaced by a Hamiltonian for a total of m edge channels. Hence, an obvious second guess is

$$\mathcal{H}_{\mathrm{edge}} = \sum_{j=1}^{m} \mathcal{H}_{\mathrm{edge}}^{(j)} \tag{4.2.9}$$

where $\mathcal{H}_{\mathrm{edge}}^{(j)}$ is the same for all j, i.e. each of the m eigenstates experiences the same white noise potential $V(x)$, just as it appears in the original problem in two spatial dimensions. This, however, is not correct and the theory with general m, (4.2.7), necessarily requires inter-channel scattering to take place. We have to start from a matrix Hamiltonian

$$\mathcal{H}_{\mathrm{edge}}^{jj'} = -iv_{\mathrm{d}}\delta_{jj'}\partial_x + V_{jj'}(x) \tag{4.2.10}$$

where V is a Hermitian matrix. The matrix elements $V_{jj'}$ connect the edge channels j, j' and are distributed with a weight

$$P[V] \propto \exp\{-\tfrac{1}{g} \oint \mathrm{d}x\ \mathrm{tr}\ V^2\}. \tag{4.2.11}$$

One can construct a generating function for the free particle Green's functions as usual, according to

$$Z = \int \mathcal{D}\bar{\psi}\psi \int \mathcal{D}V P[V] \exp \sum_{p,\alpha,jj'} \oint \mathrm{d}x\ \bar{\psi}_p^{\alpha,j} \left[(\mu + ip\eta)\delta_{jj'} - \mathcal{H}_{\mathrm{edge}}^{jj'} \right] \psi_p^{\alpha,j'}. \tag{4.2.12}$$

In appendix 4C we show that (4.2.12) and (4.2.7) describe the same physics in the limit of large distances.

4.2.3 Criticality at the edge

We next point out that the results of the previous section provide an exact solution to our topological edge theory (4.2.4–4.2.7) for all values of N_r. The simple but important observation to be made is that the random potential $V_{jj'}(x)$ in (4.2.12) can be 'gauged away', i.e. absorbed in a redefinition of the fermion fields, and all that remains is the trivial theory of 'pure' chiral edge states,

$$Z = \int \mathcal{D}\bar{\psi}\psi\ \exp \sum_{p=\pm,\alpha,j} \oint \mathrm{d}x\ \bar{\psi}_p^{\alpha,j} [-iv_{\mathrm{d}}\partial_x + i\eta p] \psi_p^{\alpha,j}. \tag{4.2.13}$$

Eq. (4.2.13) is just a formal way of saying that edge electrons do not Anderson localise, because chirality excludes backscattering processes on random impurities. However, the conclusions that are simply drawn from edge electrons (4.2.12, 4.2.13) become highly nontrivial statements when formulated in the language of Q-field theory. In this section we will use the simplicity of the former theory in order to derive explicit results for the latter. Write

$$\begin{aligned} G_{\pm}^{jj'}(x,x') &= \langle x,j|(-\mathcal{H}_{\mathrm{edge}} \pm i\eta)^{-1}|x',j'\rangle & (4.2.14)\\ G_{\pm}(x,x') &= \langle x|(iv_{\mathrm{d}}\partial_x \pm i\eta)^{-1}|x'\rangle & (4.2.15) \end{aligned}$$

to represent the single particle propagator of the dirty edge (4.2.12) and clean edge (4.2.13) respectively. Some useful identities and relations are obtained as follows,

$$\rho_{\mathrm{e}} = \tfrac{1}{2\pi i} \sum_j \left[G_-^{jj}(x,x) - G_+^{jj}(x,x) \right] = \tfrac{m}{2\pi i} \left[G_-(x,x) - G_+(x,x) \right]. \tag{4.2.16}$$

Here ρ_{e} denotes the density of edge states at the Fermi level which can be obtained explicitly from the r.h.s.,

$$\rho_{\mathrm{e}}(x) = \rho_{\mathrm{e}} = m/2\pi v_{\mathrm{d}}. \tag{4.2.17}$$

Eq. (4.2.17) shows that the density of edge electrons is a constant, independent of x and disorder, as it should be. An important conclusion now follows for the theory of Q-fields (4.2.7), namely

$$\langle Q \rangle = \Lambda \tag{4.2.18}$$

(where the expectation is with respect to (4.2.7)), which holds for arbitrary N_r. This result may be obtained e.g. by differentiating both theories (4.2.7) and (4.2.12, 4.2.13) with respect to η. Notice that (4.2.18) can be regarded as the 'order parameter' (analogous to the magnetisation in the language of the Heisenberg ferromagnet) and one would naively expect this quantity to vanish in one spatial dimension. The result $\langle Q \rangle = \Lambda$ indicates, however, that the continuous symmetry is permanently broken at the edge of the instanton vacuum for all numbers of field components N_r. This apparent violation of the Mermin-Wagner-Coleman theorem is clearly due to the lack of positive definite Boltzmann weights in our problem that is described by an imaginary action (4.2.7). Eq. (4.2.18) also indicates that the edge of the topological vacuum is critical. The simplest way of demonstrating this is by employing the background field method. For example, the replacement $t \to t \cdot t_0$ in the second term of (4.2.7) can be written as

$$\begin{aligned}\oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[\Lambda t \nabla t^{-1}] \;\; &\to \;\; \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[\Lambda t t_0 \nabla (t_0^{-1} t^{-1})] \\ &= \;\; \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[\Lambda t \nabla t^{-1}] + \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[Q t_0 \nabla t_0^{-1}]. \qquad (4.2.19)\end{aligned}$$

Here, t_0 represents a fixed and slowly varying background field. We obtain an effective action for t_0 as follows,

$$S_{\mathrm{eff}}[t_0] = \tfrac{m}{2} \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[t_0 \nabla t_0^{-1} \langle Q \rangle] = \tfrac{m'}{2} \oint \mathrm{d}\vec{x} \cdot \mathrm{tr}\,[\Lambda t_0 \nabla t_0^{-1}]. \qquad (4.2.20)$$

Eq. (4.2.20) defines an 'effective' parameter $m' = m \mathrm{tr}\; \Lambda \langle Q \rangle / 2N_r$ which can be identified as the 'Hall conductance' and which provides information on the renormalisation of the theory at large distances [66]. Apparently we have $m' = m$. The same conclusion can be drawn for the η parameter (i.e. $\eta' = \eta \mathrm{tr}\; \Lambda \langle Q \rangle / 2N_r = \eta$) and hence we are dealing with a critical fixed point theory! The full significance of this result will become clear in section 4.4 where we make contact with the theory of chiral edge bosons.

For the remainder of this section we will elaborate on several other identities that will be used later on. The most important pair correlation of the Q-fields can be obtained as follows,

$$\begin{aligned}N(x, x') \;\; &= \;\; \pi^2 \rho_{\mathrm{e}}^2 \left\langle Q_{+-}^{\alpha\beta}(x) Q_{-+}^{\beta\alpha}(x') \right\rangle = \sum_{jj'} G_{-}^{jj'}(x, x') G_{+}^{j'j}(x', x) \\ &= \;\; m G_{-}(x, x') G_{+}(x', x). \qquad (4.2.21)\end{aligned}$$

Here, the α, β are fixed but arbitrary replica channels and

$$G_{-}(x, x') G_{+}(x', x) = \frac{i}{v_{\mathrm{d}}} \int \frac{\mathrm{d}k}{2\pi} \cdot \frac{e^{ik(x'-x)}}{v_{\mathrm{d}} k + 2i\eta} = \frac{1}{v_{\mathrm{d}}^2} \theta(x' - x) \exp\left[-\frac{2\eta}{v_{\mathrm{d}}} (x' - x) \right]. \qquad (4.2.22)$$

The step function θ shows that a chiral electron, being created at position x and drifting in the positive direction, can only be destroyed at a 'later' position $x' > x$. Notice that we have the sum rule

$$\int \mathrm{d}x'\; N(x, x') = \pi \rho_{\mathrm{e}} / \eta. \qquad (4.2.23)$$

The other pair correlations of the Q-fields vanish identically. In particular, it is straightforward to show that

$$\left\langle Q_{pp}^{\alpha\beta}(x) Q_{p'p'}^{\gamma\delta}(x') \right\rangle_{\mathrm{cum}} = 0 \qquad (4.2.24)$$

for all $p, p' = \pm$ and all replica channels $\alpha, \beta, \gamma, \delta$. Finally we wish to clarify the significance of several Q-field operators that have appeared in different contexts before. First, there are the higher order corrections to the theory of (4.2.7). In appendix 4C we derive higher dimensional operators of the type

$$\mathrm{tr}\,[\tfrac{m}{2} \partial_x + \pi \eta \rho_{\mathrm{e}} \Lambda, Q]^2. \qquad (4.2.25)$$

Secondly, we mention the bilinear combinations of the form

$$A_1 \mathrm{tr}\ \Lambda Q\ \mathrm{tr}\ \Lambda Q + A_2 \mathrm{tr}\ [\Lambda, Q][\Lambda, Q] \tag{4.2.26}$$

which are known to describe the anomalous fluctuations in the density at the quantum Hall transitions, as well as in the localisation problem in $2+\varepsilon$ dimensions [60]. We have already seen, however, that the density of chiral electrons does not fluctuate and we therefore expect (4.2.26) to be irrelevant. A classification of these operators follows from the classical equations of motion of the topological action (4.2.7) which can be written as

$$[\tfrac{m}{2}\partial_x + \pi\eta\rho_{\mathrm{e}}\Lambda, Q] = 0. \tag{4.2.27}$$

This immediately implies that the higher order generators (4.2.25) are, in fact, *redundant* (i.e. they can be 'gauged' away by a suitable transformation of Q). Next, from the identity

$$\int^x \mathrm{d}x'\ \mathrm{tr}\ \Lambda Q(x')\ \mathrm{tr}\ [\tfrac{m}{2}\partial_{x'} + \pi\eta\rho_{\mathrm{e}}\Lambda, Q(x')]\Lambda = 0 \tag{4.2.28}$$

it directly follows that the first term in (4.2.26) is also redundant. Finally, from (4.2.27) one obtains

$$\int^x dx'\ \mathrm{tr}\ [\Lambda, Q(x')] \left[\tfrac{m}{2}\partial_{x'} + \pi\eta\rho_{\mathrm{edge}}\Lambda, [\Lambda, Q(x')]\right] = 0, \tag{4.2.29}$$

and it is readily seen that the second operator in (4.2.26) is redundant as well.

4.2.4 Topological principle

Next there are the difficulties in establishing a *topological principle* for Hall quantisation, as mentioned in section 4.2.1. We have to see whether the instanton vacuum approach is free of ambiguities. For this purpose we will follow up on the background field method. We write

$$\exp S_{\mathrm{eff}}[t_0] := \int \mathcal{D}Q\ \exp\left(S_0[t_0^{-1}Qt_0] + \pi\rho\eta \mathrm{Tr}\ \Lambda Q\right) \tag{4.2.30}$$

where

$$S_0[Q] = -\tfrac{1}{8}\sigma^0_{\mathrm{xx}}\mathrm{Tr}\ (\nabla Q)^2 + \tfrac{1}{8}\sigma^0_{\mathrm{xy}}\mathrm{Tr}\ \varepsilon_{ij}Q\partial_i Q\partial_j Q. \tag{4.2.31}$$

Eq. (4.2.30) defines an effective action for the fixed t_0 matrix field. By making use of the local $U(N_r)\times U(N_r)$ gauge invariance, one can show that S_{eff} is of the same form as S_0, i.e.

$$S_{\mathrm{eff}}[t_0] = -\tfrac{1}{8}\sigma_{\mathrm{xx}}\mathrm{Tr}\ (\nabla Q_0)^2 + \tfrac{1}{8}\sigma_{\mathrm{xy}}\mathrm{Tr}\ \varepsilon_{ij}Q_0\partial_i Q_0\partial_j Q_0 \tag{4.2.32}$$

with $Q_0 = t_0^{-1}\Lambda t_0$. The main problem next is to obtain explicit knowledge of the 'effective' parameters σ_{ij} in (4.2.32), which now represent the (exact) Kubo expressions for the conductances. As long as one works with spherical boundary conditions on the Q-field, the quantisation of the Hall conductance is readily established. All that one needs is in fact that the theory develops a mass gap in the limit of large distances. Under these circumstances the insertion of slowly varying, spherically symmetric background fields t_0 should be immaterial (as $\eta \to 0$) and this then necessarily implies that $\sigma_{\mathrm{xx}} = 0$ and σ_{xy} =integer. For example, the explicit renormalisation group flows obtained from instanton calculations always indicate that the quantum Hall conditions $\sigma_{\mathrm{xx}} = 0$, σ_{xy} =integer are the stable (massive) fixed points of the theory for arbitrary number of field components N_r.

It is well known, however, that spherical boundary conditions are justified in weak coupling only, due to finite action requirements. Since there is no compelling argument which says that spherical boundary conditions should be retained all the way down to the strong coupling regime, one is clearly left with ambiguities in the instanton vacuum approach, at least as far as the quantisation of the Hall conductance is concerned.

Armed with the insight gained from edge excitations in the previous sections, we next allow free boundary conditions on Q, as it should be. For the special case where the Fermi energy lies in a d.o.s. gap, (4.2.30) has already been worked out in detail. A straightforward computation of S_{eff} for general N_r leads to

$$S_{\text{eff}}[t_0] = 2\pi i m \cdot q[Q_0] - \tfrac{m^2}{32\pi\eta\rho_{\text{e}}} \oint \text{tr}\ (\partial_x Q_0)^2. \tag{4.2.33}$$

The result naively diverges as $\eta \to 0$ and this clearly reflects the effect of the massless excitations in the problem. In the limit $\eta \to 0$, the field t_0 entering (4.2.33) is forced to obey the classical equations of motion (defined along the sample edge),

$$\partial_x\{Q(\eta - \tfrac{1}{2}v_{\text{d}}\partial_x Q)\} = 0 \tag{4.2.34}$$

The solution Q_0 =constant at the edge means that spherical boundary conditions are automatically generated by the background field procedure as $\eta \to 0$. The effect of S_{eff} has now been reduced to a phase factor which is immaterial provided the Hall conductance precisely equals an integer. Physically, the phase factor describes an integer number of edge electrons which have crossed the Fermi level as a result of the background field insertion.

The same procedure can be repeated for the theory with $\sigma^0_{\text{xx}} \neq 0$, making use of the fact that a mass gap exists in the system of large wavelength excitations, i.e. a finite localisation length ξ. One expects (4.2.33) to be modified according to

$$S_{\text{eff}}[t_0] = -\tfrac{\sigma_{\text{xx}}}{8}\text{Tr}\ (\nabla Q_0)^2 + 2\pi i \sigma_{\text{xy}} q[Q_0] - g_m L_\eta \oint \text{tr}\ (\partial_x Q_0)^2 \tag{4.2.35}$$

where the σ_{ij} represent the 'conductances',

$$\sigma_{\text{xx}} = f_{\text{xx}}(\eta\xi^2) \approx \mathcal{O}(\eta\xi^2) \quad ; \quad \sigma_{\text{xy}} = f_{\text{xy}}(\eta\xi^2) \approx m + \mathcal{O}(\eta^2\xi^4), \tag{4.2.36}$$

$g_m = m/2$ is the quantised 1D conductance of the chiral edge states and $L_\eta = m/16\pi\eta\rho_{\text{e}}$ is the length scale induced by the frequency. In the limit $\eta \to 0$ the Q_0 is forced to obey not only the classical equations of motion on the edge (4.2.34), but also those arising from the bulk kinetic term in (4.2.35). The solutions are known as *instantons* and just as has happened before in the trivial example with a d.o.s. gap in the bulk, the effect of S_{eff} reduces to that of a phase factor which is immaterial as long as σ_{xy} =integer. Therefore, the quantum Hall effect can be understood in terms of a continuous symmetry which is dynamically restored in the limit of large length scales.

4.3 Derivation of the full edge theory

In this section we derive the full Q-field theory on the edge in the presence of gauge fields and interactions.

4.3.1 Preliminaries

Our starting point is the truncated $\widetilde{Q}$-field theory in the presence of the plasmon field λ and external potentials A_μ (2.2.18),

$$
\begin{aligned}
S[A,\widetilde{Q},\lambda] &= -\tfrac{1}{2g}\mathrm{Tr}\,\widetilde{Q}^2 + \mathrm{Tr}\,\ln[i\omega + i\hat{A}_0 + i\hat{\lambda} + \mu - \hat{\mathcal{H}} + i\widetilde{Q}] \\
&\quad -\tfrac{1}{2}\beta\int \mathrm{d}^2x\mathrm{d}^2x'\ \lambda^\dagger(\vec{x})U_0^{-1}(\vec{x}-\vec{x}')\lambda(\vec{x}'). \qquad (4.3.1)
\end{aligned}
$$

In order to facilitate calculations we choose our sample to be the half plane $y>0$. The notation $\int\mathrm{d}^2x$ stands for integration over this half plane. Integration over the boundary $y=0$ is denoted by $\oint\mathrm{d}x$. Integration over the infinite plane will be written as $\int_\infty \mathrm{d}^2x$.

A theory for the edge is obtained by choosing the chemical potential μ approximately halfway between Landau energies, where the bulk d.o.s. is virtually zero if the disorder is not too strong. The saddle point equation for $\widetilde{Q}$ is given by

$$\widetilde{Q}_{\mathrm{sp}} \propto \rho_0 T^{-1}\Lambda T \qquad (4.3.2)$$

where ρ_0 is the d.o.s. Since we are interested in the limit $\rho_0\to 0$, we may replace the full expression for $\widetilde{Q}$ by a much simpler one,

$$\widetilde{Q} \to \varepsilon T^{-1}\Lambda T = \varepsilon Q, \qquad \varepsilon \ll 1. \qquad (4.3.3)$$

From [63] we know that the forms $\widetilde{Q}=T^{-1}PT$ and (4.3.3) give rise to identical results as long as the bulk d.o.s. ρ_0 can be safely taken to zero. However, in order to deal with the complications of $U(1)$ gauge invariance, there is considerable advantage in working with the simplified expression εQ, and we will refer to the details of more elaborate analyses only when necessary.

In order to facilitate the expansion of the Tr ln term in (4.3.1) we perform a gauge transformation that sets $A_0+\lambda=0$ by acting with a suitable W-rotation. We introduce the following notation

$$A_\mu^{(\lambda)} = A_\mu + \delta_{\mu 0}\lambda \qquad W = \exp\left[-i\textstyle\sum_{n\alpha}\mathrm{I}_n^\alpha(\frac{A_0^{(\lambda)}}{\partial_0})_n^\alpha\right] \qquad R = WQW^{-1}. \qquad (4.3.4)$$

The λ is absorbed into A_0, and the gauge invariant quantity R represents the Q-matrix in the gauge where $A_0^{(\lambda)}=0$. We will sometimes use notation with derivatives in the denominator. For any bosonic function f we define

$$(\tfrac{f}{\partial_0})_n := \tfrac{f_n}{-i\nu_n} \quad , \quad n\neq 0 \qquad (4.3.5)$$

For functions without a zero frequency component (which is the only kind we are dealing with) it is easily checked that $\partial_0\frac{f}{\partial_0} = \frac{\partial_0 f}{\partial_0} = f$. The vector potential $\vec{A}$ in the gauge $A_0^{(\lambda)}=0$ is denoted as $\vec{z}$

$$\vec{z} = \vec{A} - \nabla\tfrac{A_0^{(\lambda)}}{\partial_0} = \tfrac{\vec{E}-\nabla\lambda}{\partial_0}, \qquad (4.3.6)$$

where $\vec{E}=\partial_0\vec{A}-\nabla A_0$. The last expression shows that $\vec{z}$ is gauge invariant. Using the notation of (4.3.4)–(4.3.6) the action (4.3.1) becomes

$$
\begin{aligned}
S[Q,\lambda,A] &= -\tfrac{1}{2}\beta\int\mathrm{d}^2x\mathrm{d}^2x'\ \lambda^\dagger(\vec{x})U_0^{-1}(\vec{x}-\vec{x}')\lambda(\vec{x}') \\
&\quad +\mathrm{Tr}\,\ln\left[i\omega+\mu-\tfrac{1}{2m_e}(\overleftarrow{\pi}^{\mathrm{st}}-\hat{\vec{z}})\cdot(\vec{\pi}^{\mathrm{st}}-\hat{\vec{z}})+i\varepsilon R\right], \qquad (4.3.7)
\end{aligned}
$$

with $\vec{\pi}^{\rm st}$ the zero frequency part of $\vec{\pi}$ (2.2.6),

$$\vec{\pi}^{\,\rm st} = -i\nabla - \vec{A}^{\rm st} \tag{4.3.8}$$

where $\nabla\times\vec{A}^{\rm st}$ is the static magnetic field. The superscript 'st' on $\vec{\pi}$ will be dropped from this point on.

4.3.2 Expansion of the Tr ln

Let us take a closer look at the last term in (4.3.7), $X := {\rm Tr}\ \ln[i\omega+\mu-\mathcal{H}_z+i\varepsilon R]$. Introducing the notation

$$D_\omega = TW^{-1}\omega WT^{-1} \qquad ; \qquad D_{\vec{z}} = TW^{-1}(\tfrac{1}{i}\nabla - \hat{\vec{z}})WT^{-1} \tag{4.3.9}$$

(where $D_{\vec{z}}$ is not a differential operator) we can write

$$X = {\rm Tr}\ \ln[iD_\omega + \mu + i\varepsilon\Lambda - \tfrac{1}{2m_{\rm e}}(\overleftarrow{\pi}\cdot\vec{\pi} + \overleftarrow{\pi}\cdot D_{\vec{z}} + D_{\vec{z}}\cdot\vec{\pi} + D_{\vec{z}}^2)]. \tag{4.3.10}$$

Expansion to first order in D_ω and $D_{\vec{z}}$ yields

$$X \approx {\rm Tr}\ \ln G_0^{-1} + i{\rm Tr}\ G_0 D_\omega - \tfrac{1}{2m_{\rm e}}{\rm Tr}\ [G_0\overleftarrow{\pi}\cdot D_{\vec{z}} + G_0 D_{\vec{z}}\cdot\vec{\pi}] \tag{4.3.11}$$

where G_0 is the bare Green's function $[\mu-\mathcal{H}_0+i\varepsilon\Lambda]^{-1}$. The Green's function can be expressed in terms of the eigenfunctions φ_{nj} of the bare Hamiltonian $\mathcal{H}_0 = \frac{1}{2m_{\rm e}}\overleftarrow{\pi}\cdot\vec{\pi}$,

$$\langle x|G_0|x\rangle = \sum_{nj}\frac{|\varphi_{nj}(x)|^2}{\mu - E_{nj} + i\varepsilon\Lambda} \tag{4.3.12}$$

$$\langle x|\frac{G_0\overleftarrow{\pi} + \vec{\pi}\,G_0}{2m_{\rm e}}|x\rangle = \sum_{nj}\frac{\varphi_{nj}^*\frac{1}{i}\nabla\varphi_{nj} - \varphi_{nj}\frac{1}{i}\nabla\varphi_{nj}^* - 2\varphi_{nj}^*\varphi_{nj}\vec{A}^{\rm st}}{2m_{\rm e}(\mu - E_{nj} + i\varepsilon\Lambda)}. \tag{4.3.13}$$

Using the general relation $\rho(x) = -\frac{1}{\pi}{\rm Im}\ G^+(x,x)$ for the d.o.s. at the Fermi energy μ, we get

$$\langle x|G_0|x\rangle = -i\pi\rho(x)\Lambda + c(x)\mathbb{1} \quad ; \quad \langle x|\frac{G_0\ \overleftarrow{\pi} + \vec{\pi}G_0}{2m_{\rm e}}|x\rangle = -i\pi\vec{\jmath}(x)\Lambda + \vec{c}(x)\mathbb{1}, \tag{4.3.14}$$

where $\vec{\jmath}(x)$ is the current density per energy interval at the Fermi energy. The c and $\vec{c}$ are real functions that disappear from the last two traces in (4.3.11). We can now write X in the form

$$\begin{aligned} X &\approx {\rm Tr}\ \ln G_0^{-1} + \pi\int{\rm d}^2x\ \rho(x){\rm tr}\ \Lambda D_\omega + i\pi\int{\rm d}^2x\ \vec{\jmath}(x)\cdot{\rm tr}\ \Lambda D_{\vec{z}} \\ &= {\rm Tr}\ \ln G_0^{-1} + \pi\int{\rm d}^2x\ \rho(x){\rm tr}\ \omega R + i\pi\int{\rm d}^2x\ \vec{\jmath}(x)\cdot{\rm tr}\ [\tfrac{1}{i}\Lambda W^{-1}T\nabla(T^{-1}W) - \hat{\vec{z}}R]. \end{aligned} \tag{4.3.15}$$

Since μ lies in a gap, the d.o.s. and the current density are nonzero only at the edge. This means that the 2D integral becomes a line integral. If we assume constant $\rho_{\rm e}$ and $\vec{\jmath}$ on the edge, the resulting expression for X is

$$X \approx {\rm Tr}\ \ln G_0^{-1} + \pi\rho_{\rm e}\oint{\rm d}x\ {\rm tr}\ \omega R - i\tfrac{m}{2}\oint{\rm d}\vec{x}\cdot{\rm tr}\ \hat{\vec{z}}R + mS_{\rm top}[R] \tag{4.3.16}$$

where we have used that $\partial I_{\text{edge}}/\partial\mu = m/2\pi$ with the filling fraction m integer valued. Eq. (4.3.16), however, is not yet the complete answer. This can be seen from a different expansion procedure which can be followed in the special case where $T=\mathbb{1}$ and $W=\mathbb{1}$. In this case we have, instead of (4.3.10),

$$\begin{aligned} X_2 &= \text{Tr}\ \ln[i\omega+\mu-\mathcal{H}_z+i\varepsilon\Lambda] \approx \text{Tr}\ \ln G^{-1} - \tfrac{1}{2}\text{Tr}\left[\frac{G(\overleftarrow{\pi}\cdot\hat{\vec{z}}+\hat{\vec{z}}\cdot\vec{\pi})}{2m_{\text{e}}}\right]^2 - \tfrac{1}{2m_e}\text{Tr}\ \hat{\vec{z}}^2 G \\ G_{nm} &= G_n\delta_{nm} = \delta_{nm}[i\omega_n+\mu-\tfrac{1}{2m_{\text{e}}}\overleftarrow{\pi}\cdot\vec{\pi}]^{-1}. \end{aligned} \tag{4.3.17}$$

This expression can be written as

$$\begin{aligned} X_2 &\approx \text{Tr}\ \ln G^{-1} - \tfrac{1}{2}\sum_{ij}\sum_{n\alpha}\int \mathrm{d}^2x\mathrm{d}^2x'\ (z_i)_n^\alpha(x)(z_j)_{-n}^\alpha(x')(\Pi_{ij})_n^\alpha(x,x') \\ &\quad -\tfrac{1}{2m_e}\int \mathrm{d}^2x\ \vec{z}^\dagger\cdot\vec{z}\ \text{tr}\ G(x,x). \end{aligned} \tag{4.3.18}$$

The 'polarisation operator' Π_{ij} is given by

$$\begin{aligned} (\Pi_{ij})_n^\alpha(x,x') &= (\tfrac{1}{2m_{\text{e}}})^2\text{tr}\ [G(x,x')(\overleftarrow{\pi}_i+\vec{\pi}_i)I_n^\alpha G(x',x)(\overleftarrow{\pi}_j+\vec{\pi}_j)I_{-n}^\alpha] \\ &= (\tfrac{1}{2m_{\text{e}}})^2\sum_k G_{k+n}(x,x')(\overleftarrow{\pi}_i+\vec{\pi}_i)G_k(x',x)(\overleftarrow{\pi}_j+\vec{\pi}_j) \end{aligned} \tag{4.3.19}$$

The frequency sum can be split in two parts: (I) k and $k+n$ have the same sign; (II) k and $k+n$ have opposite signs. Case II has been done in great detail in the context of the SCBA. The conclusion is that (II) does not contribute either to σ_{xx} or σ_{xy} when μ is in a d.o.s. gap. Case I for $i\neq j$, using the relation $\overleftarrow{\pi}+\vec{\pi} = -i2m_{\text{e}}[G^{-1},\vec{x}]$, gives rise to the familiar 'Středa' form for σ_{xy}. For $i=j$ the two contributions in (4.3.18) cancel each other. We arrive at the following expression,

$$X_2 \approx \text{Tr}\ \ln G^{-1} + \tfrac{1}{2}m\sum_{n\alpha}\int \mathrm{d}^2x\ n\ \vec{z}_n^\alpha\times\vec{z}_{-n}^\alpha. \tag{4.3.20}$$

Now we have to find a match between the first order result (4.3.16) for $T\neq\mathbb{1}$, $W\neq\mathbb{1}$ and the second order result (4.3.20) for $T=\mathbb{1}$, $W=\mathbb{1}$. Up to a constant arising from the difference between G_0 and G, this match is given by

$$\begin{aligned} &\text{Tr}\ \ln G^{-1} + \pi\rho_{\text{e}}\oint \mathrm{d}x\ \text{tr}\ \omega R + m\left(\tfrac{1}{8}\varepsilon^{ij}\text{Tr}\ R[D_i,R][D_j,R] - \tfrac{i}{2}\text{Tr}\ R\nabla\times\hat{\vec{z}}\right) \\ &= \text{Tr}\ \ln G^{-1} + \tfrac{m}{2v_{\text{d}}}\oint \mathrm{d}x\ \text{tr}\ R(\omega - iv_{\text{d}}\hat{z}_x) + mS_{\text{top}}[R] - \tfrac{im\beta}{4\pi}\int \mathrm{d}^2x\ \ \vec{z}^\dagger\times\partial_0\vec{z} \end{aligned} \tag{4.3.21}$$

with v_{d} the electron drift velocity at the edge,

$$v_{\text{d}} = m/2\pi\rho_{\text{e}}. \tag{4.3.22}$$

Now we wish to write the result (4.3.21) in terms of Q and A_μ. Using the $\mathcal{F}$-algebra relations presented in section 2.2.4, we find that the various expressions in terms of R and $\vec{z}$ take the following form

$$
\begin{aligned}
\mathrm{tr}\,\omega R &= \mathrm{tr}\,\omega Q + \mathrm{tr}\,\hat{A}_0^{(\lambda)} Q - \tfrac{\beta}{2\pi}[A_0^{(\lambda)}]^\dagger A_0^{(\lambda)} \\
\mathrm{tr}\,R\hat{z}_x &= \mathrm{tr}\,Q\hat{A}_x - \mathrm{tr}\,Q\partial_x \tfrac{A_0^{(\lambda)}}{\partial_0} - \tfrac{\beta}{\pi} A_x^\dagger A_0^{(\lambda)} + \tfrac{\beta}{\pi}[A_0^{(\lambda)}]^\dagger \partial_x \tfrac{A_0^{(\lambda)}}{\partial_0} \\
S_{\mathrm{top}}[R] &= S_{\mathrm{top}}[Q] - \tfrac{i}{2}\oint \mathrm{d}x\ \mathrm{tr}\,Q\partial_x \tfrac{A_0^{(\lambda)}}{\partial_0} + \tfrac{i\beta}{4\pi}\oint \mathrm{d}x\ [A_0^{(\lambda)}]^\dagger \partial_x \tfrac{A_0^{(\lambda)}}{\partial_0} \qquad (4.3.23) \\
\beta\int \mathrm{d}^2x\ \vec{z}^\dagger \times \partial_0 \vec{z} &= -\sum_\alpha \int [A^{(\lambda)}]^\alpha \wedge \mathrm{d}[A^{(\lambda)}]^\alpha - \beta \oint \mathrm{d}x\ [\partial_0 A_x - \partial_x A_0^{(\lambda)}]^\dagger \tfrac{A_0^{(\lambda)}}{\partial_0}.
\end{aligned}
$$

Substitution of these expressions into (4.3.21) and inclusion of the plasmon term $\propto \int U_0^{-1}\lambda^\dagger\lambda$ yields the gauge and $\mathcal{F}$-invariant effective edge action $S[Q, A, \lambda]$

$$
\begin{aligned}
S_{\mathrm{eff}}^{\mathrm{edge}}[Q, A, \lambda] &= S_{\mathrm{c}}[\lambda] + S_{\mathrm{b}}[\lambda, A] + S_{\mathrm{Q}}[Q, \lambda, A] \qquad (4.3.24) \\
S_{\mathrm{c}} &= -\tfrac{1}{2}\beta \int \mathrm{d}^2x \mathrm{d}^2x'\ \lambda^\dagger(x) U_0^{-1}(\vec{x} - \vec{x}')\lambda(x') \\
S_{\mathrm{b}} &= \tfrac{im}{4\pi}\sum_\alpha \int [A^{(\lambda)}]^\alpha \wedge \mathrm{d}[A^{(\lambda)}]^\alpha - \tfrac{m\beta}{4\pi v_{\mathrm{d}}}\oint \mathrm{d}x\ [A_0^{(\lambda)}]^\dagger A_-^{(\lambda)} \\
S_{\mathrm{Q}} &= \tfrac{m}{2v_{\mathrm{d}}}\oint \mathrm{d}x\ \mathrm{tr}\,Q(\omega + \hat{A}_-^{(\lambda)}) + mS_{\mathrm{top}}[Q].
\end{aligned}
$$

The first term is the Coulomb energy contribution from the plasmon field; the S_{b} is a 'boson' action (this adjective will become clear later on); the last expression, S_{Q}, contains the action for the Q field and the coupling of Q with λ and A_μ. We have defined a 'minus' direction as follows,

$$
\partial_- = \partial_0 - iv_{\mathrm{d}}\partial_x \quad ; \quad A_-^{(\lambda)} = A_0^{(\lambda)} - iv_{\mathrm{d}}A_x \qquad (4.3.25)
$$

reflecting the chirality inherent in the problem.

4.3.3 Combined bulk and edge theory

Eq. (4.3.24) shows us how to treat edge states and can be used to obtain a generalised action $S[Q, A]$ for the case where both bulk and edge d.o.s. are nonzero. In the limiting case of zero edge d.o.s. the bulk theory (2.3.18) is obtained, and we derive an effective theory on the edge in the opposite limit of zero bulk d.o.s. We make contact between the edge action (4.3.24) and the bulk action (2B.13) by writing for the density of states

$$
\rho(\vec{x}) = \rho_0 + \rho_{\mathrm{e}}\delta(y). \qquad (4.3.26)
$$

Outside the sample, at $y<0$, the d.o.s. isn't of course ρ_0 but zero; this fact will be encoded in the fields Q and A_μ, which will be considered to vanish outside the sample. The reason for this construction is that we wish to write everything in momentum representation without encountering the complications that step functions in $\rho(\vec{x})$ or in the integrand of $\int_\infty \mathrm{d}^2x$ would cause. We define the screened Coulomb interaction in the bulk in the usual way,

$$
U_{\mathrm{scr}}(\vec{k}) = U_0(\vec{k})/[1 + 2\pi\rho_0 U_0(\vec{k})]. \qquad (4.3.27)
$$

A quantity that will often be encountered is the Fourier transform of a function $f(\vec{x})$ at $y=0$. We write this transform as $f(k_x)$,

$$
f(\vec{x})|_{y=0} \ \leftrightarrow\ f(k_x) := (2\pi)^{-1/2}\int \mathrm{d}k_y\ f(\vec{k}) \qquad (4.3.28)
$$

Using this notation, the generalisation of (4.3.24) for nonzero $\sigma^0_{\rm xx}$ and bulk d.o.s. can be written as

$$
\begin{aligned}
S[Q,A,\lambda] &= S_0[Q,A] + S_\lambda[\lambda,Q,A] && (4.3.29)\\
S_0[Q,A] &= -\tfrac{1}{8}\sigma^0_{\rm xx}\mathrm{Tr}\,[\vec{D},Q]^2 && (4.3.30)\\
&\quad +\sigma^0_{\rm xy}S_{\rm top} - \tfrac{i}{2}\sigma^0_{\rm xy}\oint \mathrm{d}x\ \mathrm{tr}\ \hat{A}_x Q + \tfrac{i}{4\pi}\sigma^0_{\rm xy}\left[\sum_\alpha \int A^\alpha\wedge \mathrm{d}A^\alpha + \beta\oint \mathrm{d}x\ A_0^\dagger A_x\right]\\
&\quad +\tfrac{i}{2}\sigma^{\rm II}_{\rm xy}\int \mathrm{d}^2x\ B^\dagger L_2 + 2\pi\int \mathrm{d}^2x\ \rho(\vec{x})[\tfrac{\pi}{\beta}\mathrm{tr}\ \eta Q + \tfrac{1}{2}\mathrm{tr}\ \hat{A}_0 Q - \tfrac{\beta}{4\pi}A_0^\dagger A_0]\\
S_\lambda[\lambda,Q,A] &= -\tfrac{\beta}{4\pi}\int \mathrm{d}^2p\mathrm{d}^2q\ \lambda^\dagger(-\vec{p})\lambda(\vec{q})\left[\tfrac{1}{U_{\rm scr}(\vec{q})}\delta(\vec{p}-\vec{q}) + \rho_{\rm e}\delta(p_x-q_x)\right]\\
&\quad +\tfrac{1}{2}\int \mathrm{d}^2q\ \lambda^\dagger(-\vec{q})\left[2\pi\rho_0 L_1(\vec{q}) + \rho_{\rm e}\sqrt{2\pi}L_2(q_x)\right]. && (4.3.31)
\end{aligned}
$$

Here we have defined the gauge invariant quantities L_1 in the bulk and L_2 at the edge according to

$$(L_1)^\alpha_n = \mathrm{tr}\ \mathrm{I}^\alpha_{-n}Q - \tfrac{\beta}{\pi}(A_0)^\alpha_n + i\tfrac{\beta}{2\pi^2\rho_0}\sigma^{\rm II}_{\rm xy}B^\alpha_n \quad;\quad (L_2)^\alpha_n = \mathrm{tr}\ \mathrm{I}^\alpha_{-n}Q - \tfrac{\beta}{\pi}(A_0)^\alpha_n \qquad (4.3.32)$$

The $\sigma^{\rm II}_{\rm xy}$ goes to the integer value m in the limit of $\rho_0\to 0$. From the quadratic part in λ we find the following propagator in 2D momentum space

$$G(\vec{k},\vec{p}) = U_{\rm scr}(\vec{k})\delta(k_x-p_x)\left[\delta(k_y-p_y) - \rho_{\rm e}U_{\rm scr}(\vec{p})\left\{1+\rho_{\rm e}\sqrt{2\pi}U_{\rm scr}(k_x)\right\}^{-1}\right] \qquad (4.3.33)$$

Integrating out λ yields the following quadratic expression in L_1 and L_2

$$\tfrac{\pi}{4\beta}\int \mathrm{d}^2k\mathrm{d}^2p\ G(\vec{k},\vec{p})\ [2\pi\rho_0 L_1^\dagger(-\vec{k}) + \rho_{\rm e}\sqrt{2\pi}L_2^\dagger(-k_x)][2\pi\rho_0 L_1(\vec{p}) + \rho_{\rm e}\sqrt{2\pi}L_2(p_x)]. \quad (4.3.34)$$

Substitution of (4.3.33) into (4.3.34) yields four terms which can be organised in powers of L_1 and L_2. The quadratic term in L_1 consists of a normal bulk term and a complicated expression at the edge,

$$\tfrac{\pi^3\rho_0^2}{\beta}\int \mathrm{d}^2k\ U_{\rm scr}L_1^\dagger L_1 - \tfrac{\pi^3\rho_0^2}{\beta}\rho_{\rm e}\int \mathrm{d}k_x\ \left\{1+\rho_{\rm e}\sqrt{2\pi}U_{\rm scr}(k_x)\right\}^{-1}\left|\int \mathrm{d}k_y\ U_{\rm scr}(\vec{k})L_1(\vec{k})\right|^2. \qquad (4.3.35)$$

The expression $\{\cdots\}^{-1}$ is analogous to $Y(\vec{p})$ (2.3.7), with ρ_0 replaced by $\rho_{\rm e}$ and U_0 by $U_{\rm scr}$. In the following we will denote it by $Y_{\rm edge}(k_x)$. The quantity involving $\int \mathrm{d}k_y$ has the following form in coordinate space

$$\sqrt{2\pi}\int \mathrm{d}k_y\ U_{\rm scr}(\vec{k})L_1(\vec{k}) \ \leftrightarrow\ \int \mathrm{d}^2r\ U_{\rm scr}(x_0,\vec{r})L_1(\vec{r}), \qquad (4.3.36)$$

i.e. it is the potential at a point x_0 at the edge due to the screened interaction with L_1 everywhere in the bulk. The cross term $L_1\cdot L_2$ is given by

$$\tfrac{\pi^2}{\beta}\rho_0\rho_{\rm e}\sqrt{2\pi}\int \mathrm{d}k_x\ Y_{\rm edge}(k_x)L_2^\dagger(-k_x)\int \mathrm{d}k_y\ U_{\rm scr}(\vec{k})L_1(\vec{k}) \qquad (4.3.37)$$

The quadratic term in L_2 is 'doubly' screened, once by ρ_0 and then additionally by $\rho_{\rm e}$ on the edge,

$$\tfrac{\pi}{4\beta}\rho_{\rm e}^2(2\pi)^{3/2}\int\!{\rm d}k_x\,\frac{U_{\rm scr}(k_x)}{1+\rho_{\rm e}\sqrt{2\pi}U_{\rm scr}(k_x)}L_2^\dagger(-k_x)L_2(k_x). \tag{4.3.38}$$

The contributions which are hardest to interpret are (4.3.37) and the second term in (4.3.35). We expect the Coulomb interactions to cause a shift in the scalar potential A_0 of the form $\int U_{\rm scr}L_1$. Since the free action (4.3.30) contains a term quadratic in A_0, we expect to see terms where the shift occurs squared or coupled to A_0. These are given by (4.3.37) and the second term in (4.3.35) respectively. The complete expression in L_1 and L_2 can also be written in the following form in coordinate space

$$\begin{aligned} S[L_1,L_2] &= \tfrac{\pi^2\rho_0^2}{2\beta}\int\!{\rm d}^2x{\rm d}^2x'\;U_{\rm scr}(\vec{x}-\vec{x}')L_1^\dagger(\vec{x})L_1(\vec{x}')+\tfrac{\pi^2}{2\beta}\rho_{\rm e}\oint{\rm d}x\;L_2^\dagger L_2 \qquad (4.3.39)\\ &\quad -\tfrac{\pi^2}{2\beta}\tfrac{\rho_{\rm e}}{\sqrt{2\pi}}\oint{\rm d}x{\rm d}x'\;Y_{\rm edge}(x-x')\left[L_2(x)-\rho_0\int\!{\rm d}^2r\;U_{\rm scr}(x,\vec{r})L_1(\vec{r})\right]^\dagger\\ &\quad\times\left[L_2(x')-\rho_0\int\!{\rm d}^2r\;U_{\rm scr}(x',\vec{r})L_1(\vec{r})\right] \end{aligned}$$

This expression describes a quite nontrivial interplay between bulk and edge and should be able to shed some light on the full interaction problem. In this thesis, however, only the limiting scenarios of zero $\rho_{\rm e}$ or ρ_0 will be considered.

4.3.4 Limiting cases

The two limits $\rho_{\rm e}\to 0$ and $\rho_0\to 0$ are easily taken. In the former case only the first term of (4.3.39) survives. Added to S_0 (4.3.30), this precisely gives the bulk action (2.3.18). The latter case is more complicated. In the limit $\rho_0\to 0$, the first term in (4.3.39) vanishes. The expression $2\pi\rho_0 L_1$ does not vanish but becomes $i\frac{\beta}{\pi}\sigma_{\rm xy}^{\rm II}B$. At the same time we have $\sigma_{\rm xy}^{\rm II}\to\sigma_{\rm xy}^0=m$, since we are talking about a d.o.s. gap. The screened interaction in the bulk $U_{\rm scr}$ becomes U_0. Eq. (4.3.39) can then be written as

$$\begin{aligned} S[L_1,L_2] &= \tfrac{\pi^2}{2\beta}\rho_{\rm e}\oint{\rm d}x\;L_2^\dagger L_2-\tfrac{\pi^2}{2\beta}\tfrac{\rho_{\rm e}}{\sqrt{2\pi}}\sum_{n\alpha}\oint{\rm d}x{\rm d}x'\;Y_{\rm edge}^{\rho_0=0}(x-x')\left[{\rm tr}\;{\rm I}_n^\alpha Q-\tfrac{\beta}{\pi}(A_0^{\rm eff})_{-n}^\alpha\right](x)\\ &\quad\times\left[{\rm tr}\;{\rm I}_{-n}^\alpha Q-\tfrac{\beta}{\pi}(A_0^{\rm eff})_n^\alpha\right](x') \end{aligned} \tag{4.3.40}$$

where we have defined an effective scalar potential that is shifted due to the fluctuations of the magnetic field in the bulk. We will use the general notation $A_\mu^{\rm eff}$, but only the scalar component is shifted,

$$\vec{A}^{\rm eff}=\vec{A}\quad;\quad A_0^{\rm eff}(\vec{x}):=A_0(\vec{x})+\tfrac{im}{2\pi}\int\!{\rm d}^2x'\;U_0(\vec{x}-\vec{x}')B(\vec{x}'). \tag{4.3.41}$$

Adding (4.3.40) to the free action $S_0(\sigma_{\rm xx}^0=0,\rho_0=0,\sigma_{\rm xy}^{\rm II}=\sigma_{\rm xy}^0=m)$ (4.3.30) yields the full effective edge action for Q and A_μ in the presence of interactions. Here we present it in a form which most resembles the free particle result (4.3.24 with $\lambda=0$),

$$S_{\rm eff}^{\rm edge}[Q,A] = S_{\rm flux}[A]+S_{\rm b}[A]+S_Q[Q,A] \tag{4.3.42}$$

$$S_{\rm flux} = \tfrac{\beta}{2}(\tfrac{m}{2\pi})^2\int\!{\rm d}^2x{\rm d}^2x'\;B^\dagger(\vec{x})U_0(\vec{x}-\vec{x}')B(\vec{x}') \tag{4.3.43}$$

$$S_{\rm b} = \frac{im}{4\pi}\sum_\alpha \int (A^{\rm eff})^\alpha \wedge {\rm d}(A^{\rm eff})^\alpha - \frac{m\beta}{4\pi}\int {\rm d}k_x \, \frac{1}{v^{\rm eff}(k_x)}(A_0^{\rm eff})^\dagger A_c^{\rm eff} \tag{4.3.44}$$

$$S_Q = \frac{m}{2v_{\rm d}}\oint {\rm d}x \; {\rm tr}\, \omega Q + \frac{m}{2}\int {\rm d}k_x \, \frac{1}{v^{\rm eff}(k_x)} {\rm tr}\, Q(-k_x)\hat{A}_c^{\rm eff}(k_x) + mS_{\rm top}[Q]$$
$$+\frac{\rho_{\rm e}^2}{8\beta}(2\pi)^{5/2}\sum_{n\alpha}\int {\rm d}k_x \, U_{\rm scr}^{\rm edge}(k_x)\,\left|{\rm tr}\; {\rm I}_n^\alpha Q(k_x)\right|^2. \tag{4.3.45}$$

We have introduced a momentum dependent effective velocity $v^{\rm eff}$ that contains the Coulomb interaction,

$$v^{\rm eff}(k_x) := v_{\rm d} + \frac{m}{\sqrt{2\pi}}U_0(k_x). \tag{4.3.46}$$

This effective velocity changes the chiral 'minus' direction into what we call 'c'

$$\partial_c = \partial_0 - iv^{\rm eff}\partial_x \quad ; \quad A_c(k_x) = A_0(k_x) - iv^{\rm eff}(k_x)A_x(k_x). \tag{4.3.47}$$

The $U_{\rm scr}^{\rm edge}$ is the screened interaction at the edge,

$$U_{\rm scr}^{\rm edge}(k_x) = U_0(k_x)/[1+\rho_{\rm e}\sqrt{2\pi}U_0(k_x)]. \tag{4.3.48}$$

Notice that the Q-interaction part of the edge theory has exactly the same form as the bulk theory (2.3.3). The frequency term tr ωQ and the $U_{\rm scr}^{\rm edge}$ term combine to give a Finkelstein action on the edge plus a $1/v^{\rm eff}$ term that is analogous to the Y in (2.3.6). Comparison to the free action (4.3.24 without λ) shows that the effect of the interactions can be summarised as

- the replacement $A_0 \to A_0^{\rm eff}$.
- the replacement $v_{\rm d} \to v^{\rm eff}$ everywhere except in the frequency term.
- additional $U_0|B|^2$ and $U_{\rm scr}^{\rm edge}|{\rm tr}\; {\rm I}Q|^2$ interaction terms.

4.4 Chiral edge bosons

In this section we show that the iqHe edge action $S_{\rm eff}^{\rm edge}[Q,A]$ is equivalent to a theory of chiral edge bosons. We proceed as follows. First we integrate out the Q-field and obtain an effective action $S_{\rm eff}^{\rm edge}[A]$. This is an *exact* calculation, since we can use W-rotations to decouple Q from A_μ. Then we show that at $\sigma_{\rm xy} = m$, the quadratic A_μ terms in $S_{\rm eff}^{\rm edge}[A]$ can be decoupled by introducing bosonic Hubbard-Stratonovich fields $\varphi_1, \cdots, \varphi_m$. In other words, as far as the electromagnetic field is concerned, the Q-variable is equivalent to the fields φ_i. It turns out that the effective action $S[\varphi_i, A]$ describes chiral edge bosons.

In an intermediate stage, we use the effective action for A_μ to study the bulk and edge currents in the presence of Coulomb interactions. We relate Laughlin's gauge argument to the occurrence of a chiral anomaly at the edge whose amplitude is equal to the Hall conductance. The chiral anomaly persists when the bosons φ_i are introduced.

4.4.1 The noninteracting case

Let us first examine the simple case of noninteracting electrons, described by the action (4.3.24 without λ). The only coupling term is the tr $Q\hat{A}_-$. It is of course possible to expand Q in terms of V as in chapter 3 and then integrate out V in a loop expansion. However, a smarter way is possible which immediately yields the exact result. We always have the freedom to choose a certain gauge in which to perform the Q-integration. The result should be independent of the choice. A drastic simplification occurs when a gauge is chosen in which Q and A_μ decouple, since the Q-integration then leads to an unimportant constant. The gauge in which the decoupling occurs is $A_- = 0$ and it is obtained by performing a gauge transformation on A_μ according to

$$A_\mu \to A_\mu - \partial_\mu \Omega \qquad \partial_- \Omega = A_- \qquad \Omega = \partial_-^{-1} A_- + \Omega^{\rm res}. \tag{4.4.1}$$

The inverse of ∂_- has a natural definition in momentum space,

$$\begin{aligned}(\partial_-^{-1} F)(x,\tau) &= \tfrac{1}{(2\pi)^2} \int \mathrm{d}x'\mathrm{d}\tau'\; F(x',\tau') \int_{-\infty}^{\infty} \mathrm{d}k \int_{-\infty}^{\infty} \mathrm{d}\omega\; \tfrac{\exp[ik(x-x')-i\omega(\tau-\tau')]}{-i\omega+v_{\rm d}k} \\ &= \tfrac{1}{2\pi} \int \mathrm{d}x'\mathrm{d}\tau'\; F(x',\tau') \left[\tfrac{\theta(\tau-\tau')}{v_{\rm d}(\tau-\tau')-i(x-x')+\eta} + \tfrac{\theta(\tau'-\tau)}{v_{\rm d}(\tau-\tau')-i(x-x')-\eta}\right]\end{aligned} \tag{4.4.2}$$

with F an arbitrary function, θ the Heaviside step function and η a regulator. The operation ∂_-^{-1} does not commute with ∂_-. It is easily checked that

$$\partial_-(\partial_-^{-1} F) = F \quad ; \quad \partial_-^{-1}(\partial_- F) = F - F^{\rm res} \tag{4.4.3}$$

with $F^{\rm res}$ defined as that part of F which satisfies $\partial_- F = 0$. Another property of this operation is

$$\int \mathrm{d}x\mathrm{d}\tau\; F_1(\partial_-^{-1} F_2) = -\int \mathrm{d}x\mathrm{d}\tau\; (\partial_-^{-1} F_1) F_2. \tag{4.4.4}$$

The transformation (4.4.1) substituted into the free action $S^{\rm edge}_{\rm eff}[Q,A]$ yields the following effective action,

$$S[A] = \tfrac{im}{4\pi} \left[\sum_\alpha \int A^\alpha \wedge \mathrm{d}A^\alpha + \beta \oint \mathrm{d}x\; E_x^\dagger \left(\partial_-^{-1} A_- + \Omega^{\rm res}\right)\right]. \tag{4.4.5}$$

The presence of the 'residual' gauge freedom $\Omega^{\rm res}$ in (4.4.5) is necessary for gauge invariance; under a transformation $A_\mu \to A_\mu + \partial_\mu \chi$ the $\partial_-^{-1} A_-$ is shifted by an amount $\chi - \chi^{\rm res}$. The shift $\Omega^{\rm res} \to \Omega^{\rm res} + \chi^{\rm res}$ precisely compensates for the $\chi^{\rm res}$.

We can perform a check on the exact result (4.4.5) by comparing it to the result of a tree level calculation, where Q is now integrated out *without* the use of W-rotations. We parametrise T as follows

$$T = \exp i \begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix}. \tag{4.4.6}$$

(It can be checked that the Jacobian of this parametrisation is unity). This yields the tree level propagator

$$\left\langle \mathrm{tr}\,[\mathrm{I}^\beta_m Q(-p)]\mathrm{tr}\,[\mathrm{I}^\alpha_{-n} Q(q)] \right\rangle = \delta^{\alpha\beta}\delta_{mn}\delta(p-q)\tfrac{\beta v_{\rm d}}{2\pi m} \cdot \tfrac{\omega_n}{\omega_n + i v_{\rm d} q}. \tag{4.4.7}$$

This propagator is, in fact, exact, as can be checked by using the results of section 4.2.3, in particular (4.2.22) with the replacement $2\eta \to \omega_n$. Notice that the denominator $\omega_n + i v_{\rm d} q$ is the Fourier transform of ∂_-. The expression for $S[A]$ which results from (4.4.7) is exactly

(4.4.5), but Ω^{res} now arises as a residual gauge freedom in tr $\mathrm{I}Q$. The action (4.4.5) can also be written as a path integral over a total of m charge 1 bosons (for simplicity we take $m>0$),

$$S[A,\varphi_i] = \tfrac{i}{4\pi}\sum_{i=1}^{m}\left[\sum_{\alpha}\int A^\alpha\wedge \mathrm{d}A^\alpha - \beta\oint \mathrm{d}x\;(D_x\varphi_i^\dagger D_-\varphi_i - E_x^\dagger\varphi_i)\right], \tag{4.4.8}$$

where the covariant derivative D is defined as $D_\mu\varphi_i = \partial_\mu\varphi_i - A_\mu$. Under a general gauge transformation $A\to A+\mathrm{d}\chi$ the boson fields change as $\varphi_i\to\varphi_i+\chi$. When (4.4.8) is integrated over the φ_i, (4.4.5) is obtained, where the Ω^{res} now has its origin in a gauge freedom of the boson fields, $\Omega^{\text{res}} = \frac{1}{m}\sum_i\varphi_i^{\text{res}}$. Notice that the zero-momentum part of each boson field has to be excluded from the path integral, since the action does not depend on it. We mention that (4.4.8) is completely equivalent to a Chern-Simons bulk theory with gauge fields $g_1,\ldots,g_m$ that represent potentials for the electron currents, coupled to the external potentials A_μ,

$$S[A,g_i] = \tfrac{i}{4\pi}\sum_{\alpha}\sum_{i=1}^{m}\left[-\int g_i^\alpha\wedge \mathrm{d}g_i^\alpha + 2\int g_i^\alpha\wedge \mathrm{d}A^\alpha\right] \tag{4.4.9}$$

where the g_i have the gauge fixing constraint $(g_i)_-|_{\text{edge}} = 0$. In appendix 4B we explicitly show how integration over the potentials g_i leads to the action (4.4.8).

4.4.2 The Coulomb case

Now that we have seen how edge bosons are related to the free Q-theory, let us investigate the effect of interactions. It is instructive to rewrite (4.3.42) in the following form

$$\begin{aligned} S_{\mathrm{Q}}+S_{\mathrm{b}} \;=\;& \tfrac{im}{4\pi}\left[\sum_{\alpha}\int (A^{\text{eff}})^\alpha\wedge \mathrm{d}(A^{\text{eff}})^\alpha - \beta\oint \mathrm{d}x\; A_x^\dagger A_c^{\text{eff}}\right] + mS_{\text{top}}[Q]\\ &+S_{\mathrm{F}}^{\text{edge}}[Q] - \tfrac{m\pi}{4\beta}\sum_{n\alpha}\int \mathrm{d}k_x\,\tfrac{1}{v^{\text{eff}}(k_x)}\left|\text{tr }\mathrm{I}_{-n}^\alpha Q - \tfrac{\beta}{\pi}(A_c^{\text{eff}})_n^\alpha\right|^2, \end{aligned} \tag{4.4.10}$$

where $S_{\mathrm{F}}^{\text{edge}}$ is the edge analogue of the Finkelstein action (2.3.5) in the bulk theory,

$$S_{\mathrm{F}}^{\text{edge}}[Q] = \tfrac{\pi^2\rho_{\mathrm{e}}}{2\beta}\oint \mathrm{d}x\;\left[\sum_{n\alpha}\left|\text{tr }\mathrm{I}_n^\alpha Q\right|^2 + 4\text{tr }\eta Q\right]. \tag{4.4.11}$$

In (4.4.10), as in the bulk limit of (2.3.18), the gauge field couples to Q only via the $1/v^{\text{eff}}$ or 'Y' term. However, compared to the bulk case where the coupling results in the gauge invariant combination (tr $\mathrm{I}Q - \frac{\beta}{\pi}A_0$), the situation here is more subtle. The expression (tr $\mathrm{I}Q - \frac{\beta}{\pi}A_c^{\text{eff}}$) appearing in (4.4.10) is, in fact, gauge dependent, but this gauge dependence is exactly what one needs to compensate for the edge contributions resulting from gauge transformations of the 'boson' action S_{b} and the topological term. Therefore, the expression (4.4.10) is fully gauge invariant.

We now proceed as in section 4.4.1 and integrate out the Q field. This is done in the same way as for the noninteracting case; either by doing it directly or by choosing a gauge Ω such that Q decouples from A_μ in analogy with (4.4.1),

$$\partial_c\Omega = A_c^{\text{eff}} \quad;\quad \Omega = \partial_c^{-1}A_c^{\text{eff}} + \Omega^{\text{res}}. \tag{4.4.12}$$

The only difference lies in the fact that we now work with 'effective' quantities. The arguments about the 'residual' part of the electric field can again be applied, but now for the effective quantities (4.3.41, 4.3.46, 4.3.47). We then get the effective action for the external field A_μ in the presence of Coulomb interactions,

$$\begin{aligned} S[A] &= \tfrac{im}{4\pi}\left[\sum_\alpha \int (A^{\text{eff}})^\alpha \wedge \mathrm{d}(A^{\text{eff}})^\alpha + \beta \oint \mathrm{d}x\, (\partial_c^{-1} A_c^{\text{eff}} + \Omega^{\text{res}})^\dagger E_x^{\text{eff}}\right] \\ &\quad + \tfrac{\beta}{2}(\tfrac{m}{2\pi})^2 \int \mathrm{d}^2x\mathrm{d}^2x'\; B^\dagger(\vec{x}) U_0(\vec{x}-\vec{x}') B(\vec{x}'). \end{aligned} \tag{4.4.13}$$

Again, the difference with the free particle case is the appearance of a flux-flux term and various replacements by effective quantities.

4.4.3 Edge currents and Laughlin's gauge argument

The action (4.4.13) contains complete information on the response of the system to external electromagnetic fields. We define the current as $j^\mu(\vec{x}) = \delta S/\delta A_\mu(\vec{x})$. In this way we find

$$j^0(\vec{x}) = \tfrac{im}{2\pi}\left[B - \delta(y)\partial_c^{-1} E_x^{\text{eff}}\right] \tag{4.4.14}$$

$$j^1(\vec{x}) = -\tfrac{im}{2\pi}\left[E_y - \partial_y \int \mathrm{d}^2x'\; U_0(\vec{x}-\vec{x}') j^0(\vec{x}')\right] - \tfrac{mv_{\text{d}}}{2\pi}\delta(y)\partial_c^{-1} E_x^{\text{eff}} \tag{4.4.15}$$

$$j^2(\vec{x}) = \tfrac{im}{2\pi}\left[E_x - \partial_x \int \mathrm{d}^2x'\; U_0(\vec{x}-\vec{x}') j^0(\vec{x}')\right]. \tag{4.4.16}$$

It is easily verified that $\partial_\mu j^\mu = 0$. The edge currents are obtained by taking only those terms that possess a delta function. On the edge we get

$$j^0_{\text{edge}} = -\tfrac{im}{2\pi}\partial_c^{-1} E_x^{\text{eff}} \tag{4.4.17}$$

$$j^1_{\text{edge}} = -iv_{\text{d}} \cdot j^0_{\text{edge}}. \tag{4.4.18}$$

This yields for the edge anomaly

$$\partial_\mu j^\mu_{\text{edge}}(x) = -\tfrac{im}{2\pi}\left[E_x - \partial_x \int \mathrm{d}^2x'\; U_0(x,\vec{x}') j^0(\vec{x}')\right]. \tag{4.4.19}$$

Notice that the expression between square brackets is the total electric field due to externally applied field and the charge accumulations on the edge. By applying Laughlin's gauge argument [57] one can now directly relate the conductances defined by the bulk and edge theories. For example, let us do a linear response calculation for the case where N flux quanta h/e are created somewhere inside a hole in the sample. The charge q flowing from one edge into the other is found using (4.4.19),

$$\mathrm{d}q/\mathrm{d}\tau = -i\oint \partial_\mu j^\mu_{\text{edge}} = \tfrac{m}{2\pi}\mathrm{d}\Phi/\mathrm{d}\tau \tag{4.4.20}$$

where Φ is the total flux $N \cdot h/e$ enclosed by the contour integral. This yields $q = m \cdot N$, as it should.

4.4.4 Interacting chiral bosons

As in the free electron situation, we can write the theory (4.4.13) in terms of edge bosons coupled to the external field, exactly of the form (4.4.8), but now with effective quantities and an extra flux-flux term,

$$\begin{aligned} S[A,\varphi_i] &= \tfrac{i}{4\pi}\sum_{i=1}^{m}\left[\sum_{\alpha}\int (A^{\rm eff})^{\alpha}\wedge {\rm d}(A^{\rm eff})^{\alpha} - \beta\oint {\rm d}x\ (D_x\varphi_i^{\dagger}D_c^{\rm eff}\varphi_i - \varphi_i^{\dagger}E_x^{\rm eff})\right] \\ &\quad + \tfrac{\beta}{2}(\tfrac{m}{2\pi})^2\int {\rm d}^2x{\rm d}^2x'\ B^{\dagger}(\vec{x})U_0(\vec{x}-\vec{x}')B(\vec{x}'). \end{aligned} \tag{4.4.21}$$

As in the noninteracting case, this result is equivalent to a Chern-Simons bulk theory of the form (4.4.9). In this case the action for the electron currents is given by

$$\begin{aligned} S[A,g_i] &= \tfrac{i}{4\pi}\sum_{\alpha}\sum_{i=1}^{m}\left[-\int g_i^{\alpha}\wedge {\rm d}g_i^{\alpha} + 2\int g_i^{\alpha}\wedge {\rm d}(A^{\rm eff})^{\alpha}\right] \\ &\quad + \tfrac{\beta}{2}(\tfrac{m}{2\pi})^2\int {\rm d}^2x{\rm d}^2x'\ B^{\dagger}(\vec{x})U_0(\vec{x}-\vec{x}')B(\vec{x}') \end{aligned} \tag{4.4.22}$$

with the gauge fixing conditions

$$\left[(g_i^{\alpha})_-(k_x) - i\tfrac{m}{\sqrt{2\pi}}U_0(k_x)\sum_{a=1}^{m}(g_a^{\alpha})_x(k_x)\right]_{\rm edge} = 0. \tag{4.4.23}$$

It is very instructive to write (4.4.21) also in the following way

$$\begin{aligned} S &= -\tfrac{\beta}{2}(\tfrac{1}{2\pi})^2\sum_{i,j=1}^{m}\int_{\infty}{\rm d}^2x{\rm d}^2x'\ U_0(\vec{x}-\vec{x}')\nabla\times[\theta(y)\vec{D}\varphi_i(\vec{x})]^{\dagger}\nabla'\times[\theta(y')\vec{D}\varphi_j(\vec{x}')] \\ &\quad + \tfrac{i}{4\pi}\sum_{i=1}^{m}\left[\sum_{\alpha}\int A^{\alpha}\wedge {\rm d}A^{\alpha} - \beta\oint {\rm d}x\ (D_x\varphi_i^{\dagger}D_-\varphi_i - E_x^{\dagger}\varphi_i)\right], \end{aligned} \tag{4.4.24}$$

where θ is a step function. Notice that there are no effective quantities in this expression; the Coulomb interaction is completely contained in the first term. The charge density is given by $\frac{m}{2\pi}[B+\delta(y)m^{-1}\sum_i D_x\varphi_i]$. Notice also that we have written a two-dimensional integral containing φ_i, even though the boson fields only exist on the edge. This is not a problem, since the φ_i only get evaluated at the edge.

4.4.5 Tunnelling density of states

In the next chapter we are going to spend a lot of words on tunnelling exponents for 'fractional' edge quasiparticles. In order to set the stage for the discussion of tunnelling processes we now present the simple results for the iqHe. We are primarily interested in tunnelling from a Fermi liquid through 'vacuum' into the edge of a quantum Hall sample. The contribution to the tunnelling current from the Fermi liquid side is trivial. The quantity of interest at the sample's edge is the tunnelling d.o.s., which is proportional to the one particle Green's function $G(\tau_2-\tau_1,x_0)$,

$$G(\tau_2-\tau_1,x_0) = \langle\bar{\psi}^{\alpha}(x_0,\tau_2)\psi^{\alpha}(x_0,\tau_1)\rangle. \tag{4.4.25}$$

In terms of the Q-variable this expression is written as

$$\langle Q^{\alpha\alpha}(\tau_1,\tau_2,x_0)\rangle := \sum_n e^{i\nu_n(\tau_2-\tau_1)}\langle Q^{\alpha\alpha}_{nn}(x_0)\rangle, \tag{4.4.26}$$

where the expectation value is with respect to the action (4.3.42). The Green's function is a gauge dependent quantity and transforms as

$$e^{-i[\chi^\alpha(\tau_2,x_0)-\chi^\alpha(\tau_1,x_0)]}\langle Q^{\alpha\alpha}(\tau_1,\tau_2,x_0)\rangle. \tag{4.4.27}$$

Let us see what happens to (4.4.26) when the Q-field is integrated out. The gauge transformation (4.4.12) that decouples Q from A_μ in (4.3.42) changes (4.4.26) into

$$\left\langle \exp -i\left\{[\partial_c^{-1}A_c^{\text{eff}}]^\alpha(\tau_2,x_0) - [\partial_c^{-1}A_c^{\text{eff}}]^\alpha(\tau_1,x_0)\right\}\right\rangle \tag{4.4.28}$$

where the expectation value is now taken with respect to (4.4.13). When decoupling the quadratic edge term in A_μ (4.4.13) with the use of boson fields, this becomes

$$\left\langle \exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\ \partial_0\varphi_j^\alpha(\tau,x_0)\right\rangle, \qquad j=1,\ldots,m \tag{4.4.29}$$

with respect to (4.4.21). The decoupling is not a unique procedure, since combinations of the boson fields φ_i can be chosen other than (4.4.29). Any combination $\sum_i c_i\varphi_i$ will lead to the expression (4.4.28) after integration over the φ fields as long as $\sum_i c_i = 1$. However, the form $c_i=\delta_{ij}$ (4.4.29) is the only one that yields the correct fermionic result $1/(\tau_2-\tau_1)$.

$$\left\langle \exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\ \partial_0\varphi_i^\alpha(\tau,x_0)\right\rangle \propto (\tau_2-\tau_1)^{-S} \qquad S=1. \tag{4.4.30}$$

(See appendix 4A for the explicit calculation). Notice that we would have had a serious problem at this point if we had not excluded the zero-momentum components of the φ_i^α when we introduced these auxiliary fields. A redefinition of the integration measure, $\int\mathcal{D}\varphi \to \int\mathcal{D}[\varphi+f]$, with $\partial_x f(x,\tau)=0$, would yield a result depending on the arbitrary function f.

4.5 Long range disorder

4.5.1 Separation of edge channels

Long range disorder can cause the edge states of different Landau levels to become spatially separated. A potential fluctuation at the edge can lift all states in such a way that new 'edge' states are created (see Fig. 4.1). If the chemical potential lies between the shifted and unshifted energy of a Landau level, the edge states of this Landau level will be shifted from the edge into the sample. If there are several potential jumps of this kind, all the edge channels can become separated. They can also start wandering into the interior of the sample. We propose that 'edge channel separation' is the dominant effect of smooth potential fluctuations as opposed to 'inter-channel scattering' which only occurs when the potential changes abruptly. In this section we wish to embark on the problem of smooth potential fluctuations in the presence of the Coulomb interactions.

In order to fix the thought we imagine a quantum Hall sample with filling fraction $\nu=2-\varepsilon$. Fig 4.2a illustrates the equipotential contours. We may distinguish between the

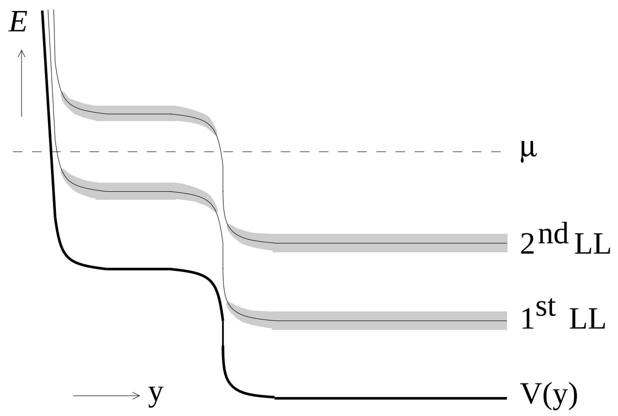

Figure 4.1: *Spatially separated edge channels.*

localised (closed) orbitals in the bulk of the sample and the extended edge states. This picture leads us to the idea of describing the chiral bosons by *one* field $\varphi(\vec{x})$ that lives on all the 'edges' instead of independent fields for every edge. The action (4.4.21) then becomes

$$\begin{aligned} S \;=\; & \tfrac{i}{4\pi}\left[\sum_\alpha \int nA^\alpha \wedge \mathrm{d}A^\alpha - \beta \sum_{a=1}^{M} s_a \oint_{C_a} \mathrm{d}x \left(D_x\varphi^\dagger[D_0\varphi - is_a v_\mathrm{d} D_x\varphi] - E_x^\dagger \varphi\right)\right] \\ & -\tfrac{\beta/2}{(2\pi)^2}\int \mathrm{d}^2x\mathrm{d}^2x'\; U_0(\vec{x}-\vec{x}')\nabla\times[n(\vec{x})\vec{D}\varphi(\vec{x})]^\dagger\nabla'\times[n(\vec{x}')\vec{D}\varphi(\vec{x}')]. \end{aligned} \tag{4.5.1}$$

The n is a function of position labelling the 'local' filling fraction: outside the sample $n(\vec{x})$ is zero; going inward, it increases by one every time you cross an edge, until it reaches its bulk value m. At the bulk orbitals, $n(\vec{x})$ jumps again. (In the case $\nu=2-\varepsilon$, depicted in Fig. 4.2a, $n(\vec{x})=1$ inside the closed orbitals).

Each edge is described by a contour labelled C_a, with $a=1,\cdots,m$ for the edge states and $a=m+1,\cdots,M$ for the closed bulk orbitals. The coordinate 'x', appearing in the edge terms, is defined on the contour and is taken in the positive (anticlockwise) direction. The symbol s_a,

$$s = (\overbrace{+1,\cdots,+1}^{m},-1,\cdots,-1) \tag{4.5.2}$$

incorporates the fact that the contours with $a\leq m$ and $a>m$ carry opposite current and charge densities. For simplicity we take the drift velocity v_d the same for all edges. We can integrate out the boson field to obtain the generalisation of (4.4.13),

$$\begin{aligned} S[A] \;=\; & \tfrac{i}{4\pi}\left[\sum_\alpha \int n(A^\mathrm{eff})^\alpha \wedge \mathrm{d}(A^\mathrm{eff})^\alpha + \beta \sum_a s_a \oint_{C_a} \mathrm{d}x\; (\partial_x\partial_c^{-1}A_c^\mathrm{eff} - A_x)^\dagger A_c^\mathrm{eff}\right] \\ & +\tfrac{\beta}{2}(\tfrac{1}{2\pi})^2\int \mathrm{d}^2x\mathrm{d}^2x'\; n(\vec{x})B^\dagger(\vec{x})U_0(\vec{x}-\vec{x}')n(\vec{x}')B(\vec{x}'). \end{aligned} \tag{4.5.3}$$

The notation ∂_c (at contour C_a) now has the sign s_a in front of the velocity and contains Coulomb interactions with *all* contours instead of just C_a itself. The definition of the 'effective' potential A_0^eff has also slightly changed,

$$A_0^\mathrm{eff}(\vec{x}) = A_0(\vec{x}) + \tfrac{i}{2\pi}\int \mathrm{d}^2x'\; U_0(\vec{x}-\vec{x}')\; n(\vec{x}')B(\vec{x}'). \tag{4.5.4}$$

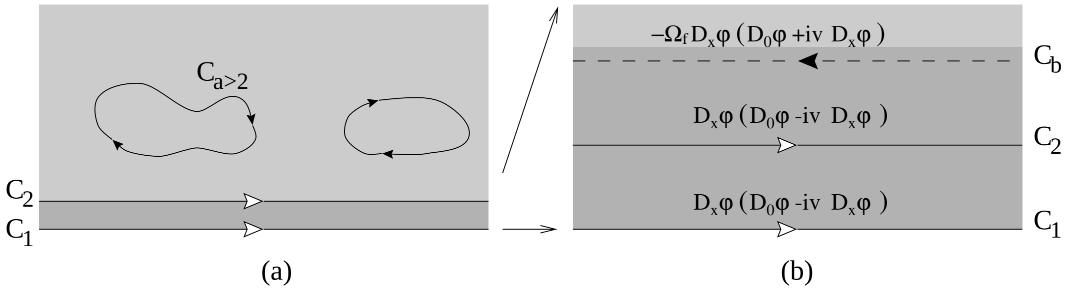

Figure 4.2: (a) *Equipotential contours corresponding to filling fraction* $\nu = 2 - \varepsilon$. (b) *Effective edge theory for filling fraction* $\nu = 2 - \varepsilon$. *The dashed line represents the (anti)chiral contribution from the bulk orbitals.*

For completeness, in appendix 4D we also present the generalisation of the action $S[Q, A]$ (4.3.42) for the case of separated edge channels. Note that we are addressing the situation where the chemical potential is away from the narrow 'percolation' regime indicated by W_0 in Fig. 4.5.

4.5.2 Hall conductance

We will next exploit the simplicity of our model and demonstrate that the Hall conductance and the tunnelling density of edge states are fundamentally different quantities that correspond to completely different physical processes. First, it is straightforward to generalise the results of section 4.4.3 to include the separated edge channels and the bulk states into Laughlin's flux argument. Differentiating the action (4.5.3) with respect to A_μ, we obtain the generalised form of the currents (4.4.14–4.4.16),

$$\begin{aligned} j^0(\vec{x}) &= \tfrac{i}{2\pi}\left[n(\vec{x})B - \sum_{a=1}^{M} s_a \delta(\vec{x} \text{ on } C_a)\tfrac{E_x^{\text{eff}}}{\partial_c}\right] \qquad (4.5.5) \\ j^i(\vec{x}) &= -i\tfrac{n(\vec{x})}{2\pi}\varepsilon^{ij}\left[E_j - \partial_j \int \mathrm{d}^2x'\, U_0(\vec{x}-\vec{x}')j^0(\vec{x}')\right] - \tfrac{v_\text{d}}{2\pi}\sum_{a=1}^{M}\delta(\vec{x} \text{ on } C_a)\partial_c^{-1}E_x^{\text{eff}}(\vec{e}_{\|}^{\,a})_i \end{aligned}$$

where the vector $\vec{e}_{\|}^{\,a}$ is tangent to the contour C_a and points in the positive direction. Again it is easy to check that $\partial_\mu j^\mu = 0$, i.e. that current conservation is respected. The edge currents are given by

$$\begin{aligned} j^0_{\text{edge}}(C_a) &= -\tfrac{i}{2\pi}s_a\partial_c^{-1}E_x^{\text{eff}} \qquad (4.5.6) \\ j^x_{\text{edge}}(C_a) &= -is_a v_\text{d} \cdot j^0_{\text{edge}}(C_a). \qquad (4.5.7) \end{aligned}$$

The edge anomaly applies to each bulk orbital and edge state separately,

$$\partial_\mu j^\mu_{\text{edge}}(C_a) = -\tfrac{i}{2\pi}s_a\left[E_x - \partial_x \int \mathrm{d}^2x'\, U_0(x,\vec{x}')j^0(\vec{x}')\right]. \qquad (4.5.8)$$

As expected, the sign s_a determines whether charge is transported into an edge or from an edge into the bulk. By repeating Laughlin's flux argument it is now demonstrated explicitly that the localised bulk orbitals do not affect the transport of charge from one sample edge to the other, independent of the electron-electron interactions; taking (4.5.8) and performing the contour integral over C_a we obtain the charge transported per unit of time from the a'th channel,

$$\mathrm{d}Q_a/\mathrm{d}\tau = -i\oint_{C_a} \partial_\mu j^\mu_{\text{edge}} = s_a \tfrac{1}{2\pi} d\Phi_a/\mathrm{d}\tau \tag{4.5.9}$$

where Φ_a is the magnetic flux enclosed by C_a. For $a\!>\!m$ this flux is obviously zero, since the localised bulk orbitals do not encircle the hole in the sample. This, then, shows that the Hall conductance is quantised (equal to m) independent of ε.

4.5.3 Continuous tunnelling exponent

Hij collaboreert met experimentatoren.
– een Leids theoreticus

Laughlin's flux argument for the Hall conductance expresses the quantum Hall state as an exact 'excited' state of the system. Tunnelling processes into the edge, on the other hand, are expressed in terms of eigenstates near the Fermi energy, i.e. the tunnelling d.o.s., and due to the Coulomb interactions this quantity is sensitive to the presence of bulk orbitals. We start our calculation of the tunnelling d.o.s. from the action (4.5.1), omitting the replica indices for notational simplicity and putting $A_\mu\!=\!0$,

$$\begin{aligned} S &= -\tfrac{i}{4\pi}\int\!\mathrm{d}\tau \sum_{j=1}^{M} s_j \oint_{C_j} \partial_x\varphi(\partial_0\varphi - i s_j v_\mathrm{d}\partial_x\varphi) \\ &\quad -\tfrac{1}{8\pi^2}\int\!\mathrm{d}\tau \sum_{j,j'=1}^{M} s_j s_{j'} \oint_{C_j}\oint_{C_{j'}} \partial_x\varphi U_0(x,x')\partial_{x'}\varphi. \end{aligned} \tag{4.5.10}$$

Following section 4.4.5, Eq. (4.4.29), the one particle Green's function can be written as follows

$$G(\tau_2-\tau_1) = \big\langle \exp -i[\varphi(\tau_2,x_0) - \varphi(\tau_1,x_0)]\big\rangle \tag{4.5.11}$$

where x_0 denotes a point on the edge contour C_1. The presence of the Coulomb interactions makes the computation of G a complicated two dimensional problem. Some procedure needs to be found that extracts the lowest energy excitations from (4.5.10). We follow the strategy of interpreting the boson fields as one two dimensional field variable and then collecting the terms with smallest momenta.

Gradient expansion

The interaction term in (4.5.10) can be written as a sum over area integrals,

$$-\tfrac{1}{8\pi^2}\int\!\mathrm{d}\tau \sum_{j,j'=1}^{M} \int_{C_j}\mathrm{d}^2x \int_{C_{j'}}\mathrm{d}^2x'\; s_j s_{j'}\; \partial_a\varphi(\vec{x})U_{ab}(\vec{x}-\vec{x}')\partial_b\varphi(\vec{x}') \tag{4.5.12}$$

with

$$U_{ab}(\vec{x}-\vec{x}') := \varepsilon_{ac}\varepsilon_{bd}\tfrac{\partial}{\partial x_c}\tfrac{\partial}{\partial x'_d}U_0(\vec{x}-\vec{x}'). \tag{4.5.13}$$

Since we are only interested in the φ with the smallest momenta, we can make the replacement

$$\sum_{j=m+1}^{M} \int_{C_j} \mathrm{d}^2x \ \to \Omega_f \int_{C_b} \mathrm{d}^2x. \tag{4.5.14}$$

The Ω_f stands for the fraction of the total area that is enclosed by all the bulk orbitals together. The contour C_b is not sharply defined and is located somewhere close to the edge (see Fig. 4.2b). It encloses the region within which the bulk orbitals are contained. The joint Coulomb effects of the bulk orbitals will effectively be comprised on this contour. For the terms in (4.5.10) containing $\partial_x\varphi\partial_0\varphi$ we can write

$$\sum_{j>m} \oint_{C_j} \partial_x\varphi\partial_0\varphi = \sum_{j>m} \int_{C_j} \mathrm{d}^2x \ \nabla\times(\nabla\varphi\partial_0\varphi) \to \Omega_f \int_{C_b} \mathrm{d}^2x \ \nabla\times(\nabla\varphi\partial_0\varphi) = \Omega_f \oint_{C_b} \partial_x\varphi\partial_0\varphi. \tag{4.5.15}$$

The expression $\sum_{j>m} \oint_{C_j} (\partial_x\varphi)^2$ averages out to $\kappa \int_{C_b} \mathrm{d}^2x \ (\nabla\varphi)^2$ with κ some positive constant related to the total length of all the bulk contours. If there are substantial stretches where a bulk orbital runs along the edge, interaction terms will arise, leading to a term $\oint_{C_b}(\partial_x\varphi)^2$.

Note that in doing the replacement (4.5.14) in (4.5.12), one also needs to introduce correction terms that compensate for the errors made when the separation $|\vec{x}-\vec{x}'|$ is 'small' (of the order of the average size of the orbitals or less) and U_{ab} does not vary slowly. These corrections are of the form $\int(\nabla\varphi)^2$. Then there are also extra correction terms that will arise if there are regions where a bulk orbital runs along the edge. This correction takes the form of a short-ranged interaction between C_b and all the other contours (including C_b). Having done the replacement (4.5.14) and writing the interaction terms again as contour integrals, we have the following action,

$$\begin{aligned} S \ = \ & -\tfrac{i}{4\pi} \int \mathrm{d}\tau \left[\sum_{j=1}^{m} \oint_{C_j} \partial_x\varphi\partial_-\varphi - \varepsilon \oint_{C_b} \partial_x\varphi\partial_0\varphi \right] \\ & -\tfrac{1}{8\pi^2} \int \mathrm{d}\tau \left[\sum_{jj'=1}^{m} \oint_{C_j} \mathrm{d}x \oint_{C_{j'}} \mathrm{d}x' \ \partial_x\varphi U_0 \partial_{x'}\varphi + \varepsilon^2 \oint_{C_b} \mathrm{d}x\mathrm{d}x' \ \partial_x\varphi(U+V_b)\partial_{x'}\varphi \right. \\ & \left. -2\varepsilon \sum_{j=1}^{m} \oint_{C_j} \mathrm{d}x \oint_{C_b} \mathrm{d}x' \ \partial_x\varphi(U_0+V_j)\partial_{x'}\varphi \right] - g \int \mathrm{d}\tau \int_{C_b} \mathrm{d}^2x \ (\nabla\varphi)^2, \end{aligned} \tag{4.5.16}$$

where g is a positive constant. We have identified Ω_f with ε, since the fraction of the area occupied by bulk states is exactly the deviation from integer filling. We have written $V_b(x,x')$ for the short-ranged interaction between two points on C_b; The $V_j(x,x')$ denotes the short-ranged interaction between a point x on C_j and x' on C_b. The precise expression for V is unknown due to the fact that it has its origin in the twilight zone near the edge, where it is unclear whether a term contributes to the bulk or edge action.

Comparing this result (4.5.16) with (4.5.10), we see that the presence of the interacting bulk states effectively leads to the appearance of an additional (anti)chiral boson on the contour C_b, an extra short-ranged interaction with this contour, and a left over bulk term $\int(\nabla\varphi)^2$.

Effect of the bulk term

In order to be able to calculate the tunnelling d.o.s. (4.5.11) we need an effective theory for the edge degrees of freedom, and therefore we have to understand how they are affected by the left over bulk term. To this end, we are going to split bulk and edge degrees of freedom. We write the bulk term as $\int_{C_b}(\nabla\Phi)^2$, where Φ represents the bulk degrees of freedom and is treated as an integration variable independent of φ. To reflect the fact that it is actually an extension of φ into the bulk, we impose some boundary condition on Φ, for instance $\Phi|_{\rm edge} = \varphi$ or $\partial_\perp\Phi|_{\rm edge} = \partial_\perp\varphi$ ($\partial_\perp$ is the derivative perpendicular to the contour). The effect of the bulk term on the edge theory is obtained by integrating out Φ, which leads to an effective action for the boundary conditions. Let us consider a general scenario and impose the boundary conditions $\Phi|_{\rm edge} = \psi_0$ and $\partial_\perp\Phi|_{\rm edge} = \psi_1$, using constraint multipliers k_0 and k_1, respectively.

$$\begin{aligned} e^{S_{\rm eff}[\psi_0(x),\psi_1(x)]} &= \int\mathcal{D}\Phi(\vec{x})\ \mathcal{D}k_0(x)\ \mathcal{D}k_1(x)\ \times \qquad (4.5.17)\\ &\times\ \ \exp\left\{i\oint k_0(\Phi-\psi_0)+i\oint k_1(\partial_\perp\Phi-\psi_1)-g\int {\rm d}^2x\ (\nabla\Phi)^2\right\}. \end{aligned}$$

For notational simplicity we have omitted time dependence and the subscript C_b under all the integrals. We first wish to integrate (4.5.17) over $\Phi(\vec{x})$ keeping k_0 and k_1 fixed. For this purpose we split Φ, which has free boundary values, into a bulk and an edge part by writing

$$\Phi = \Phi_L + \hat\Phi \qquad \partial_\perp\Phi_L|_{\rm edge} = \partial_\perp\Phi|_{\rm edge} \qquad \partial_\perp\hat\Phi|_{\rm edge} = 0 \qquad (4.5.18)$$

where Φ_L satisfies Laplace's equation

$$\nabla^2\Phi_L(\vec{x}) = 0. \qquad (4.5.19)$$

The $\Phi_L(\vec{x})$ is completely determined by $\partial_\perp\Phi_L$ on the edge, which we now take as an independent edge degree of freedom denoted by $E_1(x)$. Introducing the 2D Green's function G,

$$G(\vec{x},\vec{x}') = \tfrac{1}{2\pi}\ln|\vec{x}-\vec{x}'| \qquad ; \qquad \nabla^2G(\vec{x},\vec{x}') = \delta(\vec{x}-\vec{x}'), \qquad (4.5.20)$$

and using Green's theorem, we solve Laplace's equation and obtain for $\Phi_L(\vec{x})$

$$\Phi_L(\vec{x}) = -\oint{\rm d}x'\ \left[G(\vec{x},x')E_1(x') - \Phi_L(x')\tfrac{\partial G}{\partial y'}(x,y;x',0)\right]. \qquad (4.5.21)$$

This expression tells us that we need to now Φ_L on the edge in order to evaluate Φ_L in the bulk. Luckily, we do not need the full 2D $\vec{x}$ dependence, since due to the splitting (4.5.18) Φ_L will get evaluated at the edge only. Using a special property of the Green's function (4.5.20), namely $[\partial_{y'}G](x,0;x',0) = 0$, we can explicitly write Φ_L on the edge as a function of E_1,

$$\Phi_L(x) = -\oint{\rm d}x'\ G(x,x')E_1(x'). \qquad (4.5.22)$$

The action, written in terms of $\hat\Phi$ and E_1, is now given by

$$S = -g\int(\nabla\hat\Phi)^2 - g\oint E_1GE_1 + \oint E_1(2g\hat\Phi - iGk_0) + i\oint k_0(\hat\Phi-\psi_0) + i\oint k_1(E_1-\psi_1) \qquad (4.5.23)$$

where we have used the shorthand notation $\oint AGB$ for $\oint \mathrm{d}x \oint \mathrm{d}x' A(x)G(x,x')B(x')$. Integrating out $\hat{\Phi}$ is now simply done by replacing $\hat{\Phi}$ by its saddle point value. Varying the action with respect to $\hat{\Phi}$, keeping E_1 fixed, we get the saddlepoint equation

$$\nabla^2\hat{\Phi} + \delta(y)[E_1 + \tfrac{i}{2g}k_0] = 0. \tag{4.5.24}$$

Using the Green's function's property $[\partial_{y'}G](x,0;x',0) = 0$ again, we find the following solution on the edge

$$\hat{\Phi}(x) = -\oint \mathrm{d}x'\; G(x,x')[E_1 + \tfrac{i}{2g}k_0](x'). \tag{4.5.25}$$

In substituting this solution into (4.5.23) we do not need the full 2D $\vec{x}$-dependence of $\hat{\Phi}(\vec{x})$, since we can write $\int(\nabla\hat{\Phi})^2 = -\int \hat{\Phi}\nabla^2\hat{\Phi}$ and $\nabla^2\hat{\Phi}$ is an expression restricted to the edge. Substitution of (4.5.25) into (4.5.23) yields

$$S = -2g \oint E_1 G E_1 + \tfrac{1}{4g}\oint k_0 G k_0 - i\oint k_0(\psi_0 + 2GE_1) + i\oint k_1(E_1 - \psi_1). \tag{4.5.26}$$

Integrating out k_0 is straightforward and gives

$$S = g\oint \big(\psi_0 G^{-1}\psi_0 + 4\psi_0 E_1 + 2E_1 G E_1\big) + i\oint k_1(E_1 - \psi_1). \tag{4.5.27}$$

In the end we integrate out k_1, yielding the constraint $E_1 = \psi_1$. The final result for $S_{\text{eff}}[\psi_0,\psi_1]$ becomes

$$\begin{aligned} S_{\text{eff}}[\psi_0,\psi_1] &= g\oint \big(\psi_0 G^{-1}\psi_0 + 4\psi_0\psi_1 + 2\psi_1 G\psi_1\big) \\ &= g\oint (\psi_0,\psi_1)\begin{pmatrix} G^{-1} & 2 \\ 2 & 2G \end{pmatrix}\begin{pmatrix}\psi_0 \\ \psi_1\end{pmatrix}. \end{aligned} \tag{4.5.28}$$

We are going to put $\psi_0 = 0$ in order to avoid double counting of $(\partial_x\varphi)^2$ terms at the edge, and we put $\psi_1 = \partial_\perp\varphi$. The action (4.5.28) becomes

$$S[\partial_\perp\varphi] = 2g\oint \partial_\perp\varphi\; G\; \partial_\perp\varphi. \tag{4.5.29}$$

This edge term, derived from the interaction with the bulk orbitals, is seriously going to affect the tunnelling exponent. A quick way to see this is as follows: on the contours $C_1,\cdots,C_b$, the field $\varphi(\vec{x})$ can be written as $\varphi(x,y \text{ on } C_1)+$ perpendicular derivatives. For the tunnelling exponent, only $\varphi|_{C_1}$ is needed, so we can integrate out the perpendicular derivatives in (4.5.16 minus bulk term+4.5.29) to obtain an effective action for φ on C_1. The dominant part of the 1D propagator for $\partial_\perp\varphi$ is given by $G^{-1}(k) \propto |k|$, from which it follows that all terms introduced by the integration over $\partial_\perp\varphi$ are irrelevant. Higher powers of $\partial_\perp$ are even less relevant. Replacing all the φ in (4.5.16) by $\varphi|_{C_1}$, we get a term $\nu\oint\partial_x\varphi\partial_0\varphi$, leading to a tunnelling exponent $S = 1/\nu$ instead of the free particle result $S = 1$.

In what follows we are going to do this analysis more formally, based on a consideration of the neutral modes in the theory where the edge channels are not spatially separated.

Demise of the neutral modes

Zo, jij bent nu in het terminale stadium?
– een Amsterdams theoreticus

In the large wavelength limit, the contours $C_1, \cdots, C_b$ are lying so close together that we can effectively return to the picture where all the edge channels are sitting on top of each other. We label the channels $\varphi_1(x), \cdots, \varphi_m(x), \varphi_b(x)$. Let us for simplicity's sake first consider the case $\nu = 1-\varepsilon$, where we just have the two fields φ_1 and φ_b. In terms of these fields, the action (4.5.16), without the bulk term and the bulk effect (4.5.29), takes the form (again using abbreviated notation)

$$\begin{aligned} S_0[\varphi_1, \varphi_b] &= -\tfrac{1}{4\pi} \oint [\partial_x \varphi_1 \partial_0 \varphi_1 - \varepsilon \partial_x \varphi_b \partial_0 \varphi_b] \\ &\quad - \tfrac{1}{8\pi^2} \oint U_0 [\partial_x \varphi_1 - \varepsilon \partial_x \varphi_b]^2 - \tfrac{1}{8\pi^2} \sum_{k,l=1,b} \oint V_{kl} \partial_x \varphi_k \partial_x \varphi_l. \end{aligned} \tag{4.5.30}$$

We have put all the short range contributions into the $(m+1)\times(m+1)$ velocity matrix V. We next define a 'charged mode' Γ and a 'neutral mode' γ in such a way that only the charged mode feels the long-ranged part of the interaction,

$$\begin{aligned} \Gamma = \tfrac{1}{\nu}(\varphi_1 - \varepsilon \varphi_b) \quad &; \quad \gamma = \varphi_1 - \varphi_b \\ \varphi_1 = \Gamma - \tfrac{\varepsilon}{\nu}\gamma \quad &; \quad \varphi_b = \Gamma - \tfrac{1}{\nu}\gamma. \end{aligned} \tag{4.5.31}$$

In the basis (Γ, γ) the action (4.5.30) becomes

$$S_0[\Gamma, \gamma] = -\tfrac{1}{4\pi} \oint \left[\nu \partial_x \Gamma \partial_0 \Gamma - \tfrac{\varepsilon}{\nu} \partial_x \gamma \partial_0 \gamma\right] - \tfrac{\nu^2}{8\pi^2} \oint U_0 (\partial_x \Gamma)^2 - \tfrac{1}{8\pi^2} \oint [\partial_x \Gamma \;\; \partial_x \gamma] \hat{V} \begin{bmatrix} \partial_x \Gamma \\ \partial_x \gamma \end{bmatrix}$$

where $\hat{V}$ is the velocity matrix in the new basis. The expression $\partial_\perp \varphi$ is evidently equivalent to the neutral mode $\gamma \propto \varphi_b - \varphi_1$. The leftover bulk contribution (4.5.29) therefore translates into an extra term involving the neutral mode,

$$S_{\text{bulk}}[\gamma] = \text{constant} \cdot \oint \gamma \; G \; \gamma. \tag{4.5.32}$$

The tunnelling d.o.s. is now expressed as

$$\left\langle \exp -i\varphi_1 |_{\tau_1}^{\tau_2} \right\rangle \propto \int \mathcal{D}\Gamma \; \mathcal{D}\gamma \; \exp\Big(-i(\Gamma - \tfrac{\varepsilon}{\nu}\gamma)|_{\tau_1}^{\tau_2} + S_0[\Gamma, \gamma] + S_{\text{bulk}}[\gamma]\Big). \tag{4.5.33}$$

If we perform the integration over γ first, we see that the 'bulk' part of the action yields the following contribution to the inverse propagator: $G(k) \propto 1/|k|$, which is dominant at low momenta. The integration over γ yields Γ-Γ terms of order $k^5 \hat{V}(k)$. These are clearly irrelevant. For the tunnelling d.o.s. we can write

$$\begin{aligned} \left\langle \exp -i\varphi_1 |_{\tau_1}^{\tau_2} \right\rangle &\propto \int \mathcal{D}\Gamma \; \exp\big(-i\Gamma|_{\tau_1}^{\tau_2} + S_{\text{eff}}[\Gamma]\big) \\ S_{\text{eff}}[\Gamma] &= -\tfrac{i\nu}{4\pi} \oint \partial_x \Gamma \partial_0 \Gamma - \tfrac{\nu^2}{8\pi^2} \oint \partial_x \Gamma (U_0 + \nu^{-2} \hat{V}_{\Gamma\Gamma}) \partial_x \Gamma. \end{aligned} \tag{4.5.34}$$

For small momenta the $\hat{V}$ essentially reduces to a constant and we can use the results of appendix 4A, obtaining

$$\left\langle \exp -i\varphi_1 |_{\tau_1}^{\tau_2} \right\rangle \propto (\tau_2 - \tau_1)^{-S} \quad ; \quad S = 1/\nu. \tag{4.5.35}$$

The general case $\nu = m - \varepsilon$

The results for $\nu = 1-\varepsilon$ are easily generalised. From the 'bulk' channel φ_b and the edge channels $\varphi_1, \cdots, \varphi_m$ we construct a charged mode γ_0 and m neutral modes $\gamma_1, \cdots, \gamma_m$ as follows,

$$\begin{aligned} \gamma_0 &= \tfrac{1}{\nu}\Big(\sum_{k=1}^{m} \varphi_k - \varepsilon\varphi_b\Big) \\ \gamma_a &= \tfrac{1}{a}\Big(\sum_{k=1}^{a} \varphi_k - a\varphi_{a+1}\Big) \qquad a = 1, \cdots, m \end{aligned} \tag{4.5.36}$$

where we define φ_{m+1} as φ_b. The neutral modes $\gamma_1, \cdots, \gamma_{m-1}$ are the usual ones for a theory with m edges. They are mutually perpendicular and normal to the charged mode. The additional γ_m is normal to the other neutral modes but not to the charged mode. The φ's are expressed in terms of the γ's as follows

$$\begin{aligned} \varphi_b &= \gamma_0 - \tfrac{m}{\nu}\gamma_m \\ \varphi_k &= \gamma_0 - \tfrac{\varepsilon}{\nu}\gamma_m - (1-\tfrac{1}{k})\gamma_{k-1} + \sum_{a=k}^{m-1} \tfrac{1}{a+1}\gamma_a \qquad k \leq m. \end{aligned} \tag{4.5.37}$$

Equation (4.5.30) is generalised to

$$\begin{aligned} S[\varphi] &= -\tfrac{1}{4\pi} \oint \left[\sum_{j=1}^{m} \partial_x\varphi_j\partial_0\varphi_j - \varepsilon\partial_x\varphi_b\partial_0\varphi_b \right] \qquad (4.5.38) \\ &\quad -\tfrac{1}{8\pi^2} \oint U_0 \left[\sum_{j=1}^{m} \partial_x\varphi_j - \varepsilon\partial_x\varphi_b \right]^2 - \tfrac{1}{8\pi^2} \sum_{k,l=1}^{m+1} \oint V_{kl}\partial_x\varphi_k\partial_x\varphi_l. \end{aligned}$$

Again, all the short-ranged contributions have been put into a velocity matrix V, which now has dimension $(m+1)\times(m+1)$. Writing (4.5.38) in terms of the γ-basis, we get

$$\begin{aligned} S[\gamma] &= -\tfrac{1}{4\pi} \oint \left[\nu\partial_x\gamma_0\partial_0\gamma_0 + \sum_{a=1}^{m-1} \tfrac{a}{a+1}\partial_x\gamma_a\partial_o\gamma_a - m\tfrac{\varepsilon}{\nu}\partial_x\gamma_m\partial_0\gamma_m \right] \\ &\quad -\tfrac{\nu^2}{8\pi^2} \oint U_0(\partial_x\gamma_0)^2 - \tfrac{1}{8\pi^2} \sum_{a,c=0}^{m} \oint \hat{V}_{ac}\partial_x\gamma_a\partial_x\gamma_c \end{aligned} \tag{4.5.39}$$

where $\hat{V}$ is the velocity matrix in the basis of γ's. The argument of (4.5.32 to 4.5.35) can be applied again, in a slightly modified form; the neutral modes are equivalent to $\partial_\perp\varphi$ and higher derivatives. (A basis $\hat{\gamma}$ can be found for the neutral modes in which $\hat{\gamma}_n$ corresponds to the 1D lattice discretisation of $\partial_\perp^n\varphi$). On dimensional grounds the propagator for the n'th normal derivative of φ has to be proportional to k^{2n-1}, leading to irrelevant contributions. A more concrete way of making this statement would be to generalise the analysis presented in (4.5.17 to 4.5.28), including boundary conditions for the higher normal derivatives. However, that would also require us to take into account higher order terms in the φ-theory (4.5.16). The resulting effective action for the charged mode γ_0 is of the form (4.5.34), with $\nu = m-\varepsilon$.

We can summarise the results of section 4.5 as follows: We have seen that the Fermi liquid result $S=1$ is obtained for the tunnelling d.o.s. (I) when the Coulomb interactions are omitted, or (II) when interactions are included but only short length scales are considered. An interacting theory for the lowest lying excitations, which are slowly varying field configurations, yields completely different results. The presence of bulk orbitals, interacting mutually and with the edge states, is effectively described by an extra edge channel with prefactor $-\varepsilon$ plus a remnant of the interactions in the bulk of the form $\int(\nabla\varphi)^2$. The leftover bulk term serves to make all the neutral edge modes irrelevant, yielding an effective edge action for the one remaining, charged, mode. Due to the presence of the extra 'bulk' channel, the prefactor of this effective action $S[\Gamma]$ becomes $m-\varepsilon=\nu$, which is a continuous parameter in sharp contrast to the integer quantised m.

4.6 Plateau transitions

In this section we show how the notion of critical edge states can be used in order to gain insight into 'long range potential fluctuations'. This problem, which is very difficult to handle within the nonlinear sigma model method, plays an important role experimentally. For instance, it has been stressed that the plateau transitions as observed in the experiments of H.P. Wei *et al.* [76] are very difficult to observe in general, due to the presence of slowly varying potential fluctuations.

A slowly varying potential is the generic type of disorder in the standard GaAs heterostructures, which has historically led to semiclassical considerations (percolation picture) of delocalisation near the Landau band center [57]. It is important to recognise that also our critical system (4.2.3) is very sensitive to the presence of smooth potentials (or 'inhomogeneities') in the sample. For example, the critical magnetic field B^* may be slowly varying throughout the system due to inhomogeneities in the electron density. This means that the scaling result is valid only up to a certain fixed value for L. Beyond this value the remaining 'extended' states in the problem may be confined to the equipotential contours of the inhomogeneity potential, quite similar to the semiclassical picture of percolation.

It is generally difficult to obtain detailed knowledge on the various length- and energy scales that are involved in the cross-over problem between percolation and localisation. In what follows, we present the simplest possible scenario for crossover that enables us to deal simultaneously with interactions and such basic concepts as mean field theory and 'universality' of the plateau transition.

4.6.1 Percolation

In order to fix the thought, we imagine the equipotential contours near half filling to form a large cluster (Fig. 4.3) [13]. Since the disconnected, closed contours do not contribute to the transport, we focus our attention to an infinite backbone cluster which we take as a regular 2D array of 'saddlepoints' and we disregard all the loose hanging, finite pieces (Fig. 4.3b). The saddlepoints (the sites of the square lattice) are connected to one another by the disordered 1D chiral edge channels (links on the lattice). This network can alternatively be looked upon as a checkerboard with filling fractions alternating between the values $\nu=0$ and $\nu=1$. The kinetic part of the action for this system may be written in the form of (4.2.4)

$$S[Q] = \tfrac{1}{8}\int \mathrm{d}^2x\; m(\vec{x})\mathrm{tr}\;\varepsilon_{ij}Q\partial_i Q\partial_j Q \tag{4.6.1}$$

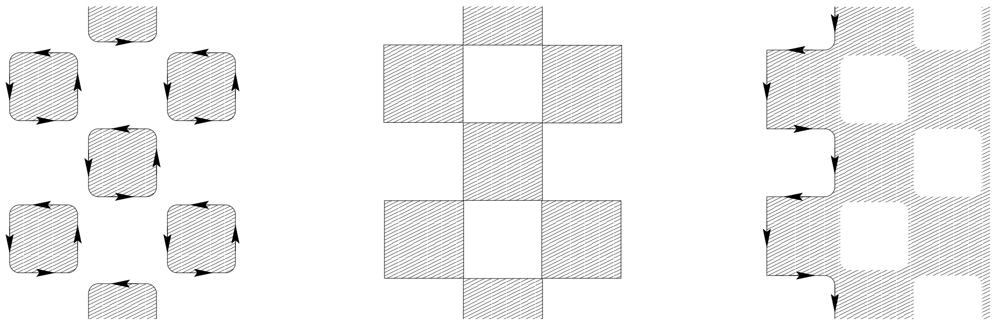

Figure 4.3: *Backbone cluster as a network of saddlepoints. Shaded areas have $\nu=1$, white areas $\nu=0$. The arrows indicate the direction of the currents.* (a) *Less than half filling*; (b) *Exactly half filling*; (c) *Filling fraction larger than one half.*

with $m(\vec{x})=0,1$ (Fig. 4.3b). Using the parametrisation of (4.2.6), the action can also be written in the form (4.2.7) which is now solely defined on the links of the square lattice,

$$S[Q] = 2\pi i \cdot q[U] + \tfrac{1}{2}\sum_i \oint_i \mathrm{d}x \;\mathrm{tr}\;(\Lambda t \partial_x t^{-1}) + \pi\eta\rho_{\mathrm{link}} \sum_i \oint_i \mathrm{d}x \;\mathrm{tr}\;\Lambda Q. \tag{4.6.2}$$

Here, the sum is over all the black squares and the integrals are over the contours of the black squares. Despite the fact that this action does not contain any dissipative (σ_{xx}) terms, it is easy enough to show that in the long distance limit, (4.6.1) reduces to the form of the sigma model action (4.2.2) with

$$\sigma^0_{\mathrm{xx}} = 1/2 \qquad ; \qquad \sigma^0_{\mathrm{xy}} = 1/2 \tag{4.6.3}$$

The reason for this is contained in the fact that the saddlepoints act like scattering centers which render the system dissipative at large distances. In order to demonstrate this, all one needs to do is to follow up on (4.2.20) where the background field t_0 now represents the 'slow modes' that are kept. The t field variables are the 'fast modes' which contain all the wavelengths smaller than the lattice constant, i.e. the average distance between the saddlepoints, and which are integrated out. This leads to an effective action for each link according to

$$\begin{aligned} S_{\mathrm{link}}[t_0] &= \tfrac{1}{2}\int_{\mathrm{link}} \mathrm{d}x \;\mathrm{tr}\;(\langle Q\rangle t_0 \partial_x t_0^{-1}) + \tfrac{1}{8}\left\langle \left[\int_{\mathrm{link}} \mathrm{d}x \;\mathrm{tr}\;(Q t_0 \partial_x t_0^{-1})\right]^2 \right\rangle_{\mathrm{cum}} \\ &= \tfrac{1}{2}\int_{\mathrm{link}} \mathrm{d}x \;\mathrm{tr}\;(\Lambda t_0 \partial_x t_0^{-1}) - \tfrac{1}{8}\sigma^0_{\mathrm{xx}} \int_{\mathrm{link}} \mathrm{d}x \;\mathrm{tr}\;(\partial_x Q_0)^2 \end{aligned} \tag{4.6.4}$$

where $Q_0 = t_0^{-1}\Lambda t_0$ and $\sigma^0_{\mathrm{xx}} = L_0/2$ is the 1D conductivity of a single channel of length L_0, a well-known result in the theory of pure metals. The expectation is with respect to the theory (4.2.7) with $m=1$. The subscript 'cum' indicates that only connected diagrams are taken. Use has been made of (4.2.18), (4.2.21), (4.2.22) with $\eta=0$ and (4.2.24). Next, by taking the sum over all links one can absorb the factor L_0 into the definition of a 2D integral,

$$-\tfrac{1}{16}\sum_{\mathrm{links}} L_0 \int_{\mathrm{link}} \mathrm{d}x \;\mathrm{tr}\;(\partial_x Q_0)^2 \longrightarrow -\tfrac{1}{16}\int \mathrm{d}^2x \;\mathrm{tr}\;(\nabla Q_0)^2. \tag{4.6.5}$$

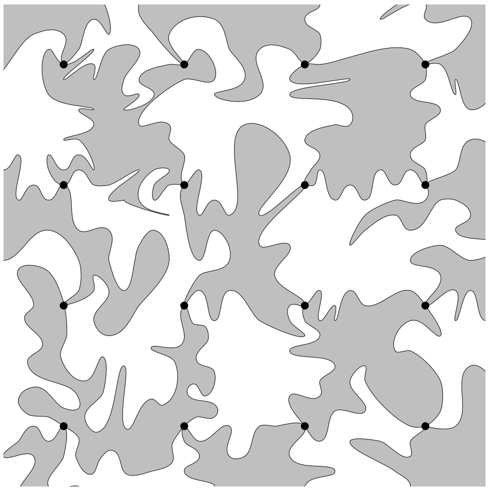

Figure 4.4: *Backbone cluster as in Fig.* 4.3b, *but with highly ramified contours between saddle points* (•).

Here we only used the fact that the Q_0 field variable varies slowly over a distance L_0. The first term in (4.6.4) can be handled in a similar way. For instance, it can be rewritten in the form of (4.6.1) with Q replaced by Q_0, which is then followed by taking the continuum limit according to

$$\begin{aligned}\tfrac{1}{2}\sum_{\text{links}}\int_{\text{link}} \mathrm{d}x \ \mathrm{tr}\ (\Lambda t_0 \partial_x t_0^{-1}) \quad &\longrightarrow \quad \tfrac{1}{8}\int \mathrm{d}^2 x\ m(\vec{x})\varepsilon_{ij}\mathrm{tr}\ Q_0\partial_i Q_0 \partial_j Q_0 \\ &\longrightarrow \quad \tfrac{1}{16}\int \mathrm{d}^2 x\ \varepsilon_{ij}\mathrm{tr}\ Q_0\partial_i Q_0 \partial_j Q_0. \qquad (4.6.6)\end{aligned}$$

The result of (4.6.4–4.6.6) is identical to the statement made in (4.6.3). Notice that (4.6.3) is precisely the point where we expect the σ model action (4.2.2) to have a critical phase. Hence, we have established a direct connection between critical 1D edge states on the one hand and the 2D delocalisation transition of the band center on the other. This connection has the following ingredients:

1. The infinite percolation cluster at the band center contains a finite density of saddlepoints. This translates into a finite density of scattering centers which, in turn, is responsible for making the sample diffusive (dissipative) at large distances.

2. The parameters $\sigma^0_{\mathrm{xx}}, \sigma^0_{\mathrm{xy}}$ (4.6.3) constitute a mean field theory of the conductances which is valid for length scales L_0.

Without going into further detail we mention that the analysis can easily be generalised to more complicated situations. For example, the links between the saddlepoints need not be straight lines. They can be taken as arbitrarily complex, non-intersecting paths reflecting the highly ramified percolation contours (Fig. 4.4). The same result (4.6.3) applies to all cases, indicating that the general result $\sigma^0_{xx} = 1/2$ actually stands for the quantised conductance in one dimension.

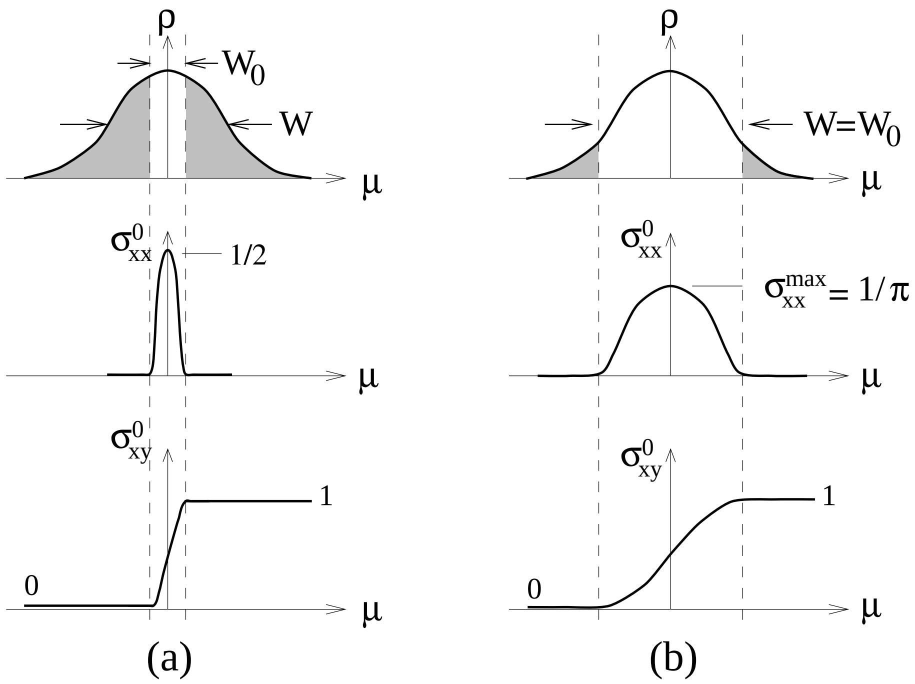

Figure 4.5: *Mean field theory for the lowest Landau level, with varying chemical potential* μ. (a) *Smooth long range disorder.* (b) *Short range disorder.*

4.6.2 Mean field theory

Next, we wish to extend our mean field analysis (4.6.3) to include also the energies away from the Landau band center. For this purpose we have to relate the range in energy W_0 within which the equipotential contours form an infinite saddlepoint cluster to the total bandwith W of the Landau band. It is understood that the phrase 'saddlepoint' actually stands for those special points where two equipotential contours approach each other at a distance of the order of the magnetic length ℓ_0 or smaller. By assuming a simple quadratic form for the potential near saddle points we obtain the following estimate,

$$W_0 \approx (\ell_0/\lambda)^2 W \tag{4.6.7}$$

where λ is the characteristic correlation length of the random potential, which we have taken to be much larger than ℓ_0, and W equals the amplitude of the potential fluctuations. The sigma model theory or, equivalently, the scaling theory of localisation only applies to the energy band W_0 about the band center. For energies just outside W_0 the network of saddlepoints is broken up into disconnected islands of size $L_0 \times L_0$ (Fig. 4.3a and c). The absence of any quantum tunnelling means that no correlation exists between the islands (they are represented by independent actions as long as one works within the free electron approach). In the language of the σ model, the situation is represented by putting $\sigma_{\rm xx} = 0$ but $\sigma_{\rm xy} = m =$integer. The latter follows from the long-ranged correlations which still exist near the edge and which can be expressed in terms of an integer number m of edge channels. In Fig. 4.5a we illustrate the behaviour of the d.o.s. ρ and the conductances σ^0_{ij} as a function of energy μ at zero temperature. The sigma model conductance parameters σ^0_{ij} can be expressed as a function of the dimensionless quantity $\triangle\mu/W_0$

$$\sigma^0_{ij} = f_{ij}(\triangle\mu/W_0) \tag{4.6.8}$$

where $\triangle\mu$ is the energy relative to the Landau band center. The f_{ij} are non-universal and generally depend on the microscopic details of the randomness. For comparison we have plotted the results of the more familiar theory of short-ranged scatterers (self-consistent Born approximation) in Fig. 4.5b. In this case, there is only a small difference between W_0 and W due to the localised states in the Gaussian tails of the Landau band.

An estimate for L_0 can be obtained as follows. Let $|\triangle\mu| \approx W_0$ denote the energies where the saddlepoint network breaks up into disconnected equipotential contours of size $L_0 \times L_0$ (Fig. 4.3a,c). According to the semiclassical picture of percolation we can relate the typical cluster size ξ_p to the energy $\triangle\mu$ according to

$$\xi_p \sim \lambda(\triangle\mu/W)^{-4/3} \tag{4.6.9}$$

where the critical index 4/3 is the exponent for semiclassical localisation. By identifying the points $|\triangle\mu| = W_0$ and $\xi_p = L_0$ in (4.6.9) we obtain the estimate

$$L_0 \approx \ell_0(\lambda/\ell_0)^{11/3} \qquad (\lambda \gg \ell_0) \tag{4.6.10}$$

or, more generally,

$$W_n \approx \frac{l_n}{\lambda} W \qquad ; \qquad L_n \approx \ell_n(\lambda/\ell_n)^{11/3} \qquad (\lambda \gg \ell_n). \tag{4.6.11}$$

The λ is an adjustable parameter in the theory and it ranges between microscopic distances ($\ell_0 \approx 100$Å) and infinity.

4.6.3 Interaction effects

It is quite possible that L_0 (4.6.10) is many times larger than the micron regime which is the typical scale for inelastic processes at low temperatures. This means that the critical behaviour (4.2.3) cannot be observed within the limitations of ordinary laboratory experiments. This, then, is the easiest and crudest explanation for the lack of scaling in many samples. As a first step toward a more quantitative understanding of transport at finite T, we come back to the distinction, made in the beginning, between the backbone cluster and the disconnected, 'loose hanging' pieces. Due to the electron-electron interactions, motion of the conducting electrons on the saddlepoint network is affected by the localised electrons. This may be expressed in terms of a *relaxation time* $\tau_{\rm in}$ which is a characteristic time for equilibration between the conducting and localised electrons. Later on in this chapter (section 4.6.5) we will address the problem of interaction effects and show that

$$1/\tau_{\rm in} = \beta_1 T + \beta_2 T^2 + \cdots \tag{4.6.12}$$

at low temperatures. This expression is determined by the collection of 'nearly saddlepoints' where quantum tunnelling is not possible but where the interactions between the conducting and localised particles are strongest nevertheless. The importance of 'nearly saddlepoints' can be seen by comparing the wavefunctions at different energies close to the Landau band center. What is a saddlepoint configuration at one energy may turn into a 'nearly saddlepoint' at another and vice versa. These abrupt changes in the configuration of the conducting network at slightly different energies blur the distinction between saddlepoints and 'nearly saddlepoint' configurations as far as finite temperatures are concerned. This means that the relaxation time $\tau_{\rm in}$ (4.6.12) determines an effective bandwidth

$W_{\rm eff} = W_0 + \tau_{\rm in}^{-1}$ of states that contribute to the conduction at finite temperatures. Eq. (4.6.8) is replaced by the expression

$$\sigma^0_{ij}(T) = f_{ij}(\triangle\mu/W_{\rm eff}) = f_{ij}(\triangle\mu/[W_0 + \tau_{\rm in}^{-1}]). \tag{4.6.13}$$

This result is a characteristic feature of long-ranged potential fluctuations and it does not occur in the problem of short-ranged scatterers. To conclude this section, we next estimate the range of validity of the result (4.6.13). We write

$$v_d \tau_{\rm in} = L_{\rm in} \quad , \quad v_d \approx 2\pi l_0^2 W/\lambda. \tag{4.6.14}$$

The $L_{\rm in}$ is the mean free path for drifting along the links of the lattice. We mentioned earlier already that the actual path between two saddle points is arbitrarily convoluted and very long. Let L_t denote the actual path length between saddle points, then the criterion for scaling is clearly given by

$$L_{\rm in} > L_t. \tag{4.6.15}$$

Next we use the ramification hypothesis [57] in order to relate L_t to the shortest distance between saddlepoints (L_0). We obtain

$$L_t \propto L_0^\sigma \tag{4.6.16}$$

with σ somewhere between 1 and 2. The criterion for scaling (4.6.15) now implies

$$\tau_{\rm in}^{-1} < (l_0/\lambda)^{8\sigma/3} W_0 \ll W_0. \tag{4.6.17}$$

This result indicates that (4.6.13) is very likely to be observed in samples with smooth disorder potentials.

The results of this section are consistent with the recently reported empirical fitting [71, 4] of the transport data in the quantum Hall regime. Since we are necessarily operating with an almost complete lack of knowledge on the microscopic details of sample disorder, it is conceivable that other types of inhomogeneity, especially those in low mobility samples, explain the same thing.

4.6.4 Independent Q-fields

The subjects of critical edge states and long-ranged disorder have left several conceptual questions that still need to be answered. For example, we have seen that short-ranged disorder causes inter-channel scattering between the chiral edge states. Since we do not expect inter-channel scattering to occur when the potential fluctuations are smooth (relative to the magnetic length), it is necessary to re-investigate the meaning of instanton vacuum theory for $\nu > 1$. Scattering between multiple edge states is avoided by writing, instead of (4.2.2)

$$\begin{aligned} S_{\rm eff}[Q^{(n)}] &= \sum_{n=0}^{\infty} \Big[-\tfrac{1}{8}\sigma_{xx}^{(n)} \int d^2x \text{ tr } [\nabla Q^{(n)}]^2 + \tfrac{1}{8}\sigma_{xy}^{(n)} \int d^2x \text{ tr } \varepsilon_{ij} Q^{(n)} \partial_i Q^{(n)} \partial_j Q^{(n)} \\ &\quad + \pi\rho^{(n)}\eta \int d^2x \text{ tr } \Lambda Q^{(n)} \Big] \end{aligned} \tag{4.6.18}$$

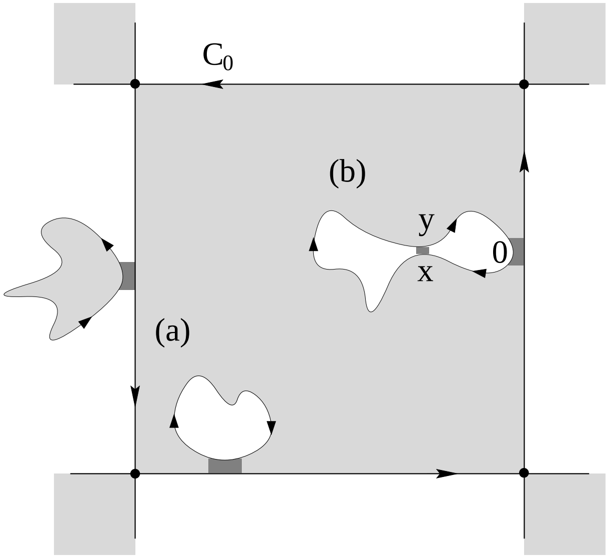

Figure 4.6: (a) *Point-like interaction between conducting electrons and the localised electrons at 'nearly saddlepoints'.* (b) *The coordinates* 0*,* x *and* y *are the points of interaction along the localised contour.*

where the sum runs over all the Landau levels n. The $Q^{(n)}$ stands for an independent field variable Q for each Landau level separately. The $\sigma_{ij}^{(n)}$ are the n'th Landau level contributions to the mean field conductances, which are now given by

$$\sigma_{ij} = \sum_{n=0}^{\infty} \sigma_{ij}^{(n)}. \tag{4.6.19}$$

The $\sigma_{ij}^{(n)}$ all have the same μ-dependence (Fig. 4.5a) except for an appropriate shift in energy. Since $0 \leq \sigma_{xy}^{(n)} \leq 1$ for each n, it is clear that (4.6.18) is the appropriate generalisation of the theory (section 4.2.1) to include filling fractions larger than one. The theories of (4.6.18) and (4.2.2) are identical as far as the critical behaviour of the plateau transitions is concerned. Eq. (4.6.18) cannot, however, be used in the limit of small magnetic field, where the Landau levels partly or completely overlap. The details of crossover require a separate analysis.

4.6.5 Computation of $\tau_{\rm in}$

En, wanneer ben jij uitgerekend?
– vraag aan een Leidse promovenda

We next return to the problem of the plateau transitions. Following section 4.6 we expect that the transport at high temperatures is dominated by interactions between the conducting electrons on the backbone saddlepoint network and those on the disconnected pieces or clusters.

The fundamental quantity to compute is the characteristic time $\tau_{\rm in}$ that is needed for the backbone electrons to equilibrate with the rest of the network. In order to set up a theory for relaxation, we consider the 'nearly saddlepoints' in the network, where tunnelling is not possible but where the Coulomb forces nevertheless produce 'sudden changes' in the motion of the conducting electrons. Fig. 4.6 illustrates the interaction of the saddlepoint network with disconnected orbitals. The 'nearly saddlepoints' where the Coulomb forces are most effective are indicated by the shaded areas. We can model the situation by introducing a delta-function potential which acts in the small areas of the nearly saddlepoints only. The action can be written as

$$S_{\text{eff}}[\varphi] = S[\varphi_0] + \sum_i S[\varphi_i] - \sum_i \int \mathrm{d}\tau \; \partial_x \varphi_0(\vec{a}_i) U_i \partial_x \varphi_i(\vec{a}_i) \tag{4.6.20}$$

where $S[\varphi_0]$ is the action for the chiral boson field on a link of the saddlepoint network that we denote as the contour C_0,

$$S[\varphi_0] = -\tfrac{i}{4\pi} \int \mathrm{d}\tau \oint_{C_0} \mathrm{d}x \; \partial_x \varphi_0 \; \partial_- \varphi_0. \tag{4.6.21}$$

This contour is taken to be very large or infinite. Similarly, we define chiral boson fields φ_i on the disconnected but large contours C_i,

$$S[\varphi_i] = \tfrac{i}{4\pi} \int \mathrm{d}\tau \oint_{C_i} \mathrm{d}x \; \partial_x \varphi_i \partial_+ \varphi_i. \tag{4.6.22}$$

The sum in the interaction term in (4.6.20) is over the discrete set of nearly saddlepoints $\vec{a}_i$ along the contour C_0 where the fields φ_0 and φ_i interact with an appropriate, random strength U_i. This problem is in many ways quite similar to the problem of interacting edge channels with a randomly varying separation between them. We proceed along the same lines as in [54] and introduce a self energy Σ for the density-density correlation of the field φ_0. If we denote the Fourier transforms of the propagators $\langle \partial_x \varphi_0(x,\tau) \partial_x \varphi_0(x',\tau') \rangle$ and $\langle \partial_x \varphi_i(x,\tau) \partial_x \varphi_i(x',\tau') \rangle$ (with x, x' parametrising the positions on the contours C_0, C_j respectively) as

$$D_0(\omega, q) = \frac{-iq}{i\omega - v_{\mathrm{d}} q} \qquad ; \qquad D_j(\omega, q) = \frac{-iq}{i\omega + v_{\mathrm{d}} q}, \tag{4.6.23}$$

then the introduction of a self energy takes the form

$$D_0(\omega, q) \to -i \frac{q}{i\omega - (v_{\mathrm{d}} + \Sigma) q}. \tag{4.6.24}$$

To lowest order in the interaction potential we may write

$$\Sigma(\omega) = -i \tfrac{z}{2\pi} \overline{U_j^2} \int \mathrm{d}q \; D_j(\omega, q) = \tfrac{z}{2 v_{\mathrm{d}}^2} \overline{U_j^2} |\omega|. \tag{4.6.25}$$

Here, the bar stands for the average over the random positions $\vec{a}_i$ along C_0 and z is the linear density of saddlepoints. The result (4.6.25) can be used to obtain an expression for $1/\tau_{\text{in}}$, i.e. the imaginary part of the self energy as it appears in the electron Green's function $\mathcal{G}(\omega, q)$, as follows

$$1/\tau_{\text{in}} = \int \mathrm{d}\omega \mathrm{d}q \; \Sigma(\omega) \mathcal{G}(\varepsilon - \omega, q). \tag{4.6.26}$$

The τ_{in} determines the rate at which the electrons on the backbone cluster equilibrate with the rest of the electronic orbitals. We find $\tau_{\text{in}}^{-1} \propto \varepsilon^2$ or T^2 at finite temperatures.

This admittedly crude approach toward electron relaxation can be improved in several ways. For example, as the most important correction to the self energy (4.6.25) we find the self-interacting orbitals as depicted in Fig. 4.6b. These corrections replace the momentum integral in (4.6.25) in the following way (in space-time notation)

$$\begin{aligned} \int \mathrm{d}q\ D_j(\omega,q) &= \int \mathrm{d}\tau\ e^{-i\omega(\tau-\tau')} D_j(0,0;\tau-\tau') \\ D_j(0,0;\tau-\tau') &\to D_j(0,0;\tau-\tau') \\ &\quad + \int \mathrm{d}\tau_0 \int_0^L \mathrm{d}x \int_x^L \mathrm{d}y\ D_j(0,x;\tau-\tau_0)\ \tilde{U}_j\ D_j(y,L;\tau_0-\tau') \end{aligned} \tag{4.6.27}$$

where x, y are the positions of the nearly saddlepoint where the self-interaction takes place. The integrals stand for the averaging over positions and all dimensional factors are absorbed into $\tilde{U}_j$. The length of the orbital is given by L and boundary conditions $x \equiv x+L$ and $y \equiv y+L$ are understood. Equation (4.6.27) can be rewritten as a shift in the chemical potential,

$$\int \mathrm{d}q\ D_j(\omega,q) \to \int \mathrm{d}q\ \frac{-iq}{i\omega - \delta\mu + v_\mathrm{d} q} \qquad \delta\mu = \tilde{U}_j. \tag{4.6.28}$$

This leads to a modified self energy according to

$$\Sigma(\omega) \to \tfrac{z}{2v_\mathrm{d}^2} \overline{U_i^2}(\omega + i\delta\mu)\mathrm{sgn}(\omega). \tag{4.6.29}$$

The shift $\delta\mu$ can be translated into a shift in the expression for τ_in^{-1} following

$$\tau_\mathrm{in}^{-1}(\varepsilon,\delta\mu) = \left(1 + i\delta\mu \tfrac{\partial}{\partial\varepsilon}\right) \tau_\mathrm{in}^{-1}(\varepsilon). \tag{4.6.30}$$

After the analytic continuation to real energies ($i\varepsilon \to \varepsilon$) has been performed, we obtain the final result $\tau_\mathrm{in}^{-1} \propto \varepsilon$ or $\tau_\mathrm{in}^{-1} \propto T$ at finite temperatures. More generally, we expect the equilibration rate to be given by a regular series expansion in powers of T which is dominated by the lowest order $\tau_\mathrm{in}^{-1} \propto T$ as T approaches absolute zero.

4A One-dimensional propagator with Coulomb interaction

In this appendix we calculate the correlation function $G(\tau, 0)$ for the charged boson fields φ_i (4.4.21),

$$G(\tau,x) = \langle \varphi_i(\tau,x)\varphi_i(0,0)\rangle \qquad \tau > 0. \tag{4A.1}$$

In momentum and frequency space this correlator is given by (we omit the Landau level label i since it is of no consequence)

$$\left\langle \varphi_a(k)\varphi_{-b}(-k')\right\rangle = \frac{2\pi i}{\beta} \frac{\delta_{ab}\delta(k-k')}{k[\omega_a + ikv^\mathrm{eff}(k)]}. \tag{4A.2}$$

We write the 1D Coulomb interaction and the effective velocity v^eff in the following form

$$U_0(k) = -c\sqrt{2\pi}\ln(k/\Lambda)^2 \quad ; \quad v^\mathrm{eff}(k) = -mc\ln(k/\Lambda D)^2 \tag{4A.3}$$

where c is a positive constant indicating the strength of the Coulomb interaction, Λ is an ultraviolet cutoff and $D = \exp(v_\mathrm{d}/2mc)$. We will only consider low momenta $|k| < \lambda\Lambda$, with $\lambda < 1$, so that we are well away from the point where the Hamiltonian becomes negative. We take the Fourier transform of (4A.2) and change the frequency sum to an integral, writing $\sum_n \to \frac{\beta}{2\pi}\int \mathrm{d}\omega$,

$$\partial_\tau G(\tau,0) = \frac{i}{2\pi}\int_{-\lambda\Lambda}^{\lambda\Lambda}\mathrm{d}k\ v^{\mathrm{eff}}(k)\int_{-\infty}^{\infty}\frac{e^{i\omega\tau}\mathrm{d}\omega}{\omega + ikv^{\mathrm{eff}}(k)} = -\int_{-\lambda\Lambda}^{\lambda\Lambda}\mathrm{d}k\ v^{\mathrm{eff}}(k)\Theta(-kv^{\mathrm{eff}})e^{kv^{\mathrm{eff}}(k)\tau}. \tag{4A.4}$$

The step function $\Theta(-kv^{\mathrm{eff}})$ constrains the integration interval to $k < 0$. We can split the last expression in (4A.4) into two parts, using $\ln k\ \mathrm{d}k = d(k\ln k - k)$, and get

$$\partial_\tau G(\tau,0) = -\tfrac{1}{\tau}[1 - (\tfrac{\lambda}{D})^{2mc\tau\lambda\Lambda}] - 2mc\Lambda D\int_0^{\lambda/D}\mathrm{d}u\ \exp[2mc\tau\Lambda D\cdot u\ln u]. \tag{4A.5}$$

The function $u\ln u$ is negative on the whole interval $(0, \lambda/D)$, since $\lambda/D < 1$. If we now send the cutoff Λ to infinity, the term with the integral in (4A.5) will go to zero as $1/\ln\Lambda$. The term $(\lambda/D)^{2mc\tau\lambda\Lambda}$ also vanishes, yielding the free particle result

$$G(\tau,0) = -\ln\tau + \text{constant}. \tag{4A.6}$$

4B Chern-Simons action for bulk currents

In this appendix we show that (4.4.8) is equivalent to the following bulk action:

$$S[A, g_i] = \tfrac{i}{4\pi}\sum_\alpha\sum_{i=1}^{m}\left[-\int g_i^\alpha\wedge\mathrm{d}g_i^\alpha + 2\int g_i^\alpha\wedge\mathrm{d}A^\alpha\right] \tag{4B.1}$$

with the condition $(g_i^\alpha)_- = 0$ on the edge. The g_i are 2+1 dimensional potentials from which the electron current density j for every Landau level can be found,

$$j_i^\mu \propto \varepsilon^{\mu\nu\lambda}\partial_\nu(g_i)_\lambda. \tag{4B.2}$$

Notice three important subtleties:

- The coupling of g with the electromagnetic gauge field is of the form $g\wedge\mathrm{d}A$ instead of the expected $A\wedge\mathrm{d}g \propto j_\mu A^\mu$. These expressions differ by an edge term. The second form is *not* invariant under the gauge transformations $A \to A + \mathrm{d}\chi$; the expression $\mathrm{d}A$ on the other hand is manifestly gauge invariant.

- Putting an arbitrary spacetime component of g zero on the edge ensures that the action is invariant under $g_\mu \to g_\mu + \partial_\mu\kappa$, a gauge transformation that does not affect the current density. Without such a condition, gauge invariance is broken at the edge.

- Because of the invariance under $g_\mu \to g_\mu + \partial_\mu\kappa$, a gauge fixing condition has to be specified for the path integration over g, for instance the Coulomb gauge $\nabla\cdot\vec{g} = 0$.

Let us now for simplicity drop the replica indices α and the Landau level index i (effectively setting $m = 1$). Having taken the condition $g_-|_{\text{edge}} = 0$, the component g_- in (4B.1) multiplies the following constraint:

$$\nabla \times (\vec{g} - \vec{A}) = 0. \tag{4B.3}$$

After integration over g_-, what remains of the action is

$$\tfrac{i}{4\pi} \int \mathrm{d}^2 x \, \left(-\vec{g} \times \partial_- \vec{g} + 2\vec{g} \times [\nabla A_- - \partial_- \vec{A}] \right) \tag{4B.4}$$

subject to the constraint (4B.3). The general solution of (4B.3) is given by

$$\vec{g} = \vec{A} - \nabla \varphi \tag{4B.5}$$

with $\varphi(\vec{x})$ a real scalar field which is now the only integration variable that is left. Substitution into (4B.4) yields an action where φ features only on the edge,

$$S[\varphi, A] = \tfrac{i}{4\pi} \left[\int A \wedge \mathrm{d}A - \oint \mathrm{d}x \mathrm{d}\tau \, (D_x \varphi D_- \varphi - \varphi E_x) \right]. \tag{4B.6}$$

This is exactly of the form (4.4.8). One may worry that the path integration over φ is ill-defined, because of the bulk degrees of freedom of φ, which do not appear in (4B.6). However, φ inherits something from the gauge fixing condition of g. This is most easily seen in the case of the Coulomb gauge; here, φ has to satisfy $\nabla^2 \varphi = 0$. This means that the bulk degrees of freedom are completely determined by $\varphi(x)$ at the edge (the well known case of Laplace's equation with Dirichlet boundary conditions) and therefore aren't independent integration variables.

One final remark on the boundary condition $g_- = 0$: The Hamiltonian (density) corresponding to (4B.6) is given by $v_\mathrm{d}(D_x \varphi)^2$. One is not allowed to choose a velocity $v_\mathrm{d} < 0$, since this would lead to energies that are unbounded from below. In general, the boundary condition has to be taken in such a way that the velocity of the chiral bosons has the same sign as the prefactor multiplying $\frac{i}{4\pi}$ in (4B.1), otherwise the integration over g is ill-defined.

4C Inter-channel scattering at the edge

In this appendix we describe the various steps of the standard Q-field approach to edge disorder. For the general case of m chiral edge channels, one can differentiate between different types of disorder, depending on whether one allows inter-channel scattering or not. Although the different scattering potentials do not give rise to fundamentally different physical results, it is nevertheless important to define the 'effective' edge Hamiltonian (4.2.10) which gives rise to the same result (4.2.7) that was previously obtained for 2D electrons. Below we show that the following m channel model satisfies our requirements

$$\mathcal{H}^{kk'}_{\text{edge}} = -i v_\mathrm{d} \delta_{kk'} \partial_x + V_{kk'}(x), \tag{4C.1}$$

where V is a hermitian random matrix and the elements $V_{kk'}$ are distributed with a Gaussian weight

$$P[V] = \exp\{-\tfrac{1}{g} \oint \mathrm{d}x \ \mathrm{tr}\ V^2\}. \tag{4C.2}$$

The indices $k, k'=1,\cdots,m$ label the edge channels. The form (4C.1) implies that single potential scattering, as described by the 2D Hamiltonian

$$\mathcal{H}_{\rm 2D} = \tfrac{1}{2m_e}(\vec{p}-\vec{A})^2 + V(\vec{x}), \tag{4C.3}$$

does not naively translate into single potential scattering for the edge states as obtained by solving (4C.3) in the presence of an edge (infinite potential wall). Rather than that, one should allow for inter-channel scattering of the 'pure' eigenstates as in (4C.1) in order to reproduce the effect of dirt in the general 2D problem (4C.3). We start from the following generating function for the averaged free particle propagators

$$Z = \int \mathcal{D}\bar{\psi}\psi \int \mathcal{D}V P[V] \exp \beta \sum_{p,\alpha,jj'} \oint {\rm d}x \; \bar{\psi}_p^{\alpha,j} \left[(\mu + ip\eta)\delta_{jj'} - \mathcal{H}_{\rm edge}^{jj'} \right] \psi_p^{\alpha,j'}. \tag{4C.4}$$

Integration over randomness and introduction of the matrix field $\widetilde{Q}_{pp'}^{\alpha\beta}(x)$ by performing the Hubbard-Stratonovich trick leads to

$$Z = \int \mathcal{D}\widetilde{Q} \; \exp\left\{ -\tfrac{1}{g}{\rm Tr}\; \widetilde{Q}^2 + m{\rm Tr}\; \ln[\mu + iv_{\rm d}\partial_x + i\widetilde{Q} + i\eta\Lambda] \right\}. \tag{4C.5}$$

Notice that the edge channel label is not present in the new field variable $\widetilde{Q}$, but it is simply contained in an overall factor m.

We will next make use of the simple analytic properties of our 1D Hamiltonian and show that the saddlepoint technique yields, in fact, exact results for all m and that therefore there is no need to rely on m to be 'large'. The stationary point equation for $\widetilde{Q}$,

$$i[\widetilde{Q}_{\rm sp}]_{pp'}^{\alpha\beta} = \delta^{\alpha\beta}\delta_{pp'}[e_0 + (-1)^p i/2\tau], \tag{4C.6}$$

can be written as

$$\tfrac{2}{g}(e_0 \pm i/2\tau) = -m \int_{-\infty}^{\infty} \tfrac{{\rm d}q}{2\pi} \left[\mu - v_{\rm d}q + e_0 \pm i(1/2\tau + \eta) \right]^{-1} = \pm im/2v_{\rm d} \tag{4C.7}$$

with the simple solution $e_0=0$, $\tau=2v_{\rm d}/mg$. One may next replace the original Q-field by the following change of variables,

$$\widetilde{Q} \to T^{-1}PT \to \tfrac{1}{2\tau}T^{-1}\Lambda T =: \tfrac{1}{2\tau}Q. \tag{4C.8}$$

Here, the $T \in U(2N_r)$ are unitary rotations and the block-diagonal hermitian $P_{pp'}^{\alpha\beta} = \delta_{pp'}P_p^{\alpha\beta}$ represent the longitudinal components. Replacing P by its saddlepoint value, as written in (4C.8), turns out to be an exact statement, valid for all m. The reason is contained in the fact that the fluctuations in P are weighted by propagators with poles in either the positive or negative imaginary momentum plane. All the momentum integrals therefore sum up to zero, giving rise to a zero weight to all orders in the P-fluctuations. The replacement of (4C.8) is exact when inserted in the Tr ln. Eq. (4C.5) factorises into

$$\begin{aligned} Z &= Z_P \cdot Z_T \\ Z_P &= \int \mathcal{D}P \; I[P] \exp\{-\tfrac{1}{g}{\rm Tr}\; P^2\} \\ Z_T &= \int \mathcal{D}T \; \exp m{\rm Tr}\; \ln[\mu + iv_{\rm d}\partial_x + \tfrac{i}{2\tau}\Lambda + iB] \end{aligned} \tag{4C.9}$$

where all T-dependence is contained in the quantity B according to

$$B = v_{\mathrm{d}} T \partial_x T^{-1} + \eta T \Lambda T^{-1} =: v_{\mathrm{d}} T D_0 T^{-1}. \tag{4C.10}$$

Equation (4C.9) can be evaluated further, and to lowest few orders in an expansion in B we obtain an effective action which can be written as

$$\begin{aligned} Z &= \int \mathcal{D}T \ \exp S_{\mathrm{eff}}[T] \\ S_{\mathrm{eff}}[T] &= \tfrac{m}{2v_{\mathrm{d}}} \mathrm{Tr}\ \Lambda B(x) - \tfrac{m\tau}{8v_{\mathrm{d}}} \mathrm{Tr}\ [B(x), \Lambda]^2 + \cdots \qquad (4C.11) \\ &= \tfrac{m}{2} \oint \mathrm{d}x\ \mathrm{tr}\ \Lambda T \partial_x T^{-1} + \tfrac{m}{2v_{\mathrm{d}}} \eta \oint \mathrm{d}x\ \mathrm{tr}\ \Lambda Q - \tfrac{m\tau v_{\mathrm{d}}}{8} \oint \mathrm{d}x\ \mathrm{tr}\ [D_0, Q]^2. \qquad (4C.12) \end{aligned}$$

The coefficients appearing in (4C.12) all have a clear physical meaning in the context of disordered edge states (see also the main text). In particular, m stands for the quantised Hall conductance σ_{xy}; $m/2\pi v_{\mathrm{d}}$ equals the total density of edge states ρ_{e}. The quantity $m\tau v_{\mathrm{d}}$ that appears in the higher dimensional operators is the 1D conductivity σ_{xx} of m channel edge states. Here, $2\tau v_{\mathrm{d}}$ is the linear dimension which sets the smallest wavelength for the Q field variables and $m/2$ is the (quantised) conductance (g_m) of the wire.

4D Action for Q and A_μ on multiple edges

The generalisation of (4.3.42) is given by

$$\begin{aligned} S[Q,A] &= \tfrac{\beta/2}{(2\pi)^2} \int \mathrm{d}^2x \mathrm{d}^2x'\ n(\vec{x}) B^\dagger(\vec{x}) U_0(\vec{x}-\vec{x}') n(\vec{x}') B(\vec{x}') + \sum_{a=1}^{M} s_a S_{\mathrm{top}}^{(a)}[Q] \\ &\quad + \tfrac{i}{4\pi} \left[\sum_\alpha \int n (A^{\mathrm{eff}})^\alpha \wedge \mathrm{d}(A^{\mathrm{eff}})^\alpha + \beta \sum_{a=1}^{M} s_a \oint_{C_a} \mathrm{d}x \left(A_x^\dagger A_0^{\mathrm{eff}} - \tfrac{2\pi}{\beta} \mathrm{tr}\ \hat{A}_x Q \right) \right] \\ &\quad + \sum_{a=1}^{M} S_{\mathrm{F}}^{(a)}[Q] - \tfrac{\pi\rho_e^2}{2\beta} \sum_{n\alpha} \sum_{a=1}^{M} \int_{C_a} \tfrac{\mathrm{d}k_x}{1+\rho_{\mathrm{e}}\sqrt{2\pi} U_0(k_x)} \left| \mathrm{tr}\ \mathrm{I}_n^\alpha Q(k_x) - \tfrac{\beta}{\pi} (A_0^{\mathrm{eff}})_{-n}^\alpha (k_x) \right|^2 \\ &\quad + \tfrac{1}{8\beta v_{\mathrm{d}}^2} \sum_{a\neq b} s_a s_b \oint_{C_a} \mathrm{d}x \oint_{C_b} \mathrm{d}x' \sum_{n\alpha} [\mathrm{tr}\ \mathrm{I}_n^\alpha Q - \tfrac{\beta}{\pi} (A_0^{\mathrm{eff}})_{-n}^\alpha](x)\ \times \\ &\quad \times\ U_0(x,x') [\mathrm{tr}\ \mathrm{I}_{-n}^\alpha Q - \tfrac{\beta}{\pi} (A_0^{\mathrm{eff}})_n^\alpha](x') \end{aligned} \tag{4D.1}$$

where $U_0(x,x')$ denotes the full 2D Coulomb interaction. All terms except those quadratic in Q arise by the obvious replacements $m \to n(\vec{x})$ and $m \oint \to \sum_a s_a \oint_{C_a}$ in (4.3.42). The terms quadratic in Q can be understood as follows. In the generalised form of (4.3.24), the quadratic term in the plasmon field is given by

$$-\tfrac{\beta}{2} \int \mathrm{d}^2k\ \lambda^\dagger U_0^{-1} \lambda - \tfrac{\beta}{4\pi v_{\mathrm{d}}} \sum_a \oint_{C_a} \lambda^\dagger \lambda, \tag{4D.2}$$

indicating that the propagator for λ between two points on the same edge will be very different from the propagator between two distinct edges. In the former case the propagator is proportional to $[U_0^{-1} + \frac{\rho_{\mathrm{e}}}{\sqrt{2\pi}}]^{-1}$, which is exactly the form obtained by combining the Finkelstein term with the next term in (4D.1). In the latter case, the propagator is simply proportional to U_0. Finally, the signs can be understood by noticing that the coupling of the plasmon field to Q is proportional to $\sum_a s_a \oint_{C_a} \mathrm{tr}\ \hat{\lambda} Q$.

Chapter 5

The Chern-Simons mapping

5.1 Introduction

In this chapter we finally come to the subject of the *fractional* quantum Hall effect. We perform the Chern-Simons mapping on the three scenarios that were discussed in the previous chapters: (1) the bulk theory, (2) the chiral edge boson theory for both short-ranged and long-ranged disorder and (3) the Q-theory for percolating edge states at the plateau transitions.

In section 2.4 we derived, at tree level, an effective bulk action for A_μ in the case of weak magnetic fields. In section 5.2 we map this action to obtain a theory for $\nu \approx 1/(2p)$. We find that the conductances in this regime transform according to the SL(2,$\mathbb{Z}$) mapping $\sigma \to \sigma/(2p\sigma+1)$. Additionally, we compute the contribution of the statistical gauge fields to the specific heat. For the problem with Coulomb interactions we find that the singularity structure of the theory is not modified. This demonstrates that the composite fermion approach to the half integral effect is free of infrared trouble. On the other hand, for a system with finite range interactions complications do arise. The statistical gauge fields in this case give rise to a divergent quasiparticle d.o.s.

In chapter 4 we established the relation between the Q-field approach to the iqHe and the theory of *chiral edge bosons*. In section 5.3 we introduce C-S fluctuations a_μ in the edge boson theory and integrate them out. In this procedure the C-S fluctuations are defined on the entire two dimensional plane while the electronic degrees of freedom are confined to the half plane. The mapping leads to the familiar K-matrix structure of the edge excitations [24, 25, 26, 8, 77]. In section 5.4 we introduce electron tunneling operators into the iqHe edge boson theory and redo the mapping of section 5.3, obtaining quasiparticle tunneling operators and tunneling exponents. The Kane-Fisher-Polchinski result [37] for tunneling between a Fermi liquid and a quantum Hall state is found.

In view of the fact that the Luttinger liquid theory of the quantum Hall edge has already been extensively investigated following the phenomenological theory of Wen [77], it may be helpful to stress some of the main differences between the standard approach and our direct microscopic way of obtaining a theory of the edge. First, our edge theory is coupled to the external electromagnetic field throughout the derivation. This obviously has considerable technical as well as conceptual advantages. For example, the appearance of the so-called *chiral anomaly* [5, 52] plays a crucial role in relating the Hall conductance as an *edge* phenomenon unambiguously to the Hall conductance as a *bulk* phenomenon. In this way we avoid mistakes in applying ad hoc pictures like the Landauer-Büttiker formalism [36]. The chiral anomaly also provides an explicit demonstration of Laughlin's gauge argument [57] and it plays a role in identifying the fractionally charged edge excitations.

Secondly, the C-S mapping is ill-defined in the absence of long range interactions between the particles. For example, without the Coulomb interactions, the Hamiltonian for counter-flowing edge modes would become unbounded from below. The significance of long range interactions is inevitably overlooked in the phenomenological approach, where one is free to choose boundary conditions (velocities) and bounded Hamiltonians even for noninteracting particles.

Thirdly, there is a difference in the way that quasiparticle operators and tunnelling exponents are obtained. Our flux attachment procedure leaves no room for ambiguities, since it directly maps the electron operators of the iqH edge into quasiparticle operators of the fqH edge. This procedure discriminates between opposite sides of an edge, thus giving a new, geometrical interpretation to fractionally charged edge excitations. Physically this means that tunnelling processes are described by different particles with different charge and statistics, depending on whether one tunnels through the vacuum or through the incompressible quantum Hall state. The *duality* transformation [5] that inverts the K-matrix (compactification radius) can be interpreted as a reversal of the bulk and the vacuum.

Finally, our velocity matrix is diagonal simultaneously with K. Since impurities have been dealt with in the first stages of our derivation, they do not make a second appearance in the form of inter-channel scattering of chiral bosons. Consequently, the calculation of tunnelling exponents is not plagued by the kind of non-universalities found by [36, 37] for counter-flowing modes.

In order to facilitate the analysis of slowly varying potential fluctuations, we introduced in chapter 4 the concept of spatially separated edge channels. We argued that smooth potential fluctuations near the sample edge decompose the multiple (iqHe) edge states into physically distinguishable and separately conserved edge modes that correspond to Landau levels. Since in our theory of the dirty, sharp edge the various edge modes are assumed to scatter strongly amongst each other, it is a priori not entirely obvious whether a system of spatially separated edge modes maps onto a physically equivalent fqH state under the flux attachment transformation. This problem is closely related to the problem of counter-flowing edge modes and to a number of puzzles [48, 43] that recently arose in the context of the conventional K-matrix approach.

In section 5.5 we apply the flux attachment transformation to the case of separated edge channels. We obtain a rich structure of decomposed fqH edge states which is reminiscent of the picture of multiple condensates in the hierarchy theory of the fqHe [68, 28, 31]. We find an *enlarged dual symmetry* of abelian quantum Hall states. We employ the enlarged duality in order to derive a complete classification of operators for fractionally charged edge excitations as well as tunnelling exponents.

In the limit of large wavelengths (i.e. for distances large compared to the spatial separation between the edge modes) our theory of separated edge modes reduces in effect to a Luttinger liquid K-matrix theory of a sharp edge. We use our explicit results for the 'clean' and 'dirty' edges to show that they describe identical physics in the limit of large length scales. This not only means that the fractional quantum hall *conductance* is the same in both cases, but also identical exponents are obtained for the tunnelling of electrons and quasiparticles. For states with only chiral edge modes these results are precisely what one would naively expect. For more complicated states with counter-flowing components, however, the relations are nontrivial.

In section 5.5.5 we embark on the problem of *arbitrary* filling fractions in the fqH regime, following up on chapter 4. We use the idea of percolating edge states and derive an effective Luttinger liquid theory for tunnelling processes into the fqH edge. Here, the main idea is again that the long range Coulomb interactions between the edge and the localised bulk orbitals give rise to an effective, chiral action of edge excitations in which the neutral modes are strongly suppressed. Only the charged mode contributes to the tunnelling processes and the effective Luttinger liquid parameter varies continuously with the filling fraction ν. This yields a tunnelling exponent $1/\nu$.

In section 5.5.6 we investigate plateau transitions between fqH states that have the same value of p. We employ the percolating network model as in chapter 4 and perform the C-S mapping on the 1+1 dimensional Q-field theory. The results indicate that the critical aspects of the plateau transitions are quite generally the same for the integral and fractional regime, independent of the type of disorder (long range vs. short range). However, our results do not exhibit the SL(2,$\mathbb{Z}$) symmetry that has been proposed in [51, 50, 38]. This symmetry, therefore, has no microscopic justification and it does not appear as a fundamental symmetry of the quantum phase transition.

Statistical gauge fields

In section 1.2.3 we showed the way in which statistical gauge fields are introduced and employed to 'map' the iqH regime into the fqH regime. Starting from a theory for the integer effect, one adds the *fluctuations* a_μ of the statistical gauge field to the electromagnetic fluctuations A_μ and includes an extra Chern-Simons term $\frac{i}{4\pi}\frac{1}{2p}\int a\wedge da$ in the action. Integration over a_μ then yields a theory for the *fractional* effect, which one can think of as the *integer* effect of composite fermions that consist of an electron plus $2p$ bound flux quanta. The filling fractions in the Jain series are obtained in this way,

$$\nu = \frac{m}{2pm+1}. \tag{5.1.1}$$

Notice that the iqHe theory that one starts with represents the saddle point theory of the sought after fqHe theory coupled to statistical gauge fields. In the Q-field formalism this mapping procedure works in an elegant way. As explained in section 2.3.3, the Q-field only couples to the fluctuations since the T-matrix commutes with the zero frequency components of gauge fields. The static component of the statistical gauge field only enters into the underlying P-field theory.

5.2 Chern-Simons mapping of the bulk theory

5.2.1 Mapping of the conductances

Let us first derive an effective action for the bulk of a fqH sample. Since we are primarily interested in the way the conductances σ_{xx} and σ_{xy} are affected by the C-S mapping, it is easiest to start from the bulk action $S[A_\mu]$ (2.4.11) or, equivalently, $S[\Psi,\Phi]$ (2.4.17). For simplicity we drop replica indices and only consider the case where $\sigma_{\mathrm{xy}}^{\mathrm{II}}=0$, i.e. weak magnetic fields. Since we are dealing with a weak magnetic field, the filling fraction we start with is large, $m\gg 1$, resulting in a mapped filling $\nu\approx 1/(2p)$. In order to bring the

C-S action for the statistical gauge field a_μ into the form of (2.4.17), we write

$$a_i = \varepsilon_{ij}\partial_j \psi \quad ; \quad a_0 = -\partial_0 \varphi. \tag{5.2.1}$$

This leads to the following extension of (2.4.17),

$$\begin{aligned} S[\Phi,\Psi] &\rightarrow S[\Phi+\varphi,\Psi+\psi] + S_{\rm cs}[\varphi,\psi] \qquad (5.2.2)\\ &= -\sigma^0_{\rm xx}\sum_{n>0}\int {\rm d}^2q\; nq^2 \left\{(\Phi^*+\varphi^*,\Psi^*+\psi^*)M\begin{pmatrix}\Phi+\varphi\\ \Psi+\psi\end{pmatrix} + (\varphi^*,\psi^*)M_{\rm cs}\begin{pmatrix}\varphi\\ \psi\end{pmatrix}\right\}\end{aligned}$$

with M as defined in (2.4.18) and $M_{\rm cs}$ given by

$$M_{\rm cs} = \frac{1}{2p\sigma^0_{\rm xx}}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}. \tag{5.2.3}$$

Integrating out a_μ (i.e. φ and ψ) is straightforward but tedious. The result is an action precisely of the form (2.4.17), where the conductances σ^0_{ij} are replaced by the following expressions,

$$\sigma_{\rm xx} = \frac{\sigma^0_{\rm xx}}{(2p\sigma^0_{\rm xx})^2 + (2p\sigma^0_{\rm xy}+1)^2} \quad ; \quad \sigma_{\rm xy} = \frac{1}{2p}\left[1 - \frac{2p\sigma^0_{\rm xy}+1}{(2p\sigma^0_{\rm xx})^2 + (2p\sigma^0_{\rm xy}+1)^2}\right]. \tag{5.2.4}$$

This result is precisely the SL(2,$\mathbb{Z}$) transformation (1.2.14). It becomes exact in the limit $\sigma^0_{\rm xx}\to\infty$, which is the weak coupling case considered here. Equation (5.2.4) applies also to the case where $\sigma^0_{\rm xx}\to 0$ but with a modified definition of $\sigma^0_{\rm xy}$ which is now integrally quantised.

It is reasonable to assume that (5.2.4) gives a good overall description provided one is not too close to the critical plateau transition. In this case one expects the conductances to be broadly distributed. This then complicates the relation between average and measured conductances, and (5.2.4) is likely to be affected by the higher order response terms which gave been neglected in (2.4.11) [15]. It is important to remark that the SL(2,$\mathbb{Z}$) mapping is neither unique nor universal, but that it depends on microscopic details of the system such as disorder. As we will see in section 5.5.6, the C-S gauge fields require a different treatment in a theory with slowly varying potential fluctuations, resulting in a different mapping between integer and fractional regimes.

5.2.2 Internal energy; specific heat

In order to decide whether the fluctuations in the C-S gauge fields are well-behaved, we next compute the free energy and extract from it the specific heat. Omitting replicas again, we write for $p=1$

$$\det[M+M_{\rm cs}] = \mathcal{G}\left[1+(\omega_c\tau)^2+\tfrac{\omega_c\tau}{\sigma^0_{\rm xx}}\right] + (2\sigma^0_{\rm xx})^{-2}. \tag{5.2.5}$$

The contribution to the free energy can be written as

$$\delta F = \sum_{n>0}\int {\rm d}^2q\; \ln\left\{\mathcal{G}\left[1+(\omega_c\tau)^2+\tfrac{\omega_c\tau}{\sigma^0_{\rm xx}}\right] + (2\sigma^0_{\rm xx})^{-2}\right\}. \tag{5.2.6}$$

In particular, we consider the derivative with respect to temperature

$$\frac{\partial \delta F}{\partial \ln T} = \sum_{n>0} \int \mathrm{d}^2 q \, \frac{1 + (\omega_c \tau)^2 + \omega_c \tau / \sigma^0_{\mathrm{xx}}}{\mathcal{G}[1 + (\omega_c\tau)^2 + \frac{\omega_c \tau}{\sigma^0_{\mathrm{xx}}}] + (2\sigma^0_{\mathrm{xx}})^{-2}} \left[\frac{q^2}{q^2 + \kappa n Y} - (\frac{q^2}{q^2 + \kappa n Y})^2 \right]$$
$$= \sum_{n>0} \int \mathrm{d}^2 q \, \frac{\kappa n [1 + (\omega_c \tau)^2 + \omega_c \tau / \sigma^0_{\mathrm{xx}}]}{\kappa n [1 + (\omega_c\tau)^2 + \frac{\omega_c \tau}{\sigma^0_{\mathrm{xx}}}] + \frac{1}{8} \rho_0 (2\sigma^0_{\mathrm{xx}})^{-2} (\kappa n + q^2/Y)} \cdot \frac{q^2}{q^2 + \kappa n Y} \tag{5.2.7}$$

with Y and κ as defined in (2.3.7) and (2.4.3). This expression is well-behaved for small frequencies ν_n and can be written in the form

$$\frac{\partial \delta F}{\partial \ln T} = \sum_{n>0} \nu_n \rho(\nu_n) \tag{5.2.8}$$

where $\rho(\nu_n)$ is of order $\nu_n \ln \nu_n$ for small ν_n. However, for free particles or short-ranged interactions the insertion of C-S gauge fields leads to singular contributions, since by putting $Y =$ constant one finds $\rho(\nu_n) \approx |\ln \nu_n|$ for small frequencies. This implies that the C-S fields lead to a singular quasiparticle d.o.s.

$$\rho_{\mathrm{qp}}(\varepsilon) = \rho(i\varepsilon) + \rho(-i\varepsilon) \tag{5.2.9}$$

entering the expression for the specific heat (see chapter 3). The exact meaning of the C-S gauge field procedure is not clear in this case. The results demonstrate the fundamental significance of $\mathcal{F}$-invariance in the problem, possibly indicating that a new saddlepoint should be found for the finite range interaction problem.

5.3 Chern-Simons mapping on the edge

5.3.1 The sharp edge

In chapter 4 we derived the complete iqHe edge theory in a d.o.s. gap in several equivalent forms and in terms of different fields. Of these, the chiral edge boson action (4.4.24) is best suited for the C-S mapping. Here we write it down again, omitting the replica indices.

$$S_{\mathrm{chiral}}[\varphi_i, A] = \tfrac{i}{4\pi}\mathrm{sgn}(m) \sum_{i=1}^{|m|} \left[\int A \wedge \mathrm{d}A - \oint (D_x \varphi_i D_- \varphi_i - E_x \varphi_i) \right] \tag{5.3.1}$$
$$- \tfrac{1}{2} (\tfrac{1}{2\pi})^2 \sum_{i,j=1}^{|m|} \int_\infty \mathrm{d}\tau \mathrm{d}^2 x \mathrm{d}^2 x' \;\; U_0(\vec{x} - \vec{x}') \; \nabla \times [\theta(y) \vec{D} \varphi_i(\vec{x})] \; \nabla' \times [\theta(y') \vec{D} \varphi_j(\vec{x}')].$$

The $\oint$ stands for $\int_0^\beta \mathrm{d}\tau \oint \mathrm{d}x$ and we again use the notation where $\int$ denotes integration over the half plane $y > 0$ and $\int_\infty$ over the whole plane. In writing (5.3.1), we have included the possibility of m being negative. The reason for this is that we want to be able to perform Jain's composite fermion mapping [34] in such a way that the electron and the composite fermion experience magnetic fields of opposite sign (see Fig. 5.1 for the example $\nu = 2/3$). We use a convention in which ν and B are always positive. Since particle densities are positive, we then have to allow negative B_{eff} and m.

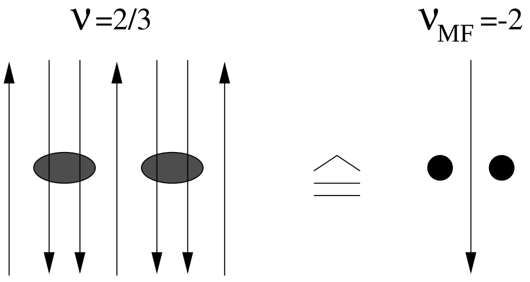

Figure 5.1: *Equivalence between filling fraction 2/3 of electrons coupled to a C-S gauge field (left) and filling −2 in mean field theory (right). Arrows denote flux quanta. The dots represent electrons.*

We implement the composite fermion picture again in the standard way by introducing statistical gauge fields a_μ. The action for the a_μ has the form of a C-S action and our starting point can be written as

$$S = S_{\text{chiral}}[\varphi_i, A + a] + \tfrac{i}{4\pi} \cdot \tfrac{1}{2p} \int_\infty a \wedge \mathrm{d}a \tag{5.3.2}$$

We stress the fact that the statistical gauge fields a_μ are defined over the entire plane. The electronic part S_{chiral} is restricted to the half plane and this, properly regarded, is the result of a confining potential added to the electron gas. The C-S mapping, however, applies to the whole system and isn't 'interested' in the fact that the electrons happen to sit only at $y > 0$. This observation will prove to be important in two ways. First, taking the integral over the whole plane leads to a proper C-S mapping in which unphysical boundary terms are avoided. Second, it means that statistical flux also gets attached to particles that tunnel into the edge through the 'vacuum'.

5.3.2 Integrating out the C-S fluctuations

We note that the action (5.3.2) can be written in such a way that $a_-(= a_0 - i v_{\mathrm{d}} a_x)$, the component in the 'preferred' chiral direction, evidently multiplies a constraint,

$$\begin{aligned} S \;=\; & \tfrac{i}{4\pi}\Bigg[\tfrac{1}{2p}\int_\infty (2a_-\nabla\times\vec{a} - \vec{a}\times\partial_-\vec{a}) \\ & +\mathrm{sgn}(m)\sum_{i=1}^{|m|}\int_\infty \Big(2\mathcal{D}_-\varphi_i\nabla\times[\theta(y)\vec{\mathcal{D}}\varphi_i] - \theta(y)\vec{\mathcal{D}}\varphi_i\times\partial_-\vec{\mathcal{D}}\varphi_i\Big)\Bigg] \\ & -\tfrac{1}{2}(\tfrac{1}{2\pi})^2\sum_{i,j=1}^{|m|}\int_\infty U_0(\vec{x}-\vec{x}')\ \nabla\times[\theta(y)\vec{\mathcal{D}}\varphi_i(\vec{x})]\ \nabla'\times[\theta(y')\vec{\mathcal{D}}\varphi_j(\vec{x}')]. \end{aligned} \tag{5.3.3}$$

Here $\mathcal{D}$ is defined as a covariant derivative containing both gauge fields, $\mathcal{D}_\mu\varphi_i = \partial_\mu\varphi_i - A_\mu - a_\mu$. The remarkable thing about (5.3.3) is that the edge bosons are completely contained in *bulk* integrals. Derivatives of the step functions with respect to the y coordinate lead to the edge terms in (5.3.2). The constraint multiplied by a_- is given by

$$0 = \nabla\times\left[\vec{a} - 2p\,\mathrm{sgn}(m)\theta(y)\sum_i\vec{\mathcal{D}}\varphi_i\right] \tag{5.3.4}$$

with general solution

$$\vec{a} = \frac{2p}{2pm\theta(y)+1}\mathrm{sgn}(m)\left[\theta(y)\sum_i \vec{D}\varphi_i + \nabla\Omega\right] \tag{5.3.5}$$

where Ω is an arbitrary function. Substitution of (5.3.5) into what remains of (5.3.3) after integration of a_- yields an expression containing only the following linear combinations

$$\tilde{\varphi}_i := \varphi_i - 2p\ \mathrm{sgn}(m)\Omega. \tag{5.3.6}$$

The resulting mapped action has a form similar to (5.3.3),

$$\begin{aligned} S &= \tfrac{i}{4\pi}\sum_{ij}\left[\delta_{ij}\ \mathrm{sgn}(m) - 2p\tfrac{\nu}{m}\right]\int_\infty \Big(2D_-\tilde{\varphi}_i\nabla\times[\theta(y)\vec{D}\tilde{\varphi}_j] - \theta(y)\vec{D}\tilde{\varphi}_i\times\partial_-\vec{D}\tilde{\varphi}_j\Big) \\ &\quad -\tfrac{1}{2}(\tfrac{\nu}{2\pi m})^2\sum_{ij}\int_\infty U_0(\vec{x}-\vec{x}')\ \nabla\times[\theta(y)\vec{D}\tilde{\varphi}_i(\vec{x})]\ \nabla'\times[\theta(y')\vec{D}\tilde{\varphi}_j(\vec{x}')]. \end{aligned} \tag{5.3.7}$$

See appendix 5A for the explicit calculation. In the expression $[\cdots]$ we recognise the inverse of the coupling matrix ('K-matrix') that was obtained in C-S theories [24, 25, 26, 8, 80],

$$K_{ij} = \delta_{ij}\ \mathrm{sgn}(m) + 2p. \tag{5.3.8}$$

Bringing the action (5.3.7) into the form (5.3.1) we obtain

$$\begin{aligned} S &= \tfrac{i\nu}{4\pi}\int A\wedge \mathrm{d}A - \tfrac{i}{4\pi}\sum_{ij}(K^{-1})_{ij}\oint (D_x\tilde{\varphi}_i D_-\tilde{\varphi}_j - \tilde{\varphi}_i E_x) \\ &\quad -\tfrac{1}{2}(\tfrac{\nu}{2\pi m})^2\sum_{ij}\int_\infty U_0(\vec{x}-\vec{x}')\ \nabla\times[\theta(y)\vec{D}\tilde{\varphi}_i(\vec{x})]\ \nabla'\times[\theta(y')\vec{D}\tilde{\varphi}_j(\vec{x}')]. \end{aligned} \tag{5.3.9}$$

Eq. (5.3.9) is the main result of this section. Even though bits and pieces of this action, such as the coupling of $\tilde{\varphi}$ to the gauge field [5], have appeared in the literature before and the K-matrix structure is wholly familiar, our complete theory is new, as many aspects of the impurity problem were previously not understood.

The Hall conductance on the edge can unambiguously be determined from (5.3.9) by using the chiral anomaly [5, 52]. The chiral anomaly enables us to use Laughlin's flux argument on the edge, since it relates the divergence of the edge current to the electric field parallel to the edge,

$$\partial_\mu J^\mu_{\mathrm{edge}} = -\tfrac{1}{2\pi}\sigma_{\mathrm{H}}E_x \qquad ; \qquad J^\mu_{\mathrm{edge}} = \tfrac{\delta S}{\delta A_\mu}\Big|_{\mathrm{edge}}. \tag{5.3.10}$$

In this way we easily confirm that the Hall conductance on the edge σ_{H} is equal to ν, as it should be. In order to make contact with phenomenological models [77], we mention that we can write down a bulk C-S action that is completely equivalent to (5.3.9). The electronic degrees of freedom are contained in fields h_i that act as potentials for the current,

$$S = \tfrac{i}{4\pi}\sum_{ij=1}^{|m|}(K^{-1})_{ij}\left[-\int h^i\wedge \mathrm{d}h^j + 2\int h^i\wedge \mathrm{d}A^{\mathrm{eff}}\right] + \tfrac{\nu^2}{8\pi^2}\int U_0(\vec{x}-\vec{x}')B(\vec{x})B(\vec{x}'). \tag{5.3.11}$$

The h^i satisfy the following gauge fixing constraint at the edge

$$\left[h^i_-(k_x) - i\tfrac{m}{\sqrt{2\pi}}U_0(k_x)\sum_{a=1}^{|m|} h^a_x(k_x)\right]_{\text{edge}} = 0. \tag{5.3.12}$$

The free part of (5.3.11) is known from the work of Wen [77]. However, lacking a microscopic edge theory, the full details of the interactions were unknown.

Charged and neutral modes

Diagonalisation of the K-matrix yields one charged mode and $|m|-1$ neutral modes. The charged mode Γ is given by

$$\Gamma = \tfrac{1}{|m|}\sum_i \tilde{\varphi}_i \tag{5.3.13}$$

where the normalisation has been chosen such that Γ has unit charge. A possible choice of basis for the neutral modes g_a is

$$\begin{aligned} g_a &= \tfrac{1}{a}\sum_{k=1}^{a}\tilde{\varphi}_k - \tilde{\varphi}_{a+1} \qquad a = 1,\cdots,|m|-1 \\ \tilde{\varphi}_k &= \Gamma - \tfrac{k-1}{k}g_{k-1} + \sum_{a=k}^{|m|-1}\tfrac{1}{a+1}g_a. \end{aligned} \tag{5.3.14}$$

The Jacobian of the transformation from the $\tilde{\varphi}_i$ to the modes (5.3.13, 5.3.14) is unity. In diagonalised form the action reads

$$\begin{aligned} S &= \tfrac{i\nu}{4\pi}\left[\int A\wedge \mathrm{d}A - \oint (D_x\Gamma D_-\Gamma - \Gamma E_x)\right] - \tfrac{i}{4\pi}\mathrm{sgn}(m)\sum_{a=1}^{|m|-1}\tfrac{a}{a+1}\oint \partial_x g_a\partial_- g_a \\ &\quad -\tfrac{1}{2}(\tfrac{\nu}{2\pi})^2\int_\infty U_0(\vec{x}-\vec{x}')\ \nabla\times[\theta(y)\vec{D}\Gamma(\vec{x})]\ \nabla'\times[\theta(y')\vec{D}\Gamma(\vec{x}')]. \end{aligned} \tag{5.3.15}$$

Naturally, the charged mode is the only one that couples to A_μ and feels interactions.

5.3.3 (In)stability of counter-flowing channels

Here we make an important observation which has been missed in the literature, namely that *for the noninteracting case or for short-ranged interactions, the C-S mapping is ill-defined at negative m.* In order to see this, we look at the Hamiltonian for the charged mode corresponding to (5.3.15) in the absence of Coulomb interactions,

$$H_\Gamma^{(\text{free})} = \frac{v_\mathrm{d}\nu}{4\pi}\oint \mathrm{d}x\ (D_x\Gamma)^2. \tag{5.3.16}$$

Recall that v_d is defined as $m/(2\pi\rho_{\text{edge}})$ and that $\nu > 0$. Hence, for negative m the free Hamiltonian is unbounded from below! The Coulomb interactions cure this situation by adding an extra term to the edge Hamiltonian,

$$H_\mathrm{c} = \tfrac{1}{2}(\tfrac{\nu}{2\pi})^2\oint \mathrm{d}x\mathrm{d}x'\ U_0(x,x')D_x\Gamma(x)D_x\Gamma(x') \tag{5.3.17}$$

effectively changing the velocity $v_{\rm d}$ to

$$\tilde{v}(k_x) = v_{\rm d} + \tfrac{\nu}{\sqrt{2\pi}} U_0(k_x) \qquad , \qquad U_0(k_x) \propto \ln(k_x) \tag{5.3.18}$$

For low momenta the logarithm in $\tilde{v}(k_x)$ is dominant, ensuring that the effective velocity $\tilde{v}$ is positive even if $v_{\rm d} < 0$ and thereby yielding a bounded Hamiltonian. We conclude that the long range Coulomb interactions are actually needed in order for the C-S mapping to be well defined. We have seen a similar situation occurring in the bulk theory with half integer filling fraction (section 5.2.2) where in the absence of long range interactions the C-S procedure gives rise to a divergent quasiparticle d.o.s.

5.4 Tunnelling density of states

5.4.1 The integer case

In chapter 4 we obtained the following results for the iqHe. In the φ-field theory, the tunnelling d.o.s. at the edge is given by

$$\left\langle \exp -i \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \; \partial_0 \varphi_j(\tau, \vec{x}_0) \right\rangle , \qquad j = 1, \ldots, |m| \tag{5.4.1}$$

where the expectation value is taken with respect to (5.3.1). Any of the fields $\varphi_1, \cdots, \varphi_{|m|}$ can be put in the exponent, since they are interchangeable. Explicit computation of (5.4.1) leads to the Fermi liquid result

$$(\tau_2 - \tau_1)^{-S} \qquad ; \qquad S = 1. \tag{5.4.2}$$

5.4.2 Mapping to the fractional case

We next wish to extend the C-S mapping procedure to include the result for the electron propagator at the edge (5.4.1). This problem is nontrivial since the amount of statistical flux that is trapped now depends on whether the path from time τ_1 to τ_2 is taken just outside or just inside the sample. Hence, the obvious translation of (5.4.1) into composite fermion language,

$$\left\langle \exp -i \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \; [\partial_0 \varphi_j - a_0](\tau, \vec{x}_0) \right\rangle , \tag{5.4.3}$$

now has an additional *geometrical* significance. In order to find out what this geometrical significance is, we must repeat the steps (5.3.3–5.3.9), but now in the presence of (5.4.3). We write $a_0 = a_- + iva_x$. The presence of a_- at $\vec{x}_0$ modifies the constraint equation (5.3.4) and we get

$$\nabla \times \left[\vec{a} + 2p \; \mathrm{sgn}(m) \theta(y) \sum_i (\vec{a} - \vec{D}\varphi_i) \right] = -2p \cdot 2\pi \delta(\vec{x} - \vec{x}_0) L \tag{5.4.4}$$

where we have defined a quantity L such that

$$\beta \sum_n L_{-n} f_n(\vec{x}) = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau f(\tau, \vec{x}) \tag{5.4.5}$$

for any bosonic function f. The factor of 2π in (5.4.4) originates from our convention $\hbar=1$, while a flux quantum is given by h/e, and the minus sign shows that the statistical flux indeed points in the negative z direction. From (5.4.4) we see that the amount of flux is given by

$$2p \text{ for } y_0 < 0 \quad ; \quad 2p\nu/m \text{ for } y_0 > 0. \tag{5.4.6}$$

This is an important result. It shows that for $y_0 > 0$ *the C-S mapping has altered the charge of the tunnelling particle* from 1 into ν/m. A physical interpretation is readily given in terms of Laughlin's gauge argument. If one flux quantum is sent through a hole in the IQH sample, one electron will leave the edge per filled Landau level, resulting in a total charge m. The C-S mapping changes the transported charge, which is equal to the Hall conductance, to ν. Since this charge is contained in m Landau levels, the charge per Landau level has to be ν/m.

We will now integrate out the C-S gauge field. The general solution to the constraint equation (5.4.4) is

$$\vec{a}(\vec{x}) = \tfrac{2p\,\mathrm{sgn}(m)}{2pm\theta(y)+1}\left[\theta(y)\sum_i \vec{D}\varphi_i + \nabla\Omega + \mathrm{sgn}(m)L\nabla\mathrm{arctg}\tfrac{x-x_0}{y-y_0}\right] \tag{5.4.7}$$

with Ω again an arbitrary gauge. Substituting this solution $\vec{a}$ into what remains of the action, we re-obtain the action (5.3.9), but now with a slightly different definition of the $\tilde{\varphi}$ fields,

$$\tilde{\varphi}_i(\vec{x}) = \varphi_i(\vec{x}) - 2p\,\mathrm{sgn}(m)\Omega(\vec{x}) - 2pL\;\mathrm{arctg}\tfrac{x-x_0}{y-y_0}. \tag{5.4.8}$$

See appendix 5B for the details of this calculation. The exp in (5.4.3) acquires the following form after the C-S mapping

$$\exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\left\{\partial_0\tilde{\varphi}_i - \theta(y_0)2p\tfrac{\nu}{|m|}\sum_j D_0\tilde{\varphi}_j\right\}(\tau,\vec{x}_0). \tag{5.4.9}$$

For further considerations we put $A_\mu=0$, since the tunnelling is done without external fluctuations. Several remarks need to be made here. First of all, we see that the outcome will indeed not depend on the gauge Ω. Secondly, the expectation value will be different for $\vec{x}_0$ going to the edge from inside or from outside of the sample. Furthermore, there is a difference between the cases $m>0$ and $m<0$. If we take $y_0\uparrow 0$, we easily obtain the result (see appendix 5C)

$$\left\langle \exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\;\partial_0\tilde{\varphi}_i\right\rangle = (\tau_2-\tau_1)^{-S_{\mathrm{out}}} \quad ; \quad S_{\mathrm{out}} = 2p+1+\left(\tfrac{1}{m}-\tfrac{1}{|m|}\right) \tag{5.4.10}$$

which is the tunnelling exponent found by [37]. For positive m we have $S_{\mathrm{out}}=K_{ii}=$odd, which clearly indicates a composite fermion. For negative m the meaning of the exponent is not clear; it does not in general represent a fermionic quantity. For $y_0\downarrow 0$ we can rewrite (5.4.9) as

$$\exp\left[-i\;\mathrm{sgn}(m)\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\;\partial_0\sum_j K_{ij}^{-1}\tilde{\varphi}_j(\tau,\vec{x}_0)\right] \tag{5.4.11}$$

Expression (5.4.11) gives us a tunnelled charge equal to $\mathrm{sgn}(m)\sum_j K_{ij}^{-1}=\nu/m$ in accordance with (5.4.6). Note that the charge is not contained in one Landau channel, as was the case in the tunnelling expression for the integer effect, but instead spread out over all the channels. In the i'th channel there is a charge $1-2p\nu/|m|$, in all the other channels $-2p\nu/|m|$. It is readily seen that the tunnelling exponent becomes

$$S_{\rm in} = 1 - 2p\tfrac{\nu}{m} + \left(\tfrac{1}{m} - \tfrac{1}{|m|}\right). \tag{5.4.12}$$

For positive m this expression takes the form $S_{\rm in}=K^{-1}_{ii}=1-2p\nu/m$.

5.4.3 Duality

Notice that the cases $y_0\uparrow 0$ and $y_0\downarrow 0$ are related by a 'T-duality' in the sense of [5], with K playing the role of compactification radius. If we define 'dual' fields ξ_i with charge ν/m as $\xi_i=\mathrm{sgn}(m)\ K^{-1}_{ij}\tilde\varphi_j$, then the action (5.3.9) takes the form

$$\begin{aligned} S[\xi] \;\;=\;\; & \tfrac{i\nu}{4\pi}\int A\wedge \mathrm{d}A - \tfrac{i}{4\pi}\sum_{ij} K_{ij}\oint \left(D_x\xi_i D_-\xi_j - \tfrac{\nu}{m}\xi_i E_x\right) \\ & -\tfrac{1}{2}(\tfrac{1}{2\pi})^2 \sum_{ij}\int_\infty U_0(\vec{x}-\vec{x}')\ \nabla\times[\theta(y)\vec{D}\xi_i(\vec{x})]\ \nabla'\times[\theta(y')\vec{D}\xi_j(\vec{x}')], \end{aligned} \tag{5.4.13}$$

with $D_\mu\xi_i=\partial_\mu\xi_i-\frac{\nu}{m}A_\mu$. The exponent (5.4.11) is written as

$$\exp[-i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\ \partial_0\xi_i], \tag{5.4.14}$$

i.e. the same form as (5.4.10) but now in terms of ξ. The action (5.4.13) for the ξ fields contains the matrix K instead of K^{-1}. In [5] it was shown that the energy spectrum for a system of 1+1D coupled bosons is invariant under $\mathrm{O}(m,m,\mathbb{Z})$ transformations acting on a specific $2m\times 2m$ matrix that expresses the Hamiltonian in terms of the momenta and winding numbers of a state. The replacement $K\to K^{-1}$ is an example of such an $\mathrm{O}(m,m,\mathbb{Z})$ transformation. In a certain way it interchanges fluxes and particles.

In the above considerations the statistics parameter is calculated of a particle with charge one; the amount of statistical flux that it acquires is either $2p$ (outside the sample) or $2p\nu/m$ (inside). If we were, on the other hand, to concentrate on the flux instead of the charge, we would consider the case of $2p\nu/m$ statistical flux quanta which acquire *charge*, namely 1 inside the sample and ν/m outside. In this sense there is charge-flux duality. Notice that the roles of inside/outside get reversed in the argument given above. We make this more explicit by defining

$$p' = -p\tfrac{\nu}{m}. \tag{5.4.15}$$

The duality relation $K\to K^{-1}$ can be written as

$$K(p',m) = K^{-1}(p,m). \tag{5.4.16}$$

In Fig. 5.2 it is shown that an exchange of p to p' interchanges the opposite sides of the edge. We will come back to this point in section 5.5.2.

In the case of the simple Laughlin fractions $\nu=1/(2p+1)$ the duality is particularly transparent. The fact that there is only one edge channel means that there is no redistribution of charge over various channels, which allows for an easy interpretation of the

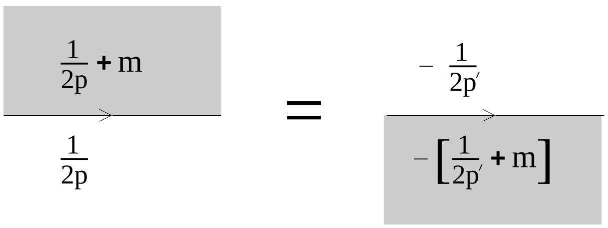

Figure 5.2: *Inverse number of statistical flux quanta inside and outside the sample. The shaded area denotes the inside. The duality $p \leftrightarrow p'$ reverses the role of the opposite sides of the edge.*

tunnelling exponent in terms of flux-charge composites. Let us consider the complex conjugate of (5.4.11); taking the complex conjugate does not change the tunnelling exponent, but it changes a particle to a hole. Expression (5.4.11) becomes $\exp -i\int d\tau(-\nu)\partial_0\tilde{\varphi}$, describing the tunnelling of a hole of charge $-\nu$. The tunnelling exponent is given by $S_{\rm in} = \nu$. We can interpret this result as follows. The hole has a statistical flux $-2p\nu$ attached to it. In addition, it has an 'intrinsic' flux of $-\nu$ coming from the fermionic statistics of the electron. The statistical parameter of a charge-flux composite is given by the product of charge and flux, in this case $(-\nu)\cdot(-2p\nu - \nu) = \nu$ for the quasihole. The relation $K \to K^{-1}$ takes the simple form $\nu^{-1} \to \nu$.

5.5 Long range disorder

5.5.1 Separation of edge channels; Edge particles

In chapter 4 we discussed how long range disorder can cause the edge states of different Landau levels to become spatially separated. A potential fluctuation at the edge generically lifts all states in such a way that new 'edges' are created. If the chemical potential lies between the shifted and unshifted energy of a Landau level, the edge states of this Landau level will be situated inside the sample, not on the outermost edge. If there are several potential jumps of this kind, all the edge channels can become separated. Viewed from above, the edge channels will mostly lie closely together at the edge, while at certain points one or more of them go venturing into the bulk.

This picture leads us to the idea of describing the chiral bosons by *one* field $\varphi(\vec{x})$ that lives on $|m|$ edges instead of $|m|$ fields that live on one edge. The action for the integer effect (5.3.1) then becomes

$$\begin{aligned} S &= \tfrac{i}{4\pi}\left[\int nA^\alpha \wedge dA^\alpha - \operatorname{sgn}(m)\sum_{a=1}^{|m|}\oint_a (D_x\varphi D_-\varphi - E_x\varphi)\right] \\ &\quad -\tfrac{1/2}{(2\pi)^2}\int_{x,x'} U_0(\vec{x}-\vec{x}')\nabla\times[n(\vec{x})\vec{D}\varphi(\vec{x})]\nabla'\times[n(\vec{x}')\vec{D}\varphi(\vec{x}')] \qquad (5.5.1)\end{aligned}$$

where $n(\vec{x})$ stands for the local filling: outside the sample $n(\vec{x})$ is zero; every time you cross an edge it changes its value by one, until it reaches its bulk value m. The $\oint_a$ denotes integration over the a'th edge. Every edge is given a label $a \in \mathbb{Z}$, equal to the maximum filling that the edge is bordering on. For positive m we have $a \in \{1, \cdots, m\}$, for negative m we have $a \in \{m+1, \cdots, 0\}$.

So we now have regions in which the integer effect occurs at different values, separated by edges of the $\nu = 1$ type on which the boson field φ lives. The C-S procedure simply maps the filling fractions in these regions from $n(\vec{x})$ to a local filling fraction $\nu(\vec{x})$,

$$\nu(\vec{x}) = n(\vec{x})/[2pn(\vec{x}) + 1]. \tag{5.5.2}$$

The mapped action reads

$$\begin{aligned} S \;=\; & \tfrac{i}{4\pi}\left[\int \nu A \wedge \mathrm{d}A - \sum_{a=1}^{|m|} \triangle\nu_{[a]} \oint_a (D_x\Phi D_-\Phi - E_x\Phi)\right] \\ & - \tfrac{1/2}{(2\pi)^2}\int_{x,x'} U_0(\vec{x}-\vec{x}')\nabla\times[\nu(\vec{x})\vec{D}\Phi(\vec{x})]\nabla'\times[\nu(\vec{x}')\vec{D}\Phi(\vec{x}')]. \end{aligned} \tag{5.5.3}$$

[A redefinition of the type (5.4.8) has occurred and the new field is denoted as Φ]. Integration over the a'th edge counting from the vacuum is written as $\oint_a$. We use the label a both for edges and for regions: the region lying 'above' the a'th edge also carries the label a. The $\triangle\nu_{[a]}$ is defined as $\nu_a - \nu_{a-1}$ for positive m and $\nu_{-a} - \nu_{-(a-1)}$ for negative m; here ν_a stands for the local filling fraction in region number a. In Fig. 5.3 we have made an overview (for samples with both signs of m in the same figure) of the parameters a, ν_a and $\triangle\nu_a$ in the simple case $p = 1$.

The tunnelling term is simply a generalisation of (5.4.9),

$$\left\langle \exp -i\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, [1 - 2p\nu(\vec{x}_0)]\partial_0\Phi(\tau,\vec{x}_0)\right\rangle, \tag{5.5.4}$$

where the expectation value is taken with respect to (5.5.3). The result again depends on the direction from which the edge is approached. The tunnelled charge is given by $q_a = \nu_a/a$ if an edge is approached from region number a. Evaluating (5.5.4) is a nontrivial exercise, since the Coulomb interaction couples the edges to each other.

If we imagine the edges to be far apart, only the intra-edge interaction is still present, drastically simplifying the calculation. Tunnelling exponents are then given by the simple expression $S = q^2/\triangle\nu$. Labelling an approach from above by 'H' and from below by 'L', we write the tunnelling exponents at the a'th edge ($a \neq 0$) as follows

$$\begin{aligned} S_a^{\mathrm{L}} &= 1 + 2p\tfrac{\nu_{a-1}}{a-1} = \tfrac{a}{\nu_a}\cdot\tfrac{\nu_{a-1}}{a-1} \\ S_a^{\mathrm{H}} &= 1 - 2p\tfrac{\nu_a}{a} = \tfrac{\nu_a}{a}\cdot\tfrac{a-1}{\nu_{a-1}} \qquad ; \qquad S_a^{\mathrm{L}}S_a^{\mathrm{H}} = 1. \end{aligned} \tag{5.5.5}$$

The edge $a = 0$ represents the outer boundary of a negative-m sample. Here we get

$$S_0^{\mathrm{H}} = S_0^{\mathrm{L}} = 2p - 1. \tag{5.5.6}$$

(See Fig. 5.3). Notice that in this approach, with spatially separated edge channels, the vacuum exponent for $m < 0$ is always *fermionic*. Notice also that the channel labelled $a = 0$, lying between the vacuum and an $m < 0$ sample, is the only one that needs long-ranged Coulomb interactions in order to be stable. In Fig. 5.3 it is seen that this channel has a chirality opposite to all the others.

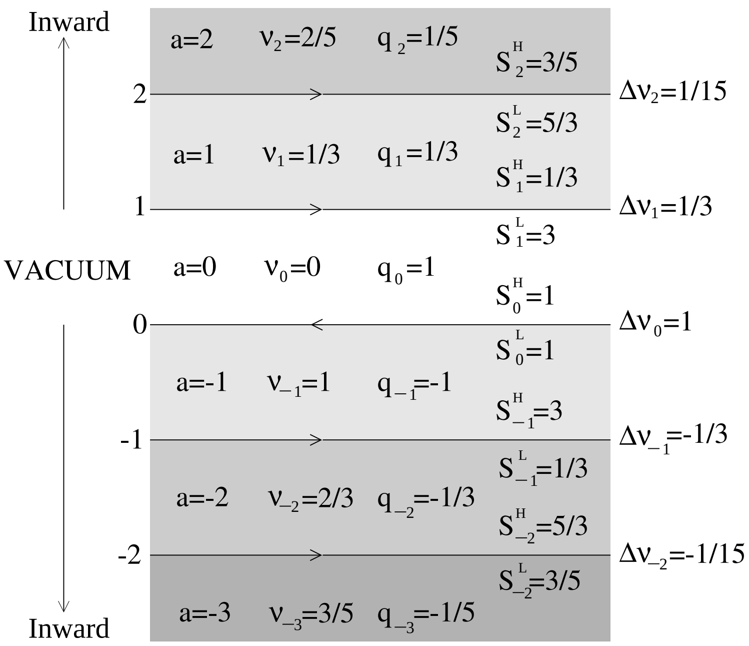

Figure 5.3: *Edge quasiparticle duality for $2p=2$. A schematic drawing is given of the regions with different $\nu(\vec{x})$; systems with $m<0$ and $m>0$ are combined in one diagram. Arrows represent the chirality of an edge. Given are the local filling fraction, the charge of the tunnelling particle as well as the tunnelling exponent S on both sides of each edge.*

5.5.2 Enlarged duality

As mentioned in section 5.4.2, the Hamiltonian of the chiral boson model has a duality symmetry under $K \to K^{-1}$. In the case of spatially separated edge channels, this duality transformation has an extra geometrical interpretation: *The order of the edge channels gets reversed.* The transformation $K \to K^{-1}$ can be accomplished by transforming $p \to p'$, with

$$2p' = -2p\nu/m. \tag{5.5.7}$$

The easiest way to visualise the reversal of channels is to consider how the transformation (5.5.7) affects the amount of statistical flux that a tunnelling particle carries (see Fig. 5.4). The inverse number of flux quanta is $1/2p$ outside the sample and increases with unit steps every time an edge is crossed, until the value $m+1/2p$ is reached in the bulk proper. Writing (5.5.7) as $-\frac{1}{2p'} = \frac{1}{2p} + m$, we see that

$$-(\tfrac{1}{2p'} + a) = \tfrac{1}{2p} + (m-a). \tag{5.5.8}$$

The minus sign preserves the chirality of the edge particles. More generally, under a transformation

$$p \to p' = -p\nu_b/b, \qquad b \in \mathbb{Z} \tag{5.5.9}$$

the whole structure of Fig. 5.3, including region labels, edge labels, filling fractions and tunnelling exponents, gets mirrored in the line $a = a_{\mathrm{mirror}}$,

$$a_{\mathrm{mirror}} = \frac{b+1}{2}. \tag{5.5.10}$$

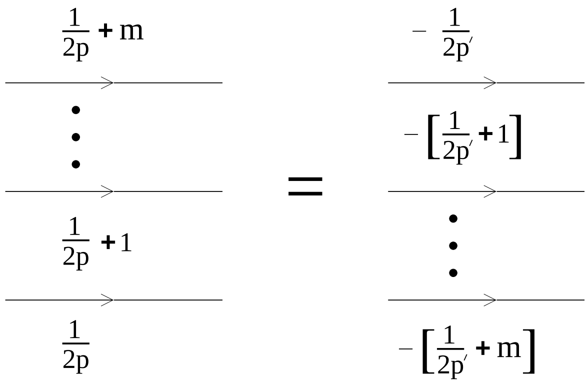

Figure 5.4: *Geometrical interpretation of the duality relation $K \to K^{-1}$ as a reversal of edge channels. For each region the inverse number of statistical flux quanta is given.* $(m > 0)$

This results in the following relations between tunnelling exponents at different edges:

$$\begin{array}{rcl} S^{\mathrm{H}}_a(p') = S^{\mathrm{L}}_{b+1-a}(p) & ; & S^{\mathrm{L}}_a(p') = S^{\mathrm{H}}_{b+1-a}(p) \\ a \neq 0 & ; & b+1-a \neq 0. \end{array} \tag{5.5.11}$$

Unless $b = -1$, expression (5.5.11) does not work for the mirrored tunnelling exponent of the counter-flowing edge $a = 0$ and its mirror image $a = b+1$.

$$S^{\mathrm{H}}_{b+1}(p') = S^{\mathrm{L}}_{b+1}(p') = 1 - 2p = -S_0(p) \tag{5.5.12}$$

$$S_0(p') = -(1 + 2p\nu_b/b) = \begin{cases} -S^{\mathrm{L}}_{b+1}(p) & \text{for } b \neq -1 \\ S_0(p) & \text{for } b = -1 \end{cases} \tag{5.5.13}$$

There are two special cases:

- $b = -1$, where the symmetry axis lies exactly on the counter-flowing channel. Every region gets mapped into minus itself, while the flux is left unchanged! $(p' = p)$. This shows that at given p, all tunnelling exponents are symmetric under $(a \to -a$, $\mathrm{L} \leftrightarrow \mathrm{H})$. The vacuum gets mapped into the $a = -1$ region, which is suggestive of an interpretation of this region as a filled $\nu = 1/(2p-1)$ 'vacuum' from which holes can be created.

- $b = 0$, with the symmetry axis at $a = 1/2$. We have $p' = -p$, so we are in a situation with reversed magnetic field. This transformation relates samples with composite fermion filling m to those with $-m$.

5.5.3 Relation between dirty and clean edges

In the previous sections we considered the various edge modes to be far apart. The results for the exponents (5.5.5) are valid provided that one probes the system at length scales short relative to the distance between the edge modes. In this section, we consider the action (5.5.3) in the opposite limit of large distances. The edge modes then strongly interact and (5.5.3) now becomes a Luttinger liquid theory of a sharp, 'clean' edge. It is not a priori obvious that this theory is physically the same as the theory of the sharp 'dirty' edge. In the following we show that this is indeed the case. In particular, the two

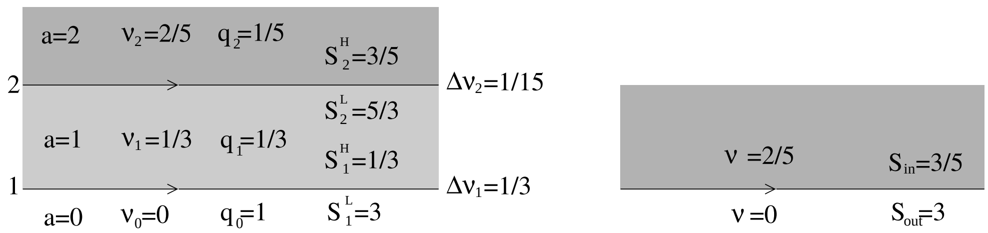

Figure 5.5: *Clean and dirty scenario for the $\nu = 2/5$ edge. Left: spatially separated channels. Right: result of the K-matrix theory. Listed are filling fractions, charges and tunnelling exponents.*

theories yield identical results for the two kinds of tunnelling exponent that can be found from (5.4.9) in the 'dirty' $\tilde{\varphi}$-theory, namely the case where the limit to the edge is taken from the vacuum and the case where one approaches the edge from inside the sample.

If one writes Φ_i for the field Φ evaluated on the i'th edge then it is possible to express the fields Φ_i in terms of a charged mode Γ and neutral modes $\gamma_1, \cdots, \gamma_{|m|-1}$ in such a way that the action (5.5.3) for the Φ_i exactly matches the action (5.3.9) for the $\tilde{\varphi}_i$. (In principle, the charged and neutral modes are not locally defined fields, since they contain parts from different edges. However, in the long distance limit all the edges are compressed together on one contour and the fields are local). The relation between the Φ_i and the $\tilde{\varphi}_i$ is nontrivial; we show that the fields are dual to each other for $m>0$ and explain why this leads to identical numbers when one calculates tunnelling exponents according to the clean picture (5.5.4) (see Fig. 5.3) or the dirty scenario (5.4.9); we also show that there is some extra structure for negative m which is the source of the differing exponents as obtained from (5.4.9) versus (5.5.4) at $m<0$.

Example for positive m: $\nu = 2/5$

First we give a simple example for $\nu = 2/5$ ($m = 2$, $p = 1$). In both $\tilde{\varphi}$ and Φ-theory the tunnelling exponents for this filling are given by $S_{\text{out}} = 3$, $S_{\text{in}} = 3/5$ (see Fig. 5.5). For $\nu = 2/5$ the $\tilde{\varphi}$-action is given by (5.3.9), with the K-matrix as defined in (5.3.8). Written explicitly we have for $A_\mu = 0$

$$S_K = -\tfrac{i}{4\pi} \sum_{ij} \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{3}{5} \end{pmatrix}_{ij} \oint \partial_x \tilde{\varphi}_i \partial_- \tilde{\varphi}_j - \tfrac{1}{2}(\tfrac{1/5}{2\pi})^2 \sum_{ij} \oint_{xx'} U_0(x-x')\ \partial_x \tilde{\varphi}_i(x) \partial_{x'} \tilde{\varphi}_j(x'). \tag{5.5.14}$$

The diagonalised form (5.3.15) of this action is given by

$$S_K = -\tfrac{i}{4\pi} \oint \left[\tfrac{2}{5}\partial_x \Gamma \partial_- \Gamma + \tfrac{1}{2}\partial_x g \partial_- g\right] - \tfrac{1}{2}(\tfrac{2/5}{2\pi})^2 \oint_{xx'} U_0(x-x')\ \partial_x \Gamma(x) \partial_{x'} \Gamma(x') \tag{5.5.15}$$

where the charged and neutral mode are defined in (5.3.14) and in this case read

$$\begin{pmatrix} g \\ \Gamma \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \tilde{\varphi}_1 \\ \tilde{\varphi}_2 \end{pmatrix} \quad ; \quad \begin{pmatrix} \tilde{\varphi}_1 \\ \tilde{\varphi}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 1 \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} g \\ \Gamma \end{pmatrix}. \tag{5.5.16}$$

On the other hand, the action for the Φ-field on the separated edges is given by (5.5.3). If we compress all the edges together we get (again for $A_\mu = 0$)

$$S_{\text{sep}} \ = \ -\tfrac{i}{4\pi} \oint \left[\tfrac{1}{3}\partial_x \Phi_1 \partial_- \Phi_1 + \tfrac{1}{15}\partial_x \Phi_2 \partial_- \Phi_2\right]$$

$$-\tfrac{1}{2}(\tfrac{1}{2\pi})^2 \oint_{xx'} U_0(x-x')\partial_x(\tfrac{1}{3}\Phi_1 + \tfrac{1}{15}\Phi_2)\partial_{x'}(\tfrac{1}{3}\Phi_1 + \tfrac{1}{15}\Phi_2). \qquad (5.5.17)$$

We define a neutral mode γ and charged mode Γ for the Φ-theory as follows

$$\begin{pmatrix}\gamma\\ \Gamma\end{pmatrix} = \begin{pmatrix}1 & -1\\ \frac{5}{6} & \frac{1}{6}\end{pmatrix}\begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} \quad ; \quad \begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} = \begin{pmatrix}\frac{1}{6} & 1\\ -\frac{5}{6} & 1\end{pmatrix}\begin{pmatrix}\gamma\\ \Gamma\end{pmatrix}. \qquad (5.5.18)$$

Whereas the $\tilde{\varphi}$ fields are all equivalent in the dirty edge theory, here the Φ_i are not, since the spatially separated edges have unequal $\triangle\nu$ values which appear in the definition of Γ. In terms of Γ and γ the Φ-action (5.5.17) takes the form

$$S_{\text{sep}} = -\tfrac{i}{4\pi}\oint \left[\tfrac{2}{5}\partial_x\Gamma\partial_-\Gamma + \tfrac{1}{18}\partial_x\gamma\partial_-\gamma\right] - \tfrac{1}{2}(\tfrac{2/5}{2\pi})^2 \oint_{xx'} U_0(x-x')\ \partial_x\Gamma\partial_{x'}\Gamma \qquad (5.5.19)$$

Comparing the two results (5.5.15) and (5.5.19), we see that there is complete equivalence given the relation

$$g = \tfrac{1}{3}\gamma. \qquad (5.5.20)$$

Having shown that the actions are equivalent, we now want to compare tunnelling operators. In order to do this, we in principle only need to rewrite the $\tilde{\varphi}$ and Φ that appear in the various operators in terms of charged and neutral modes, and then use (5.5.15) or (5.5.19). However, much insight can be gained by also expressing $\tilde{\varphi}$ and Φ in terms of each other. Combining (5.5.16), (5.5.18) and (5.5.20), we find

$$\begin{pmatrix}\tilde{\varphi}_1\\ \tilde{\varphi}_2\end{pmatrix} = \begin{pmatrix}1 & 0\\ \frac{2}{3} & \frac{1}{3}\end{pmatrix}\begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} \quad ; \quad \begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} = \begin{pmatrix}1 & 0\\ -2 & 3\end{pmatrix}\begin{pmatrix}\tilde{\varphi}_1\\ \tilde{\varphi}_2\end{pmatrix}. \qquad (5.5.21)$$

In the iqHe φ-theory we can choose which one of the fields φ_1 or φ_2 to put in the tunnelling operator. Eq. (5.4.9) then gives four possible operators, which we list in the table below, giving only the 'X' in the expression $\exp -iX|_{\tau_1}^{\tau_2}$.

Dirty	$\tilde{\varphi}$-basis	Diag. basis
Vacuum	$\varphi_1 \to \tilde{\varphi}_1$	$\Gamma + \frac{1}{6}\gamma$
	$\varphi_2 \to \tilde{\varphi}_2$	$\Gamma - \frac{1}{6}\gamma$
Bulk	$\varphi_1 \to (K^{-1}\tilde{\varphi})_1 = \frac{3}{5}\tilde{\varphi}_1 - \frac{2}{5}\tilde{\varphi}_2$	$\frac{1}{5}\Gamma + \frac{1}{6}\gamma$
	$\varphi_2 \to (K^{-1}\tilde{\varphi})_2 = -\frac{2}{5}\tilde{\varphi}_1 + \frac{3}{5}\tilde{\varphi}_2$	$\frac{1}{5}\Gamma - \frac{1}{6}\gamma$

There are only two corresponding operators in the Φ-theory. An approach from the vacuum can only involve Φ_1 because the other field does not live on the outermost edge. In the same way, the bulk operator can only contain Φ_2. From (5.5.4) we get

Clean	Φ-basis	$\tilde{\varphi}$-basis	Diag. basis
Vacuum	Φ_1	$\tilde{\varphi}_1$	$\Gamma + \frac{1}{6}\gamma$
Bulk	$\frac{1}{5}\Phi_2$	$-\frac{2}{5}\tilde{\varphi}_1 + \frac{3}{5}\tilde{\varphi}_2$	$\frac{1}{5}\Gamma - \frac{1}{6}\gamma$

In all cases the exponents obtained from the operators in these tables are $S_{\text{out}} = 3$, $S_{\text{in}} = 3/5$. There is no a priori reason why these numbers should be the same as those appearing in Fig. 5.5. In the short distance theory the actions for the two edges are uncoupled, while we have now used the action (5.5.15) where they are coupled.

Generalisation for $m>0$

The analysis for $\nu=2/5$ is easily generalised to all filling fractions resulting from positive m. Let us redo all the steps in the same order. The K-matrix action for the $\tilde{\varphi}_i$ is given by (5.3.9). Its diagonal form is given by (5.3.15) with $\text{sgn}(m)=+1$ and Γ, $g_1,\cdots,g_{m-1}$ as defined in (5.3.14). For the Φ-theory we define the charged and neutral modes as follows

$$\Gamma = \tfrac{1}{\nu}\sum_{a=1}^{m}\triangle\nu_a\Phi_a \quad ; \quad \gamma_k = \tfrac{1}{\nu_k}\sum_{a=1}^{k}\triangle\nu_a\Phi_a - \Phi_{k+1}$$
$$\Phi_a = \Gamma + \sum_{k=a}^{m-1}\tfrac{\triangle\nu_{k+1}}{\nu_{k+1}}\gamma_k - \tfrac{\nu_{a+1}}{\nu_a}\gamma_{a-1}. \tag{5.5.22}$$

The diagonalised form of the action reads

$$S = -\tfrac{i}{4\pi}\oint\left[\nu\partial_x\Gamma\partial_-\Gamma + \sum_{k=1}^{m-1}\tfrac{k}{k+1}q_k^2\partial_x\gamma_k\partial_-\gamma_k\right] - \tfrac{\nu^2}{8\pi^2}\oint_{xx'} U\ \partial_x\Gamma\partial_{x'}\Gamma \tag{5.5.23}$$

The only difference with (5.3.15) is the factor q_k^2 in front of the neutral modes. The actions for $\tilde{\varphi}$ and Φ are identical if we define

$$g_k = q_k\gamma_k. \tag{5.5.24}$$

Using (5.3.14), (5.5.22) and (5.5.24) we are now able to relate the Φ_i to the $\tilde{\varphi}_i$,

$$q_k\Phi_k = \sum_{a=1}^{k}(K^{-1}_{k\times k})_{ka}\tilde{\varphi}_a. \tag{5.5.25}$$

This is a remarkable result. It shows that the fields are dually related. Notice that the K-matrix appearing in (5.5.25) does not have the full $m\times m$ size. In order to find the k'th Φ-field one has to use the K-matrix for a truncated system which has only k edges, i.e. $(K^{-1}_{k\times k})_{ij}=\delta_{ij}-2pq_k$. If one wants to write (5.5.25) as a vector equation of the form $(q\Phi)_i=M_{ij}\tilde{\varphi}_j$, the M is *not* given by K^{-1} but by a stacking together of rows taken from matrices $K^{-1}_{k\times k}$ of increasing size.

Given below are those tunnelling operators for the clean and dirty scenario that describe the same tunnelling processes.

	$\tilde{\varphi}$-theory	Φ-theory
Vacuum	$\varphi_k\rightarrow\tilde{\varphi}_k$	$\Phi_1=\tilde{\varphi}_1$
Bulk	$\varphi_k\rightarrow(K^{-1}\tilde{\varphi})_k$	$q_m\Phi_m=(K^{-1}\tilde{\varphi})_m$

From appendix 5C we know that the $\tilde{\varphi}_1,\cdots,\tilde{\varphi}_m$ all give the same result $S_{\text{out}}=2p{+}1$. That settles it for the vacuum operators. The beautiful thing about (5.5.25) is that it ensures that the bulk operator $q_m\Phi_m$ takes the form of the truly dual field $(K^{-1}\tilde{\varphi})_m$ involving the full K-matrix. From (5.4.12) we know that all the dual fields yield the same exponent $S_{\text{in}}=1-2pq_m$.

So now we have shown for $m>0$ that the long distance limit of the short distance Φ-theory indeed reproduces the physics of the $\tilde{\varphi}$-theory. But why does the Φ-theory at positive m also give the same exponents *before the long distance limit is taken*? The

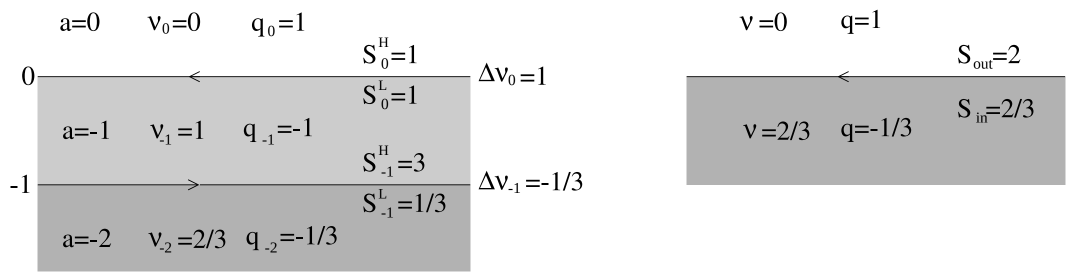

Figure 5.6: *Clean and dirty scenario for the $\nu = 2/3$ edge. Left: spatially separated counter-flowing channels. Right: K-matrix theory. Listed are filling fractions, charges and tunnelling exponents.*

answer is again provided by (5.5.25). Eq. (5.5.5) shows that the exponent on the isolated m'th edge is equal to $q_m^2/\triangle\nu_m = q_m/q_{m-1}$. On the other hand, the long distance picture leads to operators $\exp -i(K^{-1}\tilde{\varphi})_i$ in terms of fields dual to $\tilde{\varphi}_i$ which have coupling matrix K instead of K^{-1}. This leads to an exponent K^{-1}_{ii}, which is exactly equal to q_m/q_{m-1}.

Example for negative m: $\nu = 2/3$

For $m<0$ the relation between the clean and dirty scenario is more subtle. In order to get the flavour, let us first consider the easiest example, $\nu = 2/3$. In Fig. 5.6 we have shown the relevant piece of Fig. 5.3 as compared to the K-matrix picture. Our starting point for the $\tilde{\varphi}$-theory is again (5.3.9). The explicit action for $A_\mu = 0$ is given by

$$S_K = -\tfrac{i}{4\pi}\sum_{ij}\begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3}\end{pmatrix}_{ij}\oint \partial_x\tilde{\varphi}_i\partial_-\tilde{\varphi}_j - \tfrac{1}{2}(\tfrac{1/3}{2\pi})^2\sum_{ij}\oint_{x,x'} U\partial_x\tilde{\varphi}_i\partial_{x'}\tilde{\varphi}_j \tag{5.5.26}$$

and after diagonalisation it takes the form

$$-\tfrac{i}{4\pi}\oint\left[\tfrac{2}{3}\partial_x\Gamma\partial_-\Gamma - \tfrac{1}{2}\partial_x g\partial_- g\right] - \tfrac{1}{2}(\tfrac{2/3}{2\pi})^2\oint_{x,x'} U\partial_x\Gamma\partial_{x'}\Gamma. \tag{5.5.27}$$

The definition of the charged and neutral modes again follows from (5.3.14),

$$\begin{pmatrix} g \\ \Gamma\end{pmatrix} = \begin{pmatrix} 1 & -1 \\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\tilde{\varphi}_1 \\ \tilde{\varphi}_2\end{pmatrix} \quad ; \quad \begin{pmatrix}\tilde{\varphi}_1 \\ \tilde{\varphi}_2\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & 1 \\ -\frac{1}{2} & 1\end{pmatrix}\begin{pmatrix} g \\ \Gamma\end{pmatrix}. \tag{5.5.28}$$

The action (5.5.3) for the Φ-theory with compressed edges is now given by

$$S_{\rm sep} = -\tfrac{i}{4\pi}\oint\left[\partial_x\Phi_1\partial_-\Phi_1 - \tfrac{1}{3}\partial_x\Phi_2\partial_-\Phi_2\right] - \tfrac{1/2}{(2\pi)^2}\oint_{x,x'} U\partial_x(\Phi_1 - \tfrac{1}{3}\Phi_2)\partial_{x'}(\Phi_1 - \tfrac{1}{3}\Phi_2) \tag{5.5.29}$$

Charged and neutral modes are defined in a way similar to (5.5.22), taking into account the different signs of the $\triangle\nu$.

$$\begin{pmatrix}\gamma \\ \Gamma\end{pmatrix} = \begin{pmatrix} 1 & -1 \\ \frac{3}{2} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}\Phi_1 \\ \Phi_2\end{pmatrix} \quad ; \quad \begin{pmatrix}\Phi_1 \\ \Phi_2\end{pmatrix} = \begin{pmatrix}-\frac{1}{2} & 1 \\ -\frac{3}{2} & 1\end{pmatrix}\begin{pmatrix}\gamma \\ \Gamma\end{pmatrix}. \tag{5.5.30}$$

This leads to the following diagonalised form,

$$S_{\rm sep} = -\tfrac{i}{4\pi}\oint\left[\tfrac{2}{3}\partial_x\Gamma\partial_-\Gamma - \tfrac{1}{2}\partial_x\gamma\partial_-\gamma\right] - \tfrac{1}{2}(\tfrac{2/3}{2\pi})^2\oint_{x,x'} U\partial_x\Gamma\partial_{x'}\Gamma \tag{5.5.31}$$

We find that (5.5.27) and (5.5.31) are identical after defining $g = \gamma$. Using (5.5.28) and (5.5.30) we get

$$\begin{pmatrix} \tilde{\varphi}_1 \\ \tilde{\varphi}_2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} \quad ; \quad \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \tilde{\varphi}_1 \\ \tilde{\varphi}_2 \end{pmatrix}. \tag{5.5.32}$$

This relation does *not* satisfy the duality (5.5.25) for the quantities $q_{-1}\Phi_1$ and $q_{-2}\Phi_2$. (Negative indices have to be taken because the relevant regions in Fig. 5.6 are $a{=}-1$ and $a{=}-2$). Instead we have

$$\begin{aligned} q_{-1}\Phi_1 &= \tilde{\varphi}_1 - 2\Gamma \\ q_{-2}\Phi_2 &= (K^{-1}\tilde{\varphi})_2 - \tfrac{2}{3}\Gamma. \end{aligned} \tag{5.5.33}$$

Eq. (5.5.25) produces the first term on the r.h.s., an expression that has the right magnitude of the charge but the wrong sign. The $q_{-1}\Phi_1$ has charge -1, so the extra term -2Γ is precisely what is needed to flip the sign of the charge. In the same way, the term $-\frac{2}{3}\Gamma$ changes the charge from the $+1/3$ that $(K^{-1}\tilde{\varphi})_2$ yields to the $-1/3$ that is the correct value for $q_{-2}\Phi_2$.

An interesting subtlety is that even though (5.5.25) does not hold for the relation between Φ and $\tilde{\varphi}$ fields, all tunnelling exponents behave as if it did hold. The reason for this is the following. After a tunnelling operator has been expressed in terms of charged and neutral modes, it is easy to see that the only effect of the extra term involving Γ is to change the sign of the factor standing in front of Γ. This sign change has no effect whatsoever on the calculation of the tunnelling exponent. In the table below we explicitly write down all the tunnelling operators.

	$\tilde{\varphi}$-theory		Φ-theory		
	$\tilde{\varphi}$-basis	Diag. basis	Φ-basis	$\tilde{\varphi}$-basis	Diag. basis
Vacuum	$\varphi_1 \to \tilde{\varphi}_1$	$\Gamma + \frac{1}{2}\gamma$	Φ_1	$\tilde{\varphi}_2$	$\Gamma - \frac{1}{2}\gamma$
	$\varphi_2 \to \tilde{\varphi}_2$	$\Gamma - \frac{1}{2}\gamma$			
Bulk	$\varphi_1 \to \frac{1}{3}\tilde{\varphi}_1 - \frac{2}{3}\tilde{\varphi}_2$	$-\frac{1}{3}\Gamma + \frac{1}{2}\gamma$	$-\frac{1}{3}\Phi_2$	$\frac{1}{3}\tilde{\varphi}_1 - \frac{2}{3}\tilde{\varphi}_2$	$-\frac{1}{3}\Gamma + \frac{1}{2}\gamma$
	$\varphi_2 \to -\frac{2}{3}\tilde{\varphi}_1 + \frac{1}{3}\tilde{\varphi}_2$	$-\frac{1}{3}\Gamma - \frac{1}{2}\gamma$			

In both the Φ-theory and $\tilde{\varphi}$-theory the tunnelling exponents obtained from the above expressions are

$$S_{\text{out}} = 2 \quad ; \quad S_{\text{in}} = 2/3. \tag{5.5.34}$$

The discrepancy with the short distance clean scenario result $S_{\text{out}}{=}1$, $S_{\text{in}}{=}1/3$ is caused by the fact that (5.5.32) does not satisfy the duality (5.5.25).

Generalisation for arbitrary sign of m

For completeness, we now present general equations that comprise all the above cases. The $\tilde{\varphi}$-theory expressions (5.3.15) and (5.3.14) are equally valid for negative and positive m. No extra work is needed there. The Φ-theory expressions (5.5.22) and (5.5.23) do need some modification if we want to incorporate the possibility of m being negative. Eq. (5.5.22) for the charged and neutral modes can be generalised as follows,

$$\Gamma = \tfrac{1}{\nu}\sum_{a=1}^{|m|} \triangle\nu_{[a]}\Phi_a \quad ; \quad \gamma_k = \tfrac{1}{\nu_{[k]}}\sum_{a=1}^{k} \triangle\nu_{[a]}\Phi_a - \Phi_{k+1}$$
$$\Phi_a = \Gamma + \sum_{k=a}^{|m|-1} \tfrac{\triangle\nu_{[k+1]}}{\nu_{[k+1]}}\gamma_k - \tfrac{\nu_{[a+1]}}{\nu_{[a]}}\gamma_{a-1} \tag{5.5.35}$$

Here the notation $\nu_{[a]}$ (with $a>0$) means ν_a for samples with positive m and ν_{-a} for $m<0$ samples. For instance, in case of negative m we would have $\triangle\nu_{[2]}=\nu_{-2}-\nu_{-1}$. Using this definition, the generalisation of (5.5.23) becomes

$$S = -\tfrac{i}{4\pi}\oint\left[\nu\partial_x\Gamma\partial_-\Gamma + \mathrm{sgn}(m)\sum_{k=1}^{|m|-1}\tfrac{k}{k+1}q_{[k]}^2\partial_x\gamma_k\partial_-\gamma_k\right] - \tfrac{\nu^2}{8\pi^2}\oint_{xx'} U\ \partial_x\Gamma\partial_{x'}\Gamma \tag{5.5.36}$$

Comparing this expression to (5.3.15) we can make the identification

$$g_k = |q_{[k]}|\cdot\gamma_k. \tag{5.5.37}$$

This enables us to relate the fields $\tilde{\varphi}_i$ to the Φ_a.

$$\tilde{\varphi}_k = \mathrm{sgn}(m)\left[q_{[k]}\Phi_k + \sum_{a=1}^{k} 2p\triangle\nu_{[a]}\Phi_a\right] + [1-\mathrm{sgn}(m)]\cdot\Gamma \tag{5.5.38}$$

$$\begin{aligned} q_{[k]}\Phi_k &= \mathrm{sgn}(m)\tilde{\varphi}_k - 2pq_{[k]}\sum_{a=1}^{k}\tilde{\varphi}_a + [1-\mathrm{sgn}(m)]\cdot q_{[k]}\Gamma \\ &= \sum_{a=1}^{k}(K_{k\times k}^{-1})_{ka}\tilde{\varphi}_a + [1-\mathrm{sgn}(m)]\cdot q_{[k]}\Gamma \end{aligned} \tag{5.5.39}$$

where $K_{k\times k}$ again denotes the K-matrix for a system containing k edges. It is easily checked that for positive m (5.5.25) is reobtained.

Eq. (5.5.39) shows that what happened at $\nu=2/3$ happens at all negative m states: The $K^{-1}\tilde{\varphi}$ part of the tunnelling expression $q_m\Phi_{|m|}$ has the right magnitude of the charge q_m but the wrong (=positive) sign. The extra term $2q_m\Gamma$ flips the sign. The reason why we write Γ in (5.5.39) instead of expressing it in terms of the $\tilde{\varphi}_a$ is that we can now directly discern the effect on the tunnelling operators. In the diagonal basis the effect of the extra Γ term is to flip the sign of the coefficient of Γ in the tunnelling operators. Just as in the 2/3 case this leaves the exponent unaffected.

Also generalised to all $m<0$, the discrepancy between the exponents as calculated in the short and long distance limits is due to the fact that (5.5.25) does not hold,

5.5.4 Quasiparticles and the Laughlin argument

The particles probed in tunnelling experiments are not necessarily the same as those featuring in Laughlin's flux argument. In the integer case, putting a flux quantum through a hole in the sample results in the transport of one electron through each Landau level below the Fermi energy. In the language of (5.5.1), considering for simplicity $m>0$, what would happen at the a'th edge is that $a-1$ electrons enter that edge from below and that a electrons leave it on the upper side. In this way, a total of m electrons gets transported into the bulk, which is the correct number. The operator describing this process is given by

$$\exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\,\Big\{a[\partial_0\varphi-a_0](\vec{x}_{\mathrm{H}})-[a-1][\partial_0\varphi-a_0](\vec{x}_{\mathrm{L}})\Big\} \tag{5.5.40}$$

where $\vec{x}_{\mathrm{H}}$ lies infinitesimally above the a'th edge and $\vec{x}_{\mathrm{L}}$ lies below it. Both points have the same x-coordinate. Knowing the results of section 5.5.1, it is easy to do the C-S mapping. The constraint equation multiplied by a_- contains the sum of two delta functions instead of just one. Correspondingly, the redefined field Φ contains the sum of two arctg functions. The mapped expression is given by

$$\exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\,\Big\{\nu_a\partial_0\Phi(\vec{x}_{\mathrm{H}})-\nu_{a-1}\partial_0\Phi(\vec{x}_{\mathrm{L}})\Big\}\,. \tag{5.5.41}$$

As is clearly seen from the constants appearing here, the emitted charge is ν_a and the absorbed charge is ν_{a-1}. Taking all the edges together we have a total transported charge of $\sum_{a=1}^{m}\triangle\nu_a=\nu_m$, which is the correct number for a system of filling fraction ν_m. Taking the limit where $\vec{x}_{\mathrm{H}}$ and $\vec{x}_{\mathrm{L}}$ go to the same point x on the edge, the tunnelling expression takes the simple form

$$\exp -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\;\triangle\nu_a\partial_0\Phi(\tau,x), \tag{5.5.42}$$

indicating a charge $\triangle\nu_a$.

5.5.5 Continuous tunnelling exponent

Hoeveel engelen kun je op één naald prikken?
– een Leids theoreticus

In chapter 4 we discussed how the non-quantised tunnelling exponent [27] $1/\nu$ can be understood as a consequence of the Coulomb interactions between the edge modes and the localised states present in the bulk. For ν around integer filling, we explicitly showed how the interactions render the neutral modes irrelevant. Here we will generalise this result to fractional fillings, in particular to Jain's main hierarchy.

At almost integer filling $\nu=m+\varepsilon$, a fraction ε of the total area is occupied by regions with filling $m+1$, while the rest of the sample has filling m. After performing the flux attachment, a fraction $1-\varepsilon$ of the area has filling ν_m and a fraction ε has ν_{m+1}, leading to

$$\nu=m+\varepsilon\quad\stackrel{\text{C-S mapping}}{\longrightarrow}\quad\nu=\nu_m+\varepsilon(\nu_{m+1}-\nu_m). \tag{5.5.43}$$

The action for this situation is given by (5.5.3), but now including the bulk contours $C_{a>m}$. In chapter 4 we showed that in the long distance limit the bulk degrees of freedom can effectively be described by an extra chiral boson on a contour C_b near the edge, plus an edge term

$$S_{\mathrm{n}}=-\alpha\int_{C_b}\mathrm{d}k\;|k|\partial_\perp\Phi(-k)\partial_\perp\Phi(k) \tag{5.5.44}$$

(with $\partial_\perp$ the derivative normal to C_b and α a positive constant), which suppresses the neutral modes. The only difference between the integer and fractional case lies in the

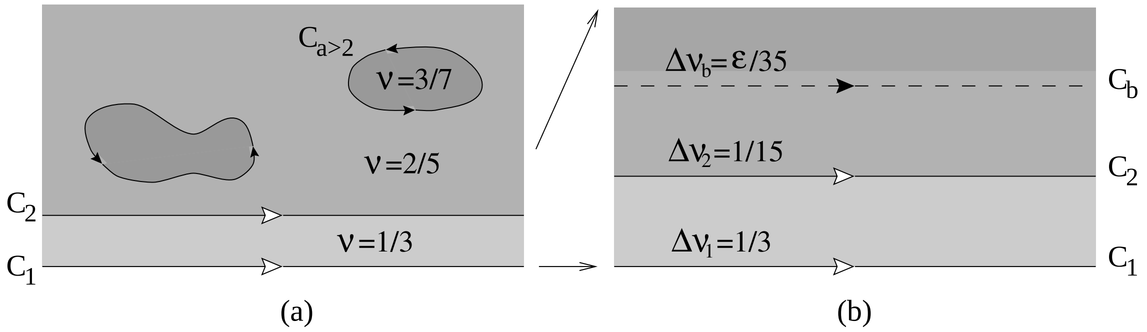

Figure 5.7: *Example of the long distance limit for filling fraction* $\nu = 2/5 + \varepsilon/35$. (a) *The* 3/7 *islands occupy a fraction* ε *of the total area.* (b) *Long distance limit. The bulk degrees of freedom are effectively represented by an extra chiral boson on the contour* C_b *with charge density* $\frac{\varepsilon}{35}\partial_x\Phi|_{C_b}$. *The filling fraction is expressed as* $\nu = \triangle\nu_1 + \triangle\nu_2 + \triangle\nu_b$.

edge parameters $\triangle\nu_1, \cdots, \triangle\nu_m$ which were all equal to 1 in the integer case and $\triangle\nu_b := \varepsilon(\nu_{m+1} - \nu_m)$ which was equal to ε. In terms of fields $\Phi_1, \ldots, \Phi_m$ and $\Phi_{m+1} := \Phi_b$ which live on *one* edge, the large wavelength limit yields the following action for the fractional case

$$\begin{aligned} S[\Phi] &= -\tfrac{i}{4\pi}\sum_{a=1}^{m+1}\delta_a \oint \partial_x\Phi_a\partial_-\Phi_a - \tfrac{1}{8\pi^2}\sum_{a,b=1}^{m+1}\delta_a\delta_b\oint_{xx'} U_0(x-x')\,\partial_x\Phi_a\partial_{x'}\Phi_b \\ &\quad -\tfrac{1}{8\pi^2}\sum_{a,b=1}^{m+1} V_{ab}\delta_a\delta_b\oint \partial_x\Phi_a\partial_x\Phi_b + S_{\mathrm{n}}. \end{aligned} \tag{5.5.45}$$

Here we have defined $\delta_a = \triangle\nu_a$ for $a = 1, \ldots, m$ and $\delta_{m+1} = \varepsilon\triangle\nu_{m+1}$. All short range effects, originating from the fact that the bulk orbitals sometimes venture near the edge, are contained in the matrix V. The definition of a charged mode Γ and neutral modes $\gamma_1, \ldots, \gamma_m$ is similar to (5.5.22),

$$\Gamma = \tfrac{1}{\nu}\sum_{a=1}^{m+1}\delta_a\Phi_a \qquad ; \qquad \gamma_k = \tfrac{1}{\nu_k}\sum_{a=1}^{k}\delta_a\Phi_a - \varphi_{k+1}. \tag{5.5.46}$$

The Φ's are expressed in terms of the γ's as follows

$$\Phi_a = \Gamma - \frac{\nu_{a-1}}{\nu_a}\gamma_{a-1} + \sum_{k=a}^{m}\frac{\delta_{k+1}}{\nu_{k+1}}\gamma_k. \tag{5.5.47}$$

Written in the new basis, the action (5.5.45) is given by

$$\begin{aligned} S[\Gamma,\gamma] &= S_{\mathrm{n}}[\gamma] - \frac{i}{4\pi}\oint\left[\nu\partial_x\Gamma\partial_-\Gamma + \sum_{k=1}^{m}\delta_{k+1}\frac{\nu_k}{\nu_{k+1}}\partial_x\gamma_k\partial_0\gamma_k\right] \\ &\quad -\frac{\nu^2}{8\pi^2}\oint_{xx'} U_0(x,x')\partial_x\Gamma\partial_{x'}\Gamma - \frac{1}{8\pi^2}\sum_{a,b=1}^{m}\oint V_{ab}\,\partial_x\gamma_a\partial_x\gamma_b. \end{aligned} \tag{5.5.48}$$

The neutral modes do not contribute to a calculation of the tunnelling d.o.s. because of (5.5.44) and consequently the parameter ν in front of the $\partial_x\Gamma\partial_-\Gamma$ term determines the tunnelling exponent. The result is $S = 1/\nu$, the same result that we obtained for the integer theory.

5.5.6 Plateau transitions

We have discussed integer plateau transitions in chapter 4. Recapitulating the results obtained there, we can sketch a transition from $\nu=m$ to $\nu=m+1$ (with m positive) as follows. One has m copies of the $\nu=1$ type Q-edge theory on the edge of the sample and one copy percolating through the bulk. Upon taking the continuum limit, the percolating 'edge' gives rise to a nonlinear sigma model with $\sigma_{\rm xx}=\frac{1}{2}$ and $\sigma_{\rm xy}=\frac{1}{2}$. Combining the edge and bulk contributions, one gets a plateau transition at $\sigma_{\rm xx}=\frac{1}{2}$, $\sigma_{\rm xy}=m+\frac{1}{2}$.

We now generalise this picture to the fractional case, in particular to the transition $\nu_n\to\nu_{n+1}$. At the boundary of the sample we have n edge channels with total contribution $(\nu_1-\nu_0)+(\nu_2-\nu_1)+\cdots+(\nu_n-\nu_{n-1})=\nu_n$ to the Hall conductance. One edge channel with parameter $\triangle\nu=\nu_{n+1}-\nu_n$ (see Fig. 5.7a) percolates through the bulk. In order to find out what kind of sigma-model arises from this percolating 'edge', we first we go from (5.5.3) back to the Q-field formalism. We write down the the action $S_{\triangle\nu}[Q,A]$ for one edge channel with parameter $\triangle\nu$,

$$\begin{aligned} S_{\triangle\nu}[Q,A] &= S_{\rm top}[Q]+\tfrac{i}{4\pi}\left[\sum_\alpha\int \nu(A^{\rm eff})^\alpha\wedge {\rm d}(A^{\rm eff})^\alpha-\triangle\nu\oint {\rm d}x\; A_x^\dagger A_c^{\rm eff}\right] \\ &\quad +\tfrac{\pi}{4\beta v_{\rm d}}S_{\rm F}[Q]-\tfrac{\pi}{4\beta}\sum_{n\alpha}\int {\rm d}k_x\,\tfrac{1}{\tilde v(k_x)}\left|{\rm tr}\;{\rm I}^\alpha_{-n}Q-\tfrac{\beta}{\pi}\sqrt{\triangle\nu}(A_c^{\rm eff})^\alpha_n\right|^2 \\ &\quad +\tfrac{\beta}{2}(\tfrac{1}{2\pi})^2\int {\rm d}^2x{\rm d}^2x'\;\nu(\vec x)B(\vec x)U_0(\vec x-\vec x')\nu(\vec x')B(\vec x') \end{aligned} \qquad (5.5.49)$$

where in the definition of $A_0^{\rm eff}$ (4.3.41) the m now has to be replaced by $\nu(\vec x)$, and in the definition of $\tilde v$ (5.3.18) one should read $\triangle\nu$ instead of ν. We have neglected here the Coulomb interaction with other channels. Notice that in (5.5.49) the charge of the Q-field is no longer 1 but $\sqrt{\triangle\nu}$. A gauge transformation takes the form

$$A_\mu\to A_\mu+\partial_\mu\chi \qquad ; \qquad Q\to e^{i\sqrt{\triangle\nu}\cdot\hat\chi}Qe^{-i\sqrt{\triangle\nu}\cdot\hat\chi}. \qquad (5.5.50)$$

Following chapter 4 we use (5.5.49) and obtain the σ model action for the percolating network. Putting $A_\mu=0$, the result for the kinetic terms is equal to

$$S_{\rm cont}[Q]=-\tfrac{1}{8}\tilde\sigma_{\rm xx}{\rm Tr}\;(\nabla Q)^2+\tfrac{1}{8}\tilde\sigma_{\rm xy}\varepsilon^{ij}{\rm Tr}\;Q\partial_iQ\partial_jQ \qquad (5.5.51)$$

where

$$\tilde\sigma_{\rm xx}=1/2 \quad ; \quad \tilde\sigma_{\rm xy}=1/2 \qquad (5.5.52)$$

now represent the mean field conductances of the *composite fermions*. The *actual* mean field values obtained from linear response in the external fields are different by a factor of $\triangle\nu$, i.e.

$$\sigma'_{\rm xx}=\tfrac{1}{2}\triangle\nu \quad ; \quad \sigma^{\rm bulk}_{\rm xy}=\tfrac{1}{2}\triangle\nu \qquad (5.5.53)$$

Adding the edge contribution to this result, we finally get for the mean field conductances at the plateau transition

$$\sigma'_{\rm xx}=\tfrac{1}{2}\cdot\tfrac{1}{(2pn+p+1)^2-p^2} \quad ; \quad \sigma'_{\rm xy}=\tfrac{1}{2}(\nu_n+\nu_{n+1})=\tfrac{(2pn+p+1)(n+1/2)-p/2}{(2pn+p+1)^2-p^2}. \qquad (5.5.54)$$

This result differs from what one would expect from acting with SL(2,$\mathbb{Z}$) on an integer plateau transition. In general, the SL(2,$\mathbb{Z}$) mapping of the conductances takes the form

$$\sigma\to\sigma'=\sigma/(2p\sigma+1), \qquad (5.5.55)$$

where the complex number σ is defined as $\sigma = \sigma_{\rm xy} + i\sigma_{\rm xx}$. In terms of $\sigma_{\rm xy}$ and $\sigma_{\rm xx}$ (5.5.55) reads

$$\sigma'_{\rm xx}(\mathrm{SL}) = \tfrac{\sigma_{\rm xx}}{(2p\sigma_{\rm xy}+1)^2+(2p\sigma_{\rm xx})^2} \quad ; \quad \sigma'_{\rm xy}(\mathrm{SL}) = \tfrac{\sigma_{\rm xy}(2p\sigma_{\rm xy}+1)+2p\sigma^2_{\rm xx}}{(2p\sigma_{\rm xy}+1)^2+(2p\sigma_{\rm xx})^2}. \tag{5.5.56}$$

Inserting the composite fermion values $\sigma_{\rm xx} = \frac{1}{2}$, $\sigma_{\rm xy} = n + \frac{1}{2}$ we now obtain

$$\sigma'_{\rm xx}(\mathrm{SL}) = \tfrac{1}{2} \cdot \tfrac{1}{(2pn+p+1)^2+p^2} \quad ; \quad \sigma'_{\rm xy}(\mathrm{SL}) = \tfrac{(2pn+p+1)(n+1/2)+p/2}{(2pn+p+1)^2+p^2}. \tag{5.5.57}$$

The different results (5.5.54) and (5.5.57) indicate that the C-S mapping is not uniquely described by SL(2,$\mathbb{Z}$) but depends on the details of disorder.

5A Integrating out the Chern-Simons field

The remnant of the noninteracting part of the action (5.3.3) after a_- has been integrated out is given by

$$\begin{aligned} -4\pi i \cdot S^{\rm free}_{\rm remnant} \;=\; & s_m \sum_i \int_\infty \Big[2D_-\varphi_i \nabla \times (\theta \vec{D}\varphi_i) - \theta \vec{D}\varphi_i \times \partial_- \vec{D}\varphi_i \Big] \\ & -\tfrac{1}{2p} \int_\infty (2pm\theta + 1)\vec{a} \times \partial_- \vec{a} - 2m \int_\infty \theta \vec{a} \times \vec{E} \end{aligned} \tag{5A.1}$$

where θ is shorthand for $\theta(y)$ and s_m stands for $\mathrm{sgn}(m)$. We now have to substitute (5.3.5) for $\vec{a}$,

$$\vec{a} = \tfrac{2p}{2pm\theta+1} s_m [\theta \textstyle\sum_i \vec{D}\varphi_i + \nabla\Omega]. \tag{5A.2}$$

Let us first concentrate on all the terms that do not depend on the arbitrary gauge Ω. From the $\vec{a}\times\partial_-\vec{a}$ in (5A.1) we obtain a term quadratic in $D_\mu\varphi$, and the last term in (5A.1) produces a product of A_μ and $\vec{D}\varphi$. Together this is

$$-2p \sum_{ij} \int_\infty \tfrac{\theta^2}{2pm\theta+1} \vec{D}\varphi_i \times \partial_- \vec{D}\varphi_j - 4p|m| \int_\infty \tfrac{\theta^2}{2pm\theta+1} \textstyle\sum_i \vec{D}\varphi_i \times \vec{E}. \tag{5A.3}$$

The expression involving step functions can be simplified to $\theta\nu/m$, since $\partial_y[\theta^2/(2pm\theta{+}1)] = \partial_y[\theta\nu/m]$. Adding (5A.3) to the first line of (5A.1), we obtain

$$-4\pi i \cdot S^{\rm free}_{\Omega=0} = \nu \int A \wedge \mathrm{d}A - \sum_{ij} [\delta_{ij} s_m - 2p\tfrac{\nu}{m}] \oint (D_x\varphi_i D_-\varphi_j - \varphi_i E_x). \tag{5A.4}$$

Now we concentrate on the Ω. From $\vec{a}\times\partial_-\vec{a}$ we get a quadratic term in Ω,

$$-2p \int_\infty \tfrac{1}{2pm\theta+1} \nabla\Omega \times \partial_- \nabla\Omega. \tag{5A.5}$$

Integrating by parts and using the relation $\partial_y[1/(2pm\theta + 1)] = -2p\nu\delta(y)$ this can be rewritten as

$$-(2p)^2\nu \oint \partial_x\Omega\, \partial_-\Omega. \tag{5A.6}$$

Both terms in the second line of (5A.1) give rise to linear terms in Ω; written together this is

$$-4p\; s_m \int_\infty \tfrac{\theta}{2pm\theta+1} \nabla\Omega \times \nabla \textstyle\sum_i D_-\varphi_i = 4p\tfrac{\nu}{|m|} \oint \partial_x\Omega \textstyle\sum_i D_-\varphi_i, \tag{5A.7}$$

where we have integrated by parts and used $\partial_y[\theta/(2pm\theta+1)]=\delta(y)\nu/m$. The free action after the C-S mapping is given by (5A.4)+(5A.6)+(5A.7). It is easily seen that the terms involving Ω can be completely absorbed into a redefinition of the φ fields according to

$$\tilde{\varphi}_i = \varphi_i - 2p\ s_m\Omega. \tag{5A.8}$$

The mapping of the interaction term is simple. We write the Coulomb contribution in the form

$$-\tfrac{1}{8\pi^2}\int_\infty U_0(\vec{x}-\vec{x}')\nabla\times[\theta{\textstyle\sum_i}\vec{\mathcal{D}}\varphi_i](\vec{x})\ \nabla'\times[\theta{\textstyle\sum_j}\vec{\mathcal{D}}\varphi_j](\vec{x}') \tag{5A.9}$$

and note that in this expression $\theta\sum_i[\vec{D}\varphi_i - \vec{a}]$, after substituting (5A.2), is equivalent to $\theta\frac{\nu}{m}\sum_i[\vec{D}\varphi_i - 2p\ s_m\Omega] = \theta\frac{\nu}{m}\sum_i \vec{D}\tilde{\varphi}_i$. With this last ingredient, the mapping is complete and yields the result (5.3.9).

5B C-S mapping in the presence of a tunnelling term

Integration over the C-S gauge fields in the presence of a tunnelling term proceeds along the same lines as in Appendix A. The remnant of the free action is now given by

$$\begin{aligned}-4\pi i\cdot S^{\text{free}}_{\text{remnant}} &= s_m\sum_i\int_\infty\Big[2D_-\varphi_i\nabla\times(\theta\vec{D}\varphi_i) - \theta\vec{D}\varphi_i\times\partial_-\vec{D}\varphi_i\Big] \\ &-\tfrac{1}{2p}\int_\infty(2pm\theta+1)\vec{a}\times\partial_-\vec{a} - 2m\int_\infty\theta\vec{a}\times\vec{E} - 4\pi\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\ (\partial_0\varphi_i - iv_{\mathrm{d}}a_x)(\tau,\vec{x}_0)\end{aligned} \tag{5B.1}$$

and we have to substitute

$$\vec{a}(\vec{x}) = \tfrac{2p\ s_m}{2pm\theta(y)+1}\left[\theta(y)\sum_i\vec{D}\varphi_i + \nabla\Omega + s_m L\nabla\text{arctg}\tfrac{x-x_0}{y-y_0}\right] \tag{5B.2}$$

with L_n as defined in (5.4.5). Eq. (5B.2) would suggest that one can obtain an action of precisely the same form as in Appendix A from (5B.1) (minus the tunnelling term at $\vec{x}_0$) by redefining Ω according to $\Omega' = \Omega + s_m L\text{arctg}\frac{x-x_0}{y-y_0}$. However, whereas Ω satisfies $\nabla\times\nabla\Omega=0$, we have $\nabla\times\nabla\text{arctg}\frac{x-x_0}{y-y_0} = -2\pi\delta(\vec{x}-\vec{x}_0)$. This means that apart from the free part of (5.3.9) (containing Ω' instead of Ω) we get extra contributions at $\vec{x}_0$. The $\vec{a}\times\partial_-\vec{a}$ part of (5B.1) gives rise to the following

$$-2p\ s_m\int_\infty\tfrac{1}{2pm\theta+1}L\nabla\text{arctg}\tfrac{x-x_0}{y-y_0}\times\partial_-\left[s_m L\nabla\text{arctg}\tfrac{x-x_0}{y-y_0} + 2\theta\sum_i\nabla\varphi_i + 2\nabla\Omega\right]. \tag{5B.3}$$

Integrating by parts and discarding all derivatives of step functions (which yield $\oint$ terms that we have already accounted for), we move all the derivatives in $[\cdots]$ to the left so that they can operate on the arctg. This gives

$$4\pi p\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\left[2\tfrac{\nu}{m}\theta\sum_i\partial_-\varphi_i + 2s_m(1-2p\nu\theta)\partial_-\{L\text{arctg}\tfrac{x-x_0}{y-y_0}\} + 2(1-2p\nu\theta)\partial_-\Omega\right](\vec{x}_0). \tag{5B.4}$$

Notice the factor of 2 that has popped up in front of the arctg term. Its origin lies in the fact that ∂_x and ∂_y do not commute when acting on arctg. We have

$$\partial_x \partial_y = \nabla\times\nabla + \partial_y \partial_x \tag{5B.5}$$

and the $\nabla\times\nabla$ produces an extra delta function. From the $\vec{a}\times\vec{E}$ part of (5B.1) we get

$$2m\int_\infty \theta A_- \nabla\times\vec{a} \longrightarrow -8\pi\, p\nu\theta \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\ A_-(\vec{x}_0). \tag{5B.6}$$

The $\int \mathrm{d}\tau$ part of (5B.1) gives

$$-4\pi\int_{\tau_1}^{\tau_2}\mathrm{d}\tau \left\{ \partial_0\varphi_i - iv_\mathrm{d} 2p\, s_m \left[\tfrac{\nu}{m}\theta \sum_j D_x\varphi_j + (1-2p\nu\theta)\partial_x(\Omega + s_m L\mathrm{arctg}\tfrac{x-x_0}{y-y_0}) \right] \right\} \tag{5B.7}$$

Adding (5B.4), (5B.6) and (5B.7) we find

$$S^{\mathrm{free}}_{\mathrm{at}\ \vec{x}_0} = -i\int_{\tau_1}^{\tau_2}\mathrm{d}\tau \left\{ \partial_0\tilde{\varphi}_i(\vec{x}_0) - 2p\tfrac{\nu}{m}\theta(y_0)\sum_j D_0\tilde{\varphi}_j(\vec{x}_0) \right\} \tag{5B.8}$$

where we have defined

$$\tilde{\varphi}_i(\vec{x}) = \varphi_i(\vec{x}) - 2ps_m\Omega(\vec{x}) - 2pL\mathrm{arctg}\tfrac{x-x_0}{y-y_0}. \tag{5B.9}$$

Now we deal with the interaction term, in the same way as in Appendix A. Again we can write

$$\theta\sum_i(\vec{D}\varphi_i - \vec{a}) \longrightarrow \theta\tfrac{\nu}{m}\sum_i \vec{D}\tilde{\varphi}_i. \tag{5B.10}$$

Only one subtlety is left: because of the presence of the arctg in $\tilde{\varphi}$, the relation $\nabla\times\vec{D}\tilde{\varphi}_i = -B$ gets altered. We have

$$\nabla\times\vec{D}\tilde{\varphi}_i = -B' \quad ; \quad B'(\vec{x}) = \nabla\times\vec{A}(\vec{x}) - 2p\delta(\vec{x}-\vec{x}_0) \tag{5B.11}$$

which reflects the fact that there is extra flux at $\vec{x}_0$.

5C Tunnelling d.o.s. resulting from neutral modes

The result S_out (5.4.10) is obtained from the expectation value $\langle \exp -i\tilde{\varphi}_k|_{\tau_1}^{\tau_2}\rangle$ as follows. The field $\tilde{\varphi}_k$ decomposes into charged and neutral modes according to (5.3.14). From the action (5.3.15) we find the correlations for the diagonal modes,

$$\langle\Gamma(\tau_2)\Gamma(\tau_1)\rangle = \tfrac{1}{\nu}\ln(\tau_2-\tau_1) \qquad ; \qquad \langle\gamma_k(\tau_2)\gamma_{k'}(\tau_1)\rangle = \delta_{kk'}\tfrac{k+1}{k}\ln(\tau_2-\tau_1) \tag{5C.1}$$

and $\langle\gamma_k\Gamma\rangle = 0$. Evaluation of $\langle \exp -i\tilde{\varphi}_k|_{\tau_1}^{\tau_2}\rangle$ for arbitrary k then yields the tunnelling exponent

$$\begin{aligned} S_\mathrm{out} &= \frac{1}{\nu} + \frac{k-1}{k} + \sum_{a=k}^{|m|-1}\frac{1}{a(a+1)} = \frac{1}{\nu} + 1 - \frac{1}{k} + \sum_{a=k}^{|m|-1}\left(\frac{1}{a} - \frac{1}{a+1}\right) \\ &= 2p + 1 + \frac{1}{m} - \frac{1}{|m|}. \end{aligned} \tag{5C.2}$$

Summary

A conclusion is the place where you got tired thinking.
– M. H. Fischer

The contents of this booklet can be summarised as follows. We have found a new symmetry in the replica formalism for disorder and interactions. This $\mathcal{F}$-invariance has enabled us to set up a unifying theory for the quantum Hall effect. Combined with an unusual new frequency truncation procedure it dictates the form of effective actions and allows for the inclusion of $U(1)$ gauge fields in the theory. The proposed theory connects edge and bulk physics and combines the Finkelstein action with the topological concepts of an instanton vacuum and statistical gauge fields.

The existence of $\mathcal{F}$-invariance has made it possible to do a perturbative RG analysis of the Finkelstein theory to two loop order. We have shown that the infrared behaviour of the theory can only be extracted from a limited class of ($\mathcal{F}$-invariant) correlation functions. We have identified the 'order parameter' of the metal-insulator transition in $2{+}2\varepsilon$ dimensions and the 'Coulomb gap'. $\mathcal{F}$-invariance also has fundamental consequences for the insulating phase, where ordinary perturbation theory is no longer valid. To this end, we exploited the analogy of the Finkelstein theory with the classical Heisenberg ferromagnet. We have shown that the transport problem in this case must be dominated by the S_{Y} part of the action which is usually discarded on the basis of naive scaling dimensions.

Fluctuations about integer quantisation of the topological charge $q[Q]$ give rise to massless excitations on the edge. These are equivalent to a model of disordered chiral fermions subject to inter-channel scattering. This equivalence leads to a better understanding of the Q-field theory in 1+1 dimensions, since the disorder can be gauged away in the chiral fermion theory, resulting in an exactly solvable model. In particular, several operators bilinear in Q turn out to be redundant. We have also shown that a topological principle for the quantisation of σ_{xy} can still be formulated, because spherical boundary conditions are dynamically generated.

We have derived a complete theory of the edge, including interactions and gauge fields, in the limit of vanishing bulk density of states. Chiral edge bosons arise through a beautiful interplay between the topological term and $\mathcal{F}$- and gauge invariance. The chiral anomaly at the edge provides a natural description of Laughlin's gauge argument, connecting σ_{xy} as a bulk quantity to σ_{xy} as an edge quantity.

Our approach to edge physics can be used to address several problems of smooth disorder and interaction effects. We model electrons in a sample with smooth disorder as a collection of spatially separated but interacting 'edge modes' that live on the equipotential contours. We have pointed out that fundamental differences exist between tunnelling at the edge and electron transport. Transport experiments inject electrons directly into edge states; these electrons do not get enough time to equilibrate with the rest of the sample and are therefore effectively decoupled from the bulk. A tunnelling measurement, however, probes eigenstates of the whole system, which involve not only edge electrons but also localised bulk orbitals. Since tunnelling processes do not probe the incompressibility of the electron gas, they are generally treated incorrectly by the theory of isolated edges. By taking into account the effect of Coulomb interactions between the edge and the localised bulk states, we have derived an effective edge theory in which the neutral modes get suppressed, while the charged mode acquires a non-quantised Luttinger liquid parameter. A tunnelling exponent $1/\nu$ is obtained, in accordance with experiments. The Hall conductance in this theory is still quantised, as can be understood by applying the Laughlin gauge argument to all edge channels separately.

For the plateau transitions we have constructed a percolation model of interacting edges. Inelastic scattering at the 'nearly saddlepoints' sets the temperature scale at which the transport coefficients cross over from mean field behaviour to critical scaling. This crossover can involve arbitrarily low temperatures and it explains the 'lack of scaling' in the transport data taken from samples with long range disorder at finite temperatures. Our mean field expression for the conductances agrees with recent empirical fits to transport data at plateau transitions.

We have applied the Chern-Simons flux attachment procedure to the bulk theory, the edge theory and the percolating edge state description of plateau transitions. We studied the bulk theory at weak magnetic fields, where the C-S mapping results in states at $\nu \approx 1/2$. The conductances are mapped according to SL(2,$\mathbb{Z}$). For free particles or short range interactions the quasiparticle density of states becomes divergent, indicating that the C-S procedure requires long range interactions in order to be well defined.

The C-S mapping applied to the iqHe theory of chiral edge bosons leads to an action that has the well known K-matrix structure but also some new properties. It offers several new insights, in particular for the $\nu = 2/3$ state. In the absence of long range interactions, the Chern-Simons procedure is ill-defined for systems with counter-flowing edge modes. The Hamiltonian of the charged mode is unbounded from below and the theory can only be made stable by the Coulomb interaction. Our description for these systems is not plagued by non-universal behaviour of tunnelling exponents and the Hall conductance. The chiral anomaly provides an elegant way of using Laughlin's gauge argument.

Our direct way of mapping tunnelling operators has shown that the C-S procedure affects the charge and statistics of tunnelling particles. An electron operator outside the sample retains its unit charge after the mapping, but inside the charge is mapped to ν/m. The operators inside and outside are related by a T-duality transformation that inverts the compactification radius (K-matrix); thus we have found a new geometrical interpretation for this duality.

Our analysis of spatially separated edge modes has revealed an even richer duality structure in the spectrum of quasiparticles, where the reversal inside/outside is generalised to a reversal of the order of edge channels. Samples with positive m and negative m are jointly described in this picture. We have shown that in the limit of large length scales, the theory of 'clean', spatially separated edges describes identical physics as the 'dirty' K-matrix theory. In the case of counter-flowing modes this happens in a quite nontrivial way.

We have made a short analysis of transitions between fqH plateaus, mapping the Q-field theory for percolating integer edges. The results indicate that the critical aspects of the plateau transitions are generally the same for both the integral and fractional regime. It turns out that the Chern-Simons mapping is not described by SL(2,$\mathbb{Z}$).

To summarise the summary, $\mathcal{F}$-invariance has quite an impact on our understanding of disorder in the qHe and on our ability to handle calculations. It is amusing to see how much of the physics occurs in the form of edge terms, not only in real space but in the 'corners' of frequency space as well.

The end of a summary is traditionally the place where one looks forward and tries to predict the future. I think it is quite safe to say that the uses we have found for $\mathcal{F}$-invariance do not exhaust the full range of its possible applications. In the description of spin degrees of freedom of electrons in quantum Hall samples, for instance, neither disorder nor interactions can be neglected. It would be very interesting to see how disorder affects the formation of skyrmions. Spin indices will of course have to be included in the Q-matrix formalism for such an exercise. Another spin-related phenomenon is the $\nu = 5/2$ effect. Perhaps an edge theory in terms of Q-fields can be found that is equivalent to one of the proposed wave functions for the $\nu = 5/2$ state. Such a Q-field theory may also be used to study the corresponding quasiparticles, which are beleived to have nonabelian statistics.

References

[1] E. Abrahams, P.W. Anderson, D.C. Licciardello and T.V. Ramakrishnan, *Phys. Rev. Lett.* **42** (1979) 673

[2] B.L. Altshuler and A.G. Aronov, *Solid St. Comm* **46** (1980) 429

[3] B.L. Altshuler, A.G. Aronov and P.A. Lee, *Phys. Rev. Lett.* **44** (1980) 1288

[4] N. Balaban, U. Meirav and I. Bar-Joseph, cond-mat/9806214

[5] A.P. Balachandran, L. Chandar and B. Sathiapalan, *Nucl. Phys.* **B443** (1995) 465; *Int. J. Mod. Phys.* **A11** (1996) 3587

[6] M.A. Baranov and A.M.M. Pruisken, to be published

[7] D. Belitz and T.R. Kirkpatrick, *Rev. Mod. Phys.* **166** (1994) 261

[8] B. Blok and X.G. Wen, *Phys. Rev.* **B42** (1990) 8133; *Phys. Rev.* **B42** (1990) 8145

[9] E. Brézin, S. Hikami and J. Zinn-Justin, *Nucl. Phys.* **B165** (1980) 528

[10] C. Castellani and C. Di Castro, *Phys. Rev.* **B34** (1986) 5935

[11] C. Castellani, C. Di Castro, P.A. Lee and M. Ma, *Phys. Rev.* **B30** (1984) 527

[12] C. Castellani, G. Kotliar, P.A. Lee, *Phys. Rev. Lett.* **59** (1987) 323

[13] J.T. Chalker and P.D. Coddington, *J. Phys.* **C21** (1988) 2665

[14] A. Chang, L.N. Pfeiffer and K.W. West, *Phys. Rev. Lett.* **77** (1996) 2538

[15] M.H. Cohen and A.M.M. Pruisken, *AIP Conf. Ser.* **286** (1993) 205; *Phys. Rev.* **B49** (1994) 4593

[16] S. Deser, R. Jackiw and S. Templeton, *Ann. Phys. (N.Y.)* **140** (1982) 372

[17] S.F. Edwards and P.W. Anderson, *J. Phys.* **F5** (1975) 965

[18] A.L. Efros and B.I. Shklovskiĭ, *J. Phys.* **C8** (1975) L49

[19] J.P. Eisenstein and H.L. Störmer, *Science* **248** (June 1990) 1510

[20] P. Fendley, A.W.W. Ludwig and H. Saleur, *Phys. Rev.* **B52** (1995) 8934; *Phys. Rev. Lett.* **75** (1995) 2196

[21] A.L. Fetter, C.B. Hanna and R.L. Laughlin, *Phys. Rev.* **B39** (1989) 9679; **43** (1991) 309

[22] A.M. Finkelstein, *JETP Lett.* **37** (1983) 517; *Soviet Phys. JETP* **59** (1984) 212; *Physica* **B197** (1994) 636

[23] E. Fradkin, *Field Theories of Condensed Matter Systems*, Addison-Wesley (Redwood City, 1991)

[24] J. Fröhlich and T. Kerler, *Nucl. Phys.* **B354** (1991) 369

[25] J. Fröhlich and U.M. Studer, *Rev. Mod. Phys.* **65** (1993) 733

[26] J. Fröhlich and A. Zee, *Nucl. Phys.* **B364** (1991) 517

[27] M. Grayson, D.C. Tsui, L.N. Pfeiffer, K.W. West and A.M. Chang, *Phys. Rev. Lett.* **80** (1998) 1062

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

[29] F.D.M. Haldane, *Phys. Rev. Lett.* **67** (1991) 937

[30] F.D.M. Haldane, *Phys. Rev. Lett.* **74** (1994) 2090

[31] B.I. Halperin, *Phys. Rev. Lett.* **52** (1984) 1583; 2290(E)

[32] B.I. Halperin, P.A. Lee and N. Read, *Phys. Rev.* **B47** (1993) 7312

[33] B. Huckestein and B. Kramer, *Phys. Rev. Lett.* **64** (1990) 1437

[34] J. K. Jain, *Phys. Rev. Lett.* **63** (1989) 199; *Phys. Rev.* **B 40** (1989) 8079; *Adv. Phys.* **41** (1992) 105

[35] C.L. Kane and M.P.A. Fisher, *Phys. Rev. Lett.* **68** (1992) 1220; *Phys. Rev.* **B46** (1992) 15233

[36] C.L. Kane and M.P.A. Fisher, *Phys. Rev.* **B51** (1995) 13449

[37] C.L. Kane, M.P.A. Fisher and J. Polchinski, *Phys. Rev. Lett.* **72** (1994) 4129

[38] S. Kivelson, D-H. Lee and S-C. Zhang, *Phys. Rev.* **B46** (1992) 2223

[39] K. von Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45** (1980) 494

[40] J.S. Langer and T. Neal, *Phys. Rev. Lett.* **16** (1966) 984

[41] R. Laughlin, *Phys. Rev. Lett.* **50** (1983) 1395

[42] R.B. Laughlin, M.L. Cohen, J.M. Kosterlitz, H. Levine, S.B. Libby and A.M.M. Pruisken, *Phys. Rev.* **B32** (1985) 1311

[43] D.H. Lee and X.G. Wen, cond-mat/9809160

[44] A. Lerda, *Anyons*, (Springer-Verlag, Berlin 1992)

[45] H. Levine, S. Libby and A.M.M. Pruisken, *Phys. Rev. Lett.* **51** (1983) 20

[46] A. Lopez and E. Fradkin, *Phys. Rev.* **B44**, 5246 (1991); *Phys. Rev. Lett.* **69** (1992) 2126

[47] A. Lopez and E. Fradkin, *Phys. Rev.* **B51** (1995) 4347

[48] A. Lopez and E. Fradkin, cond-mat/9810168

[49] J.M. Luttinger, *J. Math. Phys.* **4** (1963) 1154

[50] C.A. Lütken, *Nucl. Phys.* **B396** (1993) 670

[51] C.A. Lütken and G.C. Ross, *Phys. Rev.* **B48** (1993) 2500

[52] N. Maeda, *Phys.Lett.* **B376** (1996) 142

[53] D.C. Mattis, E.H. Lieb, *J. Math. Phys.* **6** (1965) 304

[54] Y. Oreg and A.M. Finkelstein, *Phys. Rev. Lett.* **74** (1995) 3668

[55] R. de-Picciotto et al., *Nature* vol **389** (september 1997) 162

[56] D. Pines and P. Nozières, *The Theory of Quantum Liquids: Volume 1*, (W.A. Benjamin Inc., New York 1966)

[57] For a review, see *The Quantum Hall Effect*, edited by R.E. Prange and S.M. Girvin (Springer-Verlag, Berlin, 1987)

[58] A.M.M. Pruisken, *Nucl. Phys.* **B235** (1984) [FS11] 277

[59] A.M.M. Pruisken, *Nucl. Phys.* **B285** 719 (1987); **B290** (1987) 61

[60] A.M.M. Pruisken, *Phys. Rev.* **B31** (1985) 416

[61] A.M.M. Pruisken, *Phys. Rev.* **B32** (1985) 2636

[62] A.M.M. Pruisken, *Phys. Rev. Lett.* **61** (1988) 1297

[63] See A.M.M. Pruisken in [57]

[64] A.M.M. Pruisken and M.A. Baranov, *Europhys. Lett.* **31**(9) (1995) 543

[65] A.M.M. Pruisken and K. Schoutens, *Phil. Mag.* **B76** (1997) 807

[66] A.M.M. Pruisken, Z. Wang, *Nucl. Phys.* **B322** (1989) 721

[67] See, e.g., A.M.M. Pruisken and H.P. Wei, *AIP Conf. Proc.* **286** (1994) 159

[68] N. Read, *Phys. Rev. Lett.* **65** (1990) 1502

[69] L. Saminadayar et al., *Phys. Rev. Lett.* **79** (1997) 2526

[70] J. Schonfeld, *Nucl. Phys.* **B185** (1981) 157

[71] D. Shahar et al., *Sol. State Comm.* **107** (1998) 19

[72] B. Tieke et al., *Phys. Rev. Lett.* **76** (1996) 3630; *Surf. Sci.* **361-362** (1996) 46; P.A. Crump et al., *Surf. Sci.* **361-362** (1996) 50; S.A.J. Wiegers et al., *Czech. J. Phys.* **46** (1996) 2461

[73] D.C. Tsui, H.L. Störmer and A.C. Gossard, *Phys. Rev. Lett.* **48** (1982) 1559

[74] F. Wegner, *Z. Physik* **B36** 209 (1979); **B38** (1979) 207

[75] H.P. Wei, D.C. Tsui and A.M.M. Pruisken, *Phys. Rev.* **B33** (1986) 1488

[76] H.P. Wei, D.C. Tsui, M.A. Palaanen and A.M.M. Pruisken, *Phys. Rev. Lett.* **61** (1988) 1294

[77] X.G. Wen, *Phys. Rev.* **B41** (1990) 12838; *Int. J. Mod. Phys.* **B6** (1991) 1711; cond-mat/9506066.

[78] X.G. Wen, *Phys. Rev.* **B44** (1991) 5708

[79] X.G. Wen and A. Zee, *Phys. Rev. Lett.* **69** (1992) 1811

[80] X.G. Wen and A. Zee, *Phys. Rev.* **B46** (1992) 2290

[81] F. Wilczek, *Phys. Rev. Lett.* **48** (1982) 1144; **49** (1982) 957

[82] F. Wilczek editor, *Fractional statistics and anyon superconductivity*, (World Scientific, Singapore, 1990)

[83] E. Witten, *Comm. Math. Phys.* **121** (1989) 351

[84] S.-C. Zhang, *Int. J. Mod. Phys.* **B6** (1992) 25

[85] S.-C. Zhang. H. Hansson and S. Kivelson, *Phys. Rev. Lett.* **62** (1989) 82

Samenvatting

Verantwoording aan de Nederlandse belastingbetaler.

De titel van dit proefschrift luidt in het Nederlands "$\mathcal{F}$-invariantie en de toepassing ervan op het quantum Hall effect". Wat is nu dit quantum Hall effect, waarom zijn zoveel mensen erin geïnteresseerd en wat is de strekking van al die formules in dit boekje? Ik zal proberen enige uitleg te geven in voor leken begrijpelijke taal.

Het is mogelijk om electronen op te sluiten in twee dimensies door speciaal gefabriceerde isolator/halfgeleider samples af te koelen tot onder 1 Kelvin. Bij een dergelijk lage temperatuur worden er electronen ingevangen in het grensvlak tussen de isolator en de halfgeleider. Men spreekt van een 'tweedimensionaal electrongas'. In 1980 deed Von Klitzing een verrassende ontdekking toen hij de geleidingseigenschappen van zulke tweedimensionale electrongassen in een magnetisch veld bestudeerde. De Hall weerstand (de verhouding tussen het voltage loodrecht op de stroom en de stroom zelf) bleek plateaux te vertonen als functie van de sterkte van het magneetveld loodrecht op het electrongas (zie figuur 1.1 op blz. 9). Bij het optreden van een plateau werd de gewone weerstand nul. De waarde van de Hall geleiding op een plateau bleek gequantiseerd te zijn in gehele veelvouden van e^2/h, waar e staat voor de lading van het electron en h voor de constante van Planck. Het optreden van zulke plateaux wordt het integer quantum Hall effect (iqHe) genoemd. In 1982 deden Tsui, Störmer en Gossard de verbluffende ontdekking dat er in zeer schone samples ook plateaux kunnen optreden bij *fractionele* veelvouden van e^2/h, waarvan de meest prominente 1/3, 1/5, 2/3 en 2/5. Dit wordt het fractionele quantum Hall effect (fqHe) genoemd.

Voor beide ontdekkingen zijn Nobelprijzen uitgereikt; voor het iqHe omdat het op opvallende wijze aantoonde dat de toenmalige theorieën over geleiding onvolledig waren; voor het fqHe omdat het liet zien dat zelfs de theorieën die het iqHe konden verklaren niet afdoende waren en omdat het bovendien leidt tot merkwaardige verschijnselen zoals deeltjes met fractionele lading en statistiek. Het woord statistiek slaat hier op het effect van een omwisseling van twee identieke deeltjes op de golffunctie. Bekend zijn twee soorten deeltjes, de bosonen, waarbij de golffunctie onveranderd blijft bij een omwisseling, en de fermionen, waarbij een omwisseling aanleiding geeft tot een minteken voor de golffunctie. 'Fractionele' deeltjes zijn een generalisatie hiervan, die in twee dimensies kan bestaan. In plaats van een factor ± 1 krijgt de golffunctie een fasefactor $e^{i\theta}$.

Binnen de natuurkunde van de vaste stof neemt het qHe een bijzondere plaats in omdat het een *topologisch* fenomeen is. Topologisch wil zeggen dat het slechts afhangt van een globale geometrische eigenschap. Dit is terug te voeren op het feit dat de meetopstelling equivalent is aan een torus: de meting van de Hall spanning vormt de ene lus, de stroomkring de andere. Een continue vervorming van het sample of een herverdeling van de verontreinigingen erin kunnen deze geometrie niet veranderen. Het bestaan van 'fractionele' deeltjes is in zekere zin een gevolg hiervan.

Het fqHe is een van de weinige verschijnselen waar het electron als gevolg van een collectief veeldeeltjeseffect uiteen lijkt te vallen, en het enige fysische systeem waar de fractionele lading min of meer direct is waargenomen. Het qHe is tevens bijzonder omdat het lijkt alsof het electrongas en zijn rand een gescheiden leven leiden. Deze unieke status heeft ertoe geleid dat het fqHe door theoretici wordt gebruikt als een soort testvlakte voor het uitproberen van nieuwe ideeën over fractionele deeltjes en 1+1 dimensionale fysica. Enige voorzichtigheid hierbij is geboden (maar wordt niet altijd betracht). Het is namelijk niet à priori duidelijk hoe de rand van het electrongas en de fractionele deeltjes worden beïnvloed door de verontreinigingen in het sample en de 'wanorde-potentiaal' waar de electronen zich daardoor in bevinden.

De reden voor deze onduidelijkheid is dat het heel moeilijk is om een enkele beschrijving voor het qHe te vinden die onder alle omstandigheden werkt, d.w.z. zowel op de Hall plateaux als ertussen, zowel middenin het sample (de 'bulk') als op de rand en zowel voor het iqHe als voor het fqHe. Voor een volledige beschrijving van het qHe is het noodzakelijk om een goed begrip te hebben van geladen deeltjes die door een grillig potentiaallandschap bewegen. Dit is geen eenvoudige zaak, want de onderlinge afstoting tussen de deeltjes zorgt ervoor dat de vaak gebruikte opsplitsing in ééndeeltjes-subbeschrijvingen niet werkt, terwijl de verontreinigingen de translatiesymmetrie breken die goed van pas had kunnen komen om het probleem te kraken. De aanwezigheid van een sterk magneetveld maakt het allemaal nog ingewikkelder.

Het belang van symmetrieën in de natuurkunde (ook wel invarianties genoemd) is niet te overschatten. Fundamentele natuurwetten zoals behoud van energie, impuls en lading zijn direct gerelateerd aan symmetrieën. Het feit dat de eigenschappen van de ruimte om ons heen (de lichtsnelheid, constante van Planck, π etc.) vandaag dezelfde zijn als gisteren leidt tot de wet van behoud van energie; het feit dat deze eigenschappen bij de bakker om de hoek precies eender zijn als hier heeft het behouden zijn van impuls tot gevolg. Behoud van lading is iets subtieler. Hier gaat het om een zgn. *ijksymmetrie*, een speciaal geval van een symmetrie in een andere ruimte dan de 3+1 dimensionale waar we ons door bewegen. De eenvoudigste illustratie van deze ijksymmetrie gaat als volgt. Een electrische spanning rekenen we uit als het verschil tussen twee potentialen. Als we nu overal in de ruimte de electrische potentiaal met een constante hoeveelheid verschuiven, dan zal de spanning onveranderd blijven. Meer in het algemeen wordt een ijksymmetrie gedefinieerd als een transformatie (zoals het verschuiven van de potentiaal) die geen enkele observabele grootheid verandert.

Behalve dat ze iets vertellen over behoudswetten, helpen symmetrieën ons op een tweede manier: ze brengen vaak zoveel structuur aan dat berekeningen ongelooflijk vereenvoudigd kunnen worden. Als men bijvoorbeeld een probleem gepresenteerd krijgt waar rotatiesymmetrie in zit, dan weet men vantevoren dat de oplossing gezocht moet worden in termen van rotatie-invariante formules. Sommige problemen in de natuurkunde zijn zo moeilijk dat ze zonder het gebruik van symmetrieën praktisch onoplosbaar zijn.

De $\mathcal{F}$-invariantie die centraal staat in dit proefschrift is een ijksymmetrie gerelateerd aan het behoud van lading. Het zou te ver voeren om de exacte definitie ervan hier te geven. Laat het volstaan om te zeggen dat deze invariantie aanwezig is in de zgn. Finkelstein-theorie die geladen deeltjes in een wanorde-potentiaal beschrijft. Deze theorie is in 1983 opgesteld door Finkelstein, maar het heeft tot 1995 geduurd voordat de $\mathcal{F}$-invariantie werd opgemerkt door Pruisken en Baranov. Het merkwaardige is dat de langedrachts Coulomb-interacties noodzakelijk zijn voor het bestaan van deze ijkinvariantie. In plaats van een complicerende factor te zijn maken de interacties de theorie hanteerbaar door er een symmetriestructuur in aan te brengen.

Mijn promotieonderzoek, dat in 1995 begon, heeft voor een deel bestaan uit het maximaal uitbuiten van de pas ontdekte structuur. Ten eerste werd duidelijk dat deze gebruikt kan worden om verschillende aspecten van het qHe onder één noemer te brengen. 'Unificatie' is een geliefd woord onder fysici. In tegenstelling tot wat vaak beweerd wordt is luiheid, mits strategisch toegepast, een positieve eigenschap in wetenschappers. Hoe minder denkwijzen, ingrediënten, formules en postulaten een theoretische beschrijving immers nodig heeft hoe beter. Met behulp van $\mathcal{F}$-invariantie lukte het om een gecombineerde theorie op te zetten van de rand en de bulk. Op de rand werken topologie en ijksymmetrie op wonderbaarlijk schone wijze samen. Hiernaast bleek het ook mogelijk om binnen het kader van de Finkelstein-theorie het fqHe en het iqHe op gelijke voet te behandelen. Dit is zeer bevredigend, want experimenteel zien de twee effecten er precies hetzelfde uit. Om tot een gelijke beschrijving te komen wordt standaard gebruik gemaakt van een mathematische truc die bekend staat onder de naam 'Chern-Simons flux afbeelding'. Hierbij wordt het fqHe beschouwd als het optreden van het iqHe, maar dan in termen van andere deeltjes, *samengestelde fermionen*, die bestaan uit een electron met daaraan vastgeplakt twee magnetische fluxquanta. Voor het toepassen van deze truc is het noodzakelijk dat de theoretische beschrijving van het iqHe voorziet in een koppeling van de electronen aan het electromagnetische veld. Het is precies de $\mathcal{F}$-invariante structuur van de Finkelstein-theorie die de koppeling mogelijk maakt. Als opvallend resultaat kwam naar voren dat de Chern-Simons afbeelding slechts mogelijk is in aanwezigheid van langedrachts interacties in het sample, m.a.w. wanneer er $\mathcal{F}$-invariantie is.

Behalve voor bovenstaande unificatie hebben we de $\mathcal{F}$-invariantie gebruikt om een zgn. 'twee lus renormalisatie' berekening te doen. Renormalisatie is een veelgebruikte methode in de hoge-energiefysica en vaste-stoffysica, waarmee bepaald kan worden hoe de eigenschappen van een systeem veranderen als het systeem groter wordt gemaakt. Dit is een zeer nuttige procedure, omdat eenvoudige berekeningen in de vaste stof vaak alleen kunnen worden gedaan op lengteschalen veel kleiner dan de afmetingen van een sample. Renormalisatie vormt de schakel tussen die berekeningen en daadwerkelijke experimenten. De handelingen die moeten worden verricht in de renormalisatieprocedure worden grafisch weergegeven met behulp van zgn. Feynmandiagrammen. In deze diagrammen komen lussen voor (zie figuur 3.3 op blz. 52) die tot aanzienlijke mathematische complicaties kunnen leiden. De nauwkeurigheid van de renormalisatie wordt echter groter naarmate er meer lussen in de procedure worden betrokken. Middels $\mathcal{F}$-invariantie is het gelukt om in de Finkelstein-theorie alle diagrammen met twee lussen uit te rekenen, iets wat eerder niet mogelijk was.

Een ander deel van mijn onderzoek bestond uit het onder de loep nemen van de randen van quantum Hall systemen, o.a. wat daar gebeurt wanneer de Chern-Simons afbeelding wordt toegepast. Enkele van de door ons gevonden resultaten zijn van direct belang voor het begrijpen van recente experimenten. Het mooiste voorbeeld is een *tunneling*-experiment waarbij een stroompje wordt gemeten dat loopt tussen een 'naald' van gewoon metaal en de rand van een fqH sample. Volgens de klassieke natuurkunde kan er geen stroom lopen door het isolerende materiaal ertussenin, maar vanwege het quantummechanische karakter van het systeem kunnen er toch wat electronen doorheen lekken. De sterkte van deze tunnelingstroom is niet zoals bij een gebruikelijke stroom rechtevenredig met de aangelegde spanning, omdat de rand van een quantum Hall sample zich niet als een gewoon metaal gedraagt. Het verband tussen stroom en spanning wordt gegeven door $I \sim V^{\alpha}$. Het getal α heet tunneling exponent. Experimenten hebben aangetoond dat de tunneling exponent, tegen alle verwachtingen in, lineair van het magneetveld afhangt en geen plateaux vertoont zoals de Hall geleiding. Uit onze berekeningen is de volgende verklaring naar voren gekomen. De rand en de bulk zijn *niet* ontkoppeld. Coulomb interacties tussen electronen op de rand en in de bulk zorgen ervoor dat een tunneling experiment niet alleen de rand aftast, maar ook een deel van de bulk. Het aantal electronen in de bulk dat 'gevoeld' wordt hangt af van het magneetveld, hetgeen precies leidt tot het waargenomen lineaire gedrag van α.

Bij het toepassen van de Chern-Simons flux truc gebeuren er interessante dingen aan de rand. Doorgaans wordt het fqHe aan de rand bestudeerd door eerst in een bulktheorie de Chern-Simons afbeelding toe te passen en vervolgens een randtheorie af te leiden uit de bulktheorie. Hierbij heeft men echter de vrijheid om een groot aantal constanten in te voeren (met de betekenis van snelheden van deeltjes op de rand) waar men naar believen waarden aan kan toekennen. Door deze vrijheid worden, zo hebben we aangetoond, belangrijke subtiliteiten over het hoofd gezien. We zijn uitgegaan van een beschrijving van het iqHe op de rand en hebben de Chern-Simons truc direct hierop toegepast om zonder omwegen een randtheorie voor het fqHe te verkrijgen. Op deze wijze bestaat er geen enkele keuzevrijheid. Dit werpt met name nieuw licht op het fqHe bij de fractie 2/3, waar deeltjes op de rand in tegengestelde richtingen bewegen. Wat bijvoorbeeld nooit eerder iemand was opgevallen is een stabiliteitsprobleem dat zich uit in de vorm van ongerijmdheden zoals een negatieve energie. Dit probleem kan slechts worden opgeheven door de Coulomb-interacties, hetgeen wederom aantoont dat de Chern-Simons afbeelding alleen toegepast kan worden op een $\mathcal{F}$-invariante theorie.

Ondanks het feit dat ik in deze samenvatting veel details heb weggelaten en enkele onderzoeksrichtingen niet heb vermeld, hoop ik een duidelijk beeld te hebben geschetst van de bijdrage die dit onderzoek heeft geleverd aan de vaste-stoffysica. Vele vragen zijn uiteraard nog onbeantwoord. Noch de qHe bulk noch de rand zijn nu volledig begrepen. Er valt ook nog veel te verbeteren aan ons begrip van spin-effecten en van de overgangen tussen de Hall plateaux, belangrijke aspecten waar ik het in deze samenvatting niet eens over gehad heb. Het qHe blijft voor theoretici een lastig fenomeen, maar $\mathcal{F}$-invariantie zal ook in de toekomst zeker een steentje bijdragen aan de vereenvoudiging ervan.

Dankwoord

Welkom tot dit proefschrift, beste lezer! Dit is het enige hoofdstuk dat door iedereen zorgvuldig wordt gelezen; men wil immers weten of Boris wel naar behoren dankbaar is en of men belangrijk genoeg is om bij naam genoemd te worden. Ik zal mijn best doen.

Op de eerste plaats ben ik mijn promotor Aad Pruisken erg dankbaar voor de mogelijkheid om bij hem te promoveren. Zonder zijn Fingerspitzengefühl en zijn jarenlange ervaring met quantum Hall fysica zou het verschijnen van dit boekje ondenkbaar zijn. De correctheid van zijn motto “If the physics is right, the formalism will take care of itself” bleek telkens opnieuw. Ik heb het geluk gehad om het onderzoek binnen te stappen net op het moment dat het in een stroomversnelling raakte. Daardoor kreeg ik de kans om, zonder concessies wat wetenschappelijk niveau betreft, mijn promotie af te ronden in een tijdsbestek korter dan de vier jaar die mij als bursaal gegund waren. Dit ging natuurlijk niet moeiteloos. Ik zal onze zenuwslopende decembereindsprints nooit vergeten.

I also thank Misha Baranov, whose expertise was indispensable for our project. Working together with someone of such computational prowess was quite a humbling experience. Спасибо, Миша.

Ik heb erg genoten van de quasi-periodieke werkbesprekingen van ons werkgroepje ‘Theorie v.d. Gecondenseerde Materie’. Ronald van Elburg, Kareljan Schoutens, Eddy Ardonne en Sathya Guruswamy, bedankt voor alle nuttige dingen die ik zonder de TGM meetings nooit zou hebben opgestoken.

Dan zijn daar natuurlijk alle medepromovendi aan het Instituut. Ik zal altijd met genoegen terugdenken aan de aangename sfeer en discussies in de koffiekamer en tijdens de lunch. Het is moeilijk om een selecter gezelschap jongelui te vinden! Bij naam noem ik Alain Verberkmoes, Mischa Sallé en Mark Kettenis, die nooit te beroerd zijn om een niet-goeroe te helpen met TeX-geKnuthsel en andere computerperikelen, en Jonathan Ross, die me met vele plaatjes heeft geholpen en me altijd graag inmaakt met pool als hij tijd heeft. Verder zijn zeker het vermelden waard mijn kamergenoten Roel van der Plank en later Ad Maas, die het overleefd hebben om jarenlang met mij een kamertje op het Instituut te delen.

Iedereen die mijn webpages wel eens heeft bezocht weet dat ik niet honderd procent van de tijd met natuurkunde bezig ben. “All work and no play makes Jack a dull boy”, immers. Om mijn geestelijke gezondheid te waarborgen doe ik ook nog wel eens iets anders, en daarbij heb ik gelukkig een hoop gezelschap. Ten eerste Anjali Pandit, Wout Broere en Matijs van Zuijlen, mijn oude schoolvrienden en jarenlange kameraden in twee ‘poelen des verderfs’, het zwembad en het poolcafé.

Ten tweede (met enige overlap) mijn fantasy role playing metgezellen van de (ge)wel(d)dadige P&P en AD&D sessies. Jongens, bedankt voor de vele onvergetelijke chopjes.

Dan zijn er de heren van het Geofictief Interactieproject Metaphoor. Met tevredenheid kan ik vaststellen dat we sinds de oprichting van het project desinformatie en bureaucratische ongerijmdheden tot een zeldzaam absurditeitsniveau hebben verheven.

Ook bedank ik de leden van de Babbelbox. Het geeft zo'n warm gevoel als je je e-mail een paar dagen niet hebt gelezen en plotseling merkt dat je mailbox is bezweken onder een lawine van berichten waaruit je de zorgvuldig verborgen zinvolle mededelingen moet redden.

Last but not least ben ik mijn ouders zeer dankbaar voor de nimmer aflatende zorg al die jaren.