The Modular Symmetric Renormalization Group Flow of the Quantum Hall Effect

by

Henrik Sverre Limseth

Thesis

for the degree of

Master of Science

Faculty of Mathematics and Natural Sciences
University of Oslo

October 2017

Abstract

A comprehensive introduction to the theory of modular symmetry in the renormalization group flow of the quantum Hall effect is given. The theory is tested with all available experimental scaling data, across a wide range of different materials. The data generally fits the scaling flow predicted from modular symmetry. Future experiments to further test the theory are suggested.

Acknowledgements

First and foremost, I would like to thank my supervisor Carsten Andrew Lütken for all his guidance and patience[1], and for lending me insights from his seemingly bottomless pool of knowledge these past two years. I would also like to thank Kristian Stølevik Olsen for invaluable discussions and insights, as they relate to the topics of this thesis, physics, and, somewhat reluctantly from my side, mathematics in general.

I would like to thank all the people who have my time here at Blindern *awesome*. Special thanks goes out to Torgeir, Vetle, Jon Vegard, Anne-Marthe, Sigve, Jørgen and especially Wilhelm and Anders, who have done all my chores around the apartment for the past couple of months. Thanks to the great student environment at the theory group and Lillefy. I would name you all, but I fear this thesis is already way too long.

Thank you mamma, pappa, Olaves and Ari for your continued support and belief in me.

Thank you onkel Kjetil for the original push into the natural sciences.

Thanks to Stig, Hævvan, Halvard and Lars for friendships that do not fade.

Thanks to Lars Jørgen Røed for taking on all the responsibilities of our startup business while I finished this thesis. I look forward to continuing our work, as well as our perpetual endeavour to find new and obscure Liam Neeson movies.

Lastly, this is for whatever reason not canonical, but I would like to extend my deepest gratitude to Lånekassen, and by extension every single Norwegian tax payer for financing me for the past six years.

[1] Seriously.

Contents

Chapter 1

Introduction

In classical physics, a symmetry is a coordinate transformation leaving the equations of motion unaltered. In the Lagrangian formulation of analytical mechanics, a symmetry will manifest as any coordinate transformation leaving the action invariant. This notion of symmetry was generalized in quantum physics, where actions also are invariant under phase multiplication of the fields representing particles. In modern terms, we speak of a symmetry as any transformation leaving the action, or in greater generality, the quantum partition function, unchanged.

Modern physics' fascination with symmetries is due to Noether's theorem: a conserved quantity may be found from every continuous symmetry of a system, and likewise, to every conserved quantity in the system, there is a corresponding continuous symmetry. A symmetry is continuous if it includes transformations arbitrarily close to the identity transformation.

The most famous examples of continuous symmetries are the Poincaré transformations: translations, rotations and boosts to moving inertial frames. Invariance under time translation corresponds to the conservation of energy. Invariance under spatial translations and rotations corresponds to the conservation of linear and angular momentum, respectively. Another important example comes from quantum electrodynamics, where multiplying the fields representing the electrons by arbitrary complex phases leaves the action invariant. This symmetry corresponds to the conservation of electric charge.

As all closed systems are observed to satisfy these conservation laws, any action describing a theory of a closed system should possess these symmetries. Indeed, if we take the regularities of Nature to be encoded in conserved quantities, they are equally well encoded in symmetry. As such, the modern philosophy is that symmetry should be taken as the starting point in constructing new theories. This mindset has given us the immensely successful standard model of particle physics, and the immensely popular theory of supersymmetry.

In addition to continuous symmetries, some systems also exhibit discrete symmetries, such as time reversal and space inversion (parity). However, whereas the continuous symmetries are taken to be exact features of Nature itself, the discrete symmetries are approximative,[1] and they do not predict conservation laws. Also, while there are infinitely many distinct transformations for every continuous symmetry, there is usually a finite number of discrete symmetry transformations.

With the advent of the renormalization group formalism, it became apparent that coupling constants are not constant. Indeed, restricting the validity of a theory by transformations in length scale or energy scale will yield a new theory, better suited for the new scale. However, it is often the case that this new, scaled theory is of the exact same form as the old one, with new values of parameters such as mass and charge. The manner in which a parameter changes its value under scale transformations is encoded in the renormalization group, and can be visualized as a flow in the parameter space of the theory.

As such, Lagrangians should not only be taken to be functions of coordinates or fields, but also of the relevant parameters of the theory. But even though symmetries of coordinates and fields are well studied in the literature, symmetries in the parameter space of coupling constants are sparsely mentioned.

In this thesis, a theory of symmetry in the parameter space of the quantum Hall effect is presented. This symmetry greatly distinguishes itself from the commonly known symmetries in physics, in that it is discrete, yet includes infinitely many distinct transformations. Furthermore, while the renormalization group flow for most physical systems must be calculated by perturbation theory, the symmetry presented here is so rigid that it uniquely determines an exact flow in the parameter space of the quantum Hall effect. The theory also provides a natural unification of the integer and fractional quantum Hall effects.

1.1 Outline

This thesis covers three distinct theoretical topics, spread out over equally many chapters. A theme thoughout all of them is to motivate the theory of modular symmetry in the renormalization group flow of the quantum Hall parameter space.

The aim of the second chapter is to gain familiarity with the concepts of universality, the renormalization group, parameter flow, and the context in which they are relevant. Some results from conformal field theory are needed, and as that is

[1] With the exeption of CPT: simultaneous charge conjugation, space inversion and time reversal. Even though there exists systems that breaks each of these separately, the simltaneous action of all three are always observed to be a symmetry.

generally not a topic covered by a standard physics education, an introduction to the formalism of conformal field theory is also presented. The chapter concludes with some general remarks on the β-functions, which determine the parameter flows. Special attention will be given to the 2-dimensional case, since this is the case most relevant to this thesis, both in real space and parameter space.

The third chapter gives a comprehensive introduction to the quantum Hall effect. Both theoretical and experimental considerations are presented. The aim is to give the reader an intuitive understanding of the effect, introduce all relevant terminology and explain how the most successful theory of the effect − Chern-Simons theory − is insufficient in dealing with scaling phenomena. Lastly, some aspects of the β-functions for the quantum Hall effect are considered, in order to motivate the development in the fourth chapter.

In the fourth chapter, what is believed to be the behaviour of the quantum Hall parameter space is derived. It begins with an introduction to a model of hyperbolic geometry: the Poincaré half plane, which will be identified with the quantum Hall parameter space. Then, the theory of the modular group, its maximal subgroups and modular forms is reviewed. The β-functions for the various quantum Hall effects are identified with the modular forms for the maximal subgroups of the modular group. This leads to very specific phase diagrams for the quantum Hall effects.

The final chapter deals with experimentally testing the hypothesis of this very specific symmetry in the quantum Hall parameter space. First, a review of the assumptions that had to be made throughout the thesis in order to derive the theory of modular symmetry is presented. Then, the method we have employed to test the theory is introduced. The results are presented in a paper, written in collaboration with Carsten Andrew Lütken and Kristian Stølevik Olsen.

The thesis concludes with some considerations regarding future relevant work to further test the universality of modular symmetry in the quantum Hall effect.

1.2 A Note on Units

The parameters of the quantum Hall effect are the transversal (Hall) and longitudinal conductivities (or resistivities). As such, we will frequently work in units $e^2/h = 1$, making conductivities (and resistivities) dimensionless.

Chapter 2

Critical Phenomena, Universality and the Renormalization Group

In all of physics, truly in all of science, the notion of *scale* is imperative. Every theory is limited by some sort of scale. In physics this often manifests as constraints on energy, temperature or system size, outside of which the theory breaks down and loses its ability to predict experimental outcomes. On different scales, different theories are needed to describe observable phenomena. This chapter introduces some scaling-related core concepts important to this thesis.

The choice was made to introduce the renormalization group (RG) concept through the language of statistical mechanics, and make the connection to quantum field theory (QFT), rather than the direct QFT approach. This is because, in statistical mechanics, the idea of universality and the renormalization group arises naturally from a discussion of critical phenomena. This context is more relevant to this thesis than the usual motivation of ultraviolet divergence control of loop diagrams from relativistic QFT we know and love. Even though the statistical mechanics method is more instructive and intuitive, a discussion of the renormalization group is of course incomplete without a discussion of field theory. From statistical mechanics we may then infer the behaviour of the RG of a Euclidean QFT. This is needed to introduce the β-functions, which are of paramount importance.

2.1 Emergence

The notion of scale-dependent theories is most obvious in the case of scaling with system size. For instance, one does not need to know atomic physics to develop a theory for biological cells. The collective behaviour of atoms constituting a cell is

radically different from the behaviour of a single atom. Likewise, the behaviour of a human is radically different from the behaviour of a single cell and the behaviour of large groups of people may differ significantly from the behaviour of a single person.

Of course, there may be advantages in knowing some aspects of atomic physics in order to explain the inner workings of a cell, knowing some aspects of biology to explain the behaviour of humans and knowing some psychology in order to explain the behaviour of large groups of people. But social science is not just applied psychology, psychology is not just applied biology, biology is not just applied chemistry and chemistry is not just applied physics. At each step of that ladder, new assumptions and generalizations must be made and new laws governing the field are discovered, not derived from the previous step of the ladder. The bottom line is: *You do not need atomic physics in order to study group psychology.*

The same core concept applies to purely physical systems. The collective behaviour of the multitude of particles constituting condensed matter systems may differ significantly from the behaviour of a single particle. In these systems, the degrees of freedom governing the micro-physics of each particle are effectively invisible to us as observers. Any theory aiming to describe the system should ignore these microscopic quantities and focus on the system as a whole. *The interactions of a host of particles gives rise to phenomena that can not be understood in terms of single-particle models.* These macroscopic phenomena are indeed independent of many quantum level details.

This is known as the principle of *emergence*. Aristotle wrote in his Metaphysics that "*The whole is greater than the sum of its parts*", though in quantum mechanics, Philip W. Anderson may have said it best, simply: "*More is different*" [1].

Closely related to the concept of emergence is that of *universality*. Before exploring that topic, we first need a detour into phase transitions and critical phenomena.

2.2 Statistical Mechanics and Phase Transitions

This section provides a quick recap of some aspects of statistical mechanics so that the idea of universality may be introduced. This and the following few sections largely follow the discussion in corresponding sections of [2] and [3], and chapter 8 of [4].

We study a system allowed to exchange heat with the environment at fixed tem-

perature: the canonical ensemble. Assuming the energy states of the system is determined by a microscopic multiparticle Hamiltonian $\mathcal{H}$, we start from the partition function:

$$Z = \sum_r e^{-\beta \mathcal{H}(r)}, \tag{2.2.1}$$

where the sum is taken over all microscopic states r, $\mathcal{H}(r)$ is the energy of state r and $\beta = 1/kT$, with k Boltzmann's constant and T the temperature of the system. The partition function contains all the statistical information about the system. The probability of the system being in a particular microstate r is

$$P(r) = e^{\beta(F - \mathcal{H}(r))}, \tag{2.2.2}$$

with F the free energy, given by

$$F = -\frac{1}{\beta} \ln Z. \tag{2.2.3}$$

All the macroscopic thermodynamic properties of the system (i.e. pressure, magnetization, internal energy) are found by differentiating the free energy with respect to the thermodynamic parameters (i.e. volume, applied magnetic field, temperature). Interchanging the thermodynamic properties and parameters is possible through Legendre transformations of the free energy. Lastly, the ensemble average of any given quantity, A, is given in terms of the partition function:

$$\begin{aligned} \langle A \rangle &= \frac{1}{Z} \sum_r A(r) e^{-\beta \mathcal{H}(r)} \\ &= \frac{1}{\beta Z} \left(\frac{\partial}{\partial y} Z[y] \right)_{y=0}, \end{aligned} \tag{2.2.4}$$

with the definition $Z[y] = \sum \exp\{-\beta[\mathcal{H}(r) - yA(r)]\}$, i.e. including a source term in the Hamiltonian.

Another important statistical function is the correlation function. For a continuous system:

$$\Gamma(\mathbf{x}, \mathbf{y}) = \langle \sigma(\mathbf{x}) \sigma(\mathbf{y}) \rangle - \langle \sigma(\mathbf{x}) \rangle \langle \sigma(\mathbf{y}) \rangle, \tag{2.2.5}$$

measuring the correlation among the fluctuations of the variable σ at positions $\mathbf{x}$ and $\mathbf{y}$. As the spatial separation $r = |\mathbf{x} - \mathbf{y}|$ increases, these fluctuations are expected to be less and less correlated, and Γ approaches zero. In general (see section 2.4 of [3]):

$$\Gamma(\mathbf{x}, \mathbf{y}) \sim r^{-\tau} e^{-r/\xi}, \tag{2.2.6}$$

where τ is some number and ξ is called the *correlation length* of the system. The correlation length gives a measure of the size of correlated portions of the system. For a discrete system, we use the notation $\Gamma_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$.

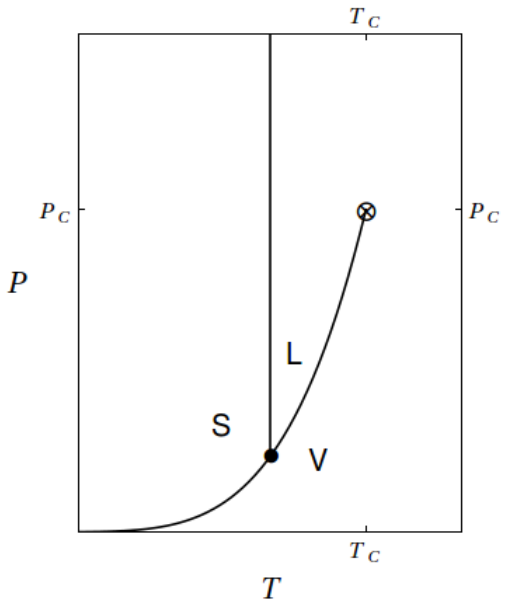

Figure 2.2.1: Illustration of a pressure-temperature phase diagram for a system with solid (S), liquid (L) and vapor (V) phases. The triple point (•), where all three phases can coexist simultaneously, and critical point (⊗), where the liquid and vapor phases become indistinguishable, are shown.

In studying how a system behaves under variation of the thermodynamic parameters, a powerful tool is the *phase diagram*, see Fig. 2.2.1. Now, a phase is a region in the phase diagram where the thermodynamic properties are analytic functions of the thermodynamic parameters. A *phase transition* is a point, line or surface in the phase diagram where the thermodynamic properties are nonanalytic [2]. For the system in Fig. 2.2.1, a phase transition will occur by crossing one of the solid lines and will manifest as a sudden change in volume per particle (density) and release or absorption of latent heat.

At sufficiently high temperatures and pressures, the liquid and gas phases of the system in Fig. 2.2.1 become indistinguishable. The point (T_C, P_C) at which this occurs is called the *critical point*. This point marks the conception of the gas and liquid phases coexisting as a solitary phase. Further increase in temperature or pressure will not lead to the discontinuous behaviour that characterize a phase transition and the system is said to be in the *supercritical phase*.

At the critical point itself, a phase transition occurs, but without latent heat or abrupt volume change. In general, a phase transition where latent heat is involved is called a first order phase transition, and a phase transition without latent heat is said to be continuous (sometimes referred to as second order)[1] [5].

[1]The terms first and second order comes from the Ehrenfest method of classifying phase

It is the continuous phase transitions, which occur at critical points, that are of interest to us here as the next sections will elucidate.

2.3 Critical Exponents

In order to properly define the term universality, the critical exponents must be introduced. The critical exponents measure the qualitative asymptotic behaviour of a system near a critical point. First, define

$$t = \frac{T - T_C}{T_C}, \tag{2.3.1}$$

as a measure of the deviation from the critical temperature. Then, two quantities must be identified: a thermodynamic *order parameter*, m, and the corresponding *ordering field*, h, with the property that

$$\lim_{h\to 0} m = 0 \qquad \text{for } t > 0,$$
$$\lim_{h\to 0} m \neq 0 \qquad \text{for } t \leq 0.$$

Furthermore, h must be defined so that the statement $t, h \approx 0$ describes the vicinity of a critical point.

For examle, in the liquid-gas transition in Fig. 2.2.1, an appropriate order parameter is the density differential $m = (\rho_l - \rho_C)$, ρ_l and ρ_C being the liquid and critical densities, and the corresponding ordering field would be the pressure differential $h = (P - P_C)$. For a ferromagnetic system (see the next section), the order parameter is naturally identified with the net magnetization, $m \propto M$, and the ordering field with the applied magnetic field $h \propto H$. [2]

Then, the critical exponent, λ_F, for a thermodynamic function F is defined as [3]:

$$\lambda_F = \lim_{t,h\to 0} \frac{\ln|F(t)|}{\ln|t|}, \tag{2.3.2}$$

such that

$$F(t) \sim |t|^{\lambda_F} \qquad (t, h \approx 0), \tag{2.3.3}$$

whenever the limit (2.3.2) exists.[2] A precise temperature-dependence to arbitrary order would be of the form:

$$F(t) = C|t|^{\lambda_F}\left(1 + a_1 t^{\lambda_1} + a_2 t^{\lambda_2} + ...\right), \tag{2.3.4}$$

transitions. A first order transition is recognized by the discontinuity of a first derivative of the free energy (i.e. volume or magnetization), while a second order transition is recognized by discontinuity of a second order derivative (i.e. heat capacity or susceptibility).

[2]It is conventional to define two critical exponents; one for the limit $t \to 0^+$ and another for the limit $t \to 0^-$, labeled by a prime. As it is generically the case that these are equal, this convention will not be followed here. See [6] for a counter example.

for a set of positive exponents $\{\lambda_1, \lambda_2, ...\}$.

The critical exponent describing the way the order parameter approaches zero close to a critical point is called β:

$$m \sim |t|^{\beta} \quad (h \to 0,\ t \to 0^{-}). \tag{2.3.5}$$

The one describing the divergence of the isothermal compressibility κ_{T_C} (liquid-gas system) or the low-field susceptibility χ_0 (magnetic system) is called γ:

$$\kappa_{T_C},\ \chi_0 \sim \left(\frac{\partial m}{\partial h}\right)_{|t|,h\to 0} \sim |t|^{-\gamma}, \quad (h,\ |t| \to 0). \tag{2.3.6}$$

Likewise, the behaviour of specific heats are described by the critical exponent α:

$$C_X \sim |t|^{-\alpha} \quad (h,\ |t| \to 0), \tag{2.3.7}$$

where $X = H$ with $H = 0$ for a magnetic system and $X = V$with $V = V_C$, the critical volume, for a liquid-gas system. The exponent δ is defined by

$$m \sim h^{1/\delta} \quad (t = 0,\ h \to 0), \tag{2.3.8}$$

so that it describes the shape of the critical isotherm.

Furthermore, as will be shown in the next section, at criticality, the correlation length diverges:

$$\xi \sim |t|^{-\nu} \quad (h,\ |t| \to 0), \tag{2.3.9}$$

defining the exponent ν. This means that the correlation function dependence on the correlation length, eq. (2.2.6) is invalid at criticality. At criticality the correlation function shows a power law dependence on r, defining another critical exponent η [3]:

$$\Gamma(r) \sim \frac{1}{r^{2-d+\eta}} \quad (|t| = 0,\ h \to 0), \tag{2.3.10}$$

where d is the dimension of the space we are studying.

2.4 Landau Theory of Phase Transitions

A theory exploring the critical phenomena of phase transitions is Landau's phenomenological theory. Though simple in form, this theory serves to illustrate one of the most powerful concepts in physics: universality. A keyword here is "phenomenological"; the analysis is carried out with minimal knowledge of the microscopic degrees of freedom.

Assume the system of study is a ferromagnetic one, consisting of spins, localized on a simple cubic lattice, in the presence of an applied magnetic field H. The thermodynamic identity for the free energy, F, of such a magnetic system, assuming constant volume, is given as

$$\mathrm{d}F = -M\mathrm{d}H - S\mathrm{d}T, \tag{2.4.1}$$

M being the net magnetization of the system, and S the entropy. The free energy thus is a function of only H and T, and by differentiating at constant temperature, one finds the magnetization:

$$M = -\left(\frac{\partial F}{\partial H}\right)_T, \tag{2.4.2}$$

and the isothermal susceptibility:

$$\chi_T = \left(\frac{\partial M}{\partial H}\right)_T = -\left(\frac{\partial^2 F}{\partial H^2}\right)_T. \tag{2.4.3}$$

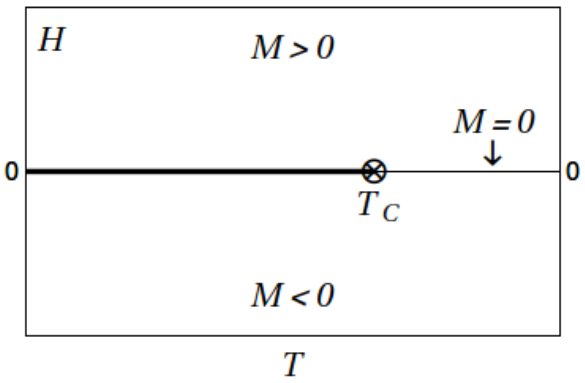

Figure 2.4.1: The phase diagram for a ferromagnet in the plane of applied magnetic field and temperature. At low temperatures the system undergoes a first order phase transition when the magnetic field is tuned through zero-field strength. Above the critical temperature the two phases become indistinguishable.

In this system, the onset of a first order phase transition would be recognized by a discontinuity in the magnetization, however, the system also exhibits a second order phase transition corresponding to a discontinuity in the susceptibility. The phase diagram is given in Fig. 2.4.1. At temperatures below T_C, the system undergoes a first-order phase transition when H is tuned through zero field strength, characterized by a discontinuous jump in M. At higher temperatures, the system may transmute continuously from positive to negative magnetization, and the notion of different phases becomes meaningless. In summation: $M \neq 0$ for $H = 0$ at $t \leq 0$ and $M = 0$ for $H = 0$ at $t > 0$.

As such, M possesses the properties of an order parameter and H the properties of an ordering field, defined in the previous section. We want to calculate the critical exponents of the system in order to extract the qualitative behaviour of the system near a critical point. In the analysis, we want M to be a thermodynamic parameter and must use the Gibbs free energy, G, as our potential for the system. The thermodynamic identity for G is in this case given by:

$$\mathrm{d}G = -S\mathrm{d}T + H\mathrm{d}M. \tag{2.4.4}$$

The applied magnetic field is thus given by differentiation of G at constant temperature:

$$\left(\frac{\partial G}{\partial M}\right)_T = H. \tag{2.4.5}$$

As Fig. 2.4.1 shows, close to the critical point, $M \approx 0$, so we may represent G as a power series in M. Following section 12.9 of [2] and chapter 8 of [4], the free energy is symmetrical with respect to the reflection $M \mapsto -M$, so only even powers of M may contribute:

$$G(M,t) = A(t) + B(t)M^2 + C(t)M^4 + ..., \tag{2.4.6}$$

The expansion of G is to be truncated at the M^4-term. This is not only necessary in order to find solutions, but indeed, only these terms may contribute to the critical behaviour [3].

Now:

$$H = \left(\frac{\partial G}{\partial M}\right)_T = 2B(t)M + 4C(t)M^3. \tag{2.4.7}$$

As the system is in the vicinity of a critical point, $|t| \ll 1$, and the coefficients B and C may be expanded about $t = 0$. Only non-negative powers of t may contribute in order to keep the free energy finite, so

$$B(t) = \sum_{k=0}^{\infty} b_k t^k, \qquad C(t) = \sum_{k=0}^{\infty} c_k t^k. \tag{2.4.8}$$

In order to get physical results (M real, positive susceptibility and $\mathrm{sgn}(H) = \mathrm{sgn}(M)$) we also need $b_0 = 0$, $b_1 > 0$ and $c_0 > 0$, so to lowest order:

$$H = 2b_1 tM + 4c_0 M^3. \tag{2.4.9}$$

At $H = 0$, we then get two possible solutions for M: $M = 0$ and $M = \pm\sqrt{-b_1 t/2c_0}$. The former is the already known solution for $t > 0$, whereas the latter is the value of spontaneous magnetization at zero field strength and $t < 0$.

Then,

$$|M| \sim |t|^{1/2} \quad (H \to 0, t \to 0^-), \tag{2.4.10}$$

yielding $\beta = \frac{1}{2}$.

The susceptibility is given by eq. (2.4.3):

$$\chi = \left(\frac{\partial H}{\partial M}\right)^{-1} = \frac{1}{2b_1 t + 12c_0 M^2}. \tag{2.4.11}$$

For $t \to 0^+$, $M \to 0$, so $\chi \approx 1/2b_1 t$. For $t \to 0^-$, $M \to \sqrt{-b_1 t/2c_0}$, so $\chi \approx -1/4b_1 t$. In either case:

$$\chi \sim |t|^{-1} \quad (H \to 0, t \to 0), \tag{2.4.12}$$

yielding $\gamma = 1$.

The heat capacity is given as

$$C \sim \left(\frac{\partial^2 G}{\partial^2 t}\right), \tag{2.4.13}$$

and may be found through tedious differentiation of eq. (2.4.6). The result includes a non-zero constant term, thus giving $\alpha = 0$.

The exponent δ is found by setting $t = 0$ in eq. (2.4.9), yielding

$$H = 4c_0 M^3, \tag{2.4.14}$$

so $\delta = 3$.

To study correlations, we use a field theoretic approach outlined in chapter 8 of [4]. Consider the magnetization to be the integral of a local spin field $s(\mathbf{x})$, and redefine the free energy to be a functional of this field:

$$G = \int d^3x \left(\frac{1}{2}(\nabla s)^2 + bts^2 + cs^4\right), \tag{2.4.15}$$

the first term is just the simplest way to introduce the tendency of nearby spins to align. The applied magnetic field is then, analogously, found from the variational problem

$$H(\mathbf{x}) = \frac{\delta G}{\delta s(\mathbf{x})} = -\nabla^2 s(\mathbf{x}) + 2bts(\mathbf{x}) + 4cs^3(\mathbf{x}), \tag{2.4.16}$$

in the thermodynamic limit.[3] For $t > 0$ and $H = 0$, we have seen that $M = 0$, so close to the critical point for $t > 0$, it is fair to assume that s will be small

[3]Remember that the *functional derivative* is defined by $\frac{\delta}{\delta s(\mathbf{x})} \int d^3y[s(\mathbf{y})\phi(\mathbf{y})] = \phi(\mathbf{x})$. For terms involving ∇s, one must integrate by parts before applying the functional derivative: $\frac{\delta}{\delta s(\mathbf{x})} \int d^3y[\nabla s(\mathbf{y})\phi(\mathbf{y})] = \nabla\phi(\mathbf{x})$, taking the surface term to vanish.

and we may ignore the s^3 term. Furthermore, we wish to study correlations of spins, so we let H be localized at the point $\mathbf{y}$: $H(\mathbf{x}) = H_0\delta^{(3)}(\mathbf{x}-\mathbf{y})$. Then, eq. (2.4.16) simplifies to

$$\left(-\nabla^2 + 2bt\right) s(\mathbf{x}) = H_0\delta^{(3)}(\mathbf{x}-\mathbf{y}), \tag{2.4.17}$$

with $s(\mathbf{x})$ now measuring the response at position $\mathbf{x}$ from an applied magnetic field at $\mathbf{y}$, meaning it is proportional to $\Gamma(\mathbf{x},\mathbf{y})$. The solution is found through Fourier transformation:

$$\begin{aligned}
\Gamma(\mathbf{x},\mathbf{y}) &\sim \int \frac{d^3k}{(2\pi)^3}\frac{H_0 e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{|\mathbf{k}|^2+2bt} \\
&= \frac{H_0}{4\pi^2}\int_0^\pi\int_0^\infty d\theta dk k^2 \frac{e^{ikr\cos\theta}}{k^2+2bt}\sin\theta \\
&= \frac{H_0}{i4\pi r}\int_{-\infty}^{\infty} dk \frac{k e^{ikr}}{k^2+2bt} \\
&= \frac{H_0}{4\pi}\frac{1}{r}e^{-r\sqrt{2bt}},
\end{aligned} \tag{2.4.18}$$

where we have transformed to polar coordinates and closed the contour in the upper half plane, picking up the residue at $k = i\sqrt{2bt}$, and $r = |\mathbf{x}-\mathbf{y}|$. This gives (compare eq. (2.2.6)) $\xi = 1/\sqrt{2bt}$, and thus

$$\nu = \frac{1}{2}. \tag{2.4.19}$$

Finally, as the system is 3-dimensional ($d = 3$), we also get (eq. (2.3.10)):

$$\eta = 0. \tag{2.4.20}$$

2.5 Universality

Calculating the critical exponents for this specific system is a nice exercise, but what is the use? First of all, the important realization is the fact that our only assumption in calculating the critical exponents of this system is the reflection symmetry of the free energy with respect to the order parameter: $G(m) = G(-m)$. There is a host of systems with this property and so, they should all have critical exponents of a similar form.

Next is an important observation: none of the critical exponents calculated here shows any dependence on the microscopic parameters of the system. Even though

the exact temperature dependence of any thermodynamic property near a critical point is a function of some parameters that describes the microscopic structure of the system (b_1, c_0, b, c), the critical exponents themselves do not show this dependence. As such, they should in fact be *equal* in each system exhibiting the reflection symmetry.

This is the famous *universality* (of the critical exponents).

Alas, the concept of universality is not generous enough to have every critical phenomenon described by a single set of critical exponents. However, Nature has seen it fit to divide its critical phenomena into *universality classes*, each class containing a number of physically distinct systems sharing a common set of critical exponents. Notable examples are the mean field class, Ising class and, more generally the N-state Potts class (Ising with N possible spin-values) in a given dimension. We return to the systematics of dividing critical phenomena into universality classes in our discussion of conformal field theory.

2.5.1 An Illustration of Universality

To illustrate just how different the systems can be, all the while being qualitatively described by the same theory, Table 2.1 displays experimentally estimated values for the critical exponents for some systems in the 3-dimensional Ising universality class [7]. Also included are theoretical estimates, done in greater detail than the simplistic Landau theory used as an example in the last section.

The uniaxial magnet is the system used as an example in the calculation from the last section. With reference to Table 2.1: in [8], the magnet used for experiments was a ferromagnet ($SrRuO_3$), and in [9] and [10] it was an antiferromagnet (FeF_2). The liquid-gas system is heavy water (D_2O) in [11], carbon dioxide (CO_2) in [12] and [13], and pentafluoroethane (CF_3CF_2) in [14].

In the mixing transition in binary fluid mixtures the order parameter is the molar concentration fraction. The mixing phase transition refers to the mixing of the two separated fluids into one mixed fluid, which happens spontaneously at critical values of temperature and concentration [7]. In [15], the mixing of aniline ($C_6H_5NH_2$) and cyclohexane ($C_{12}H_6$) is studied. In [16], it is mixing of dodecane ($C_{12}H_{26}$) and ethanol (C_2H_6O) and in [17] it is cyclohexane and polysterene.

Micellar systems are particularly exotic. These consist of lipids, such as fatty acids: molecules with a hydrophilic end, soluble in water, and a hydrophobic end which is near insoluble in water. When a large number of these molecules are submerged in water, they tend to form structures such as spheres or cylinders, with the hydrophobic end pointing inwards. These structures are collectively

System [References]	γ	ν	η	α	β
Uniaxial magnet [8], [9],—, [10], [8]	1.14(7)	0.64(1)	—	0.11(3)	0.34(2)
Liquid-gas [11], [11], [12], [13], [14]	1.14(5)	0.62(3)	0.042(6)	0.111(1)	0.341(2)
Binary fluids [15], [15], [15], [16], [17]	1.244(42)	0.636(31)	0.045(11)	0.12(1)	0.335(5)
Micellar system [7]	1.26(5)	0.63(2)	0.039(4)	—	0.34(1)
Ionic solutions [18]	—	0.641(3)	—	—	0.34(1)
Theoretical [19]	1.2373(2)	0.63012(16)	0.03639(15)	0.110(2)	0.32653(10)
Landau theory (mean field)	1	0.5	0	0	0.5

Table 2.1: Experimentally estimated values for some systems in the 3-dimensional Ising universality class. For each measured exponent, the corresponding reference has been placed in the first column, in an attempt to minimize the cluttering of the table. For the micellar system, the original sources were inaccessible and the reader is instead referred to the review article [7]. See the text for an explanation of the different systems. For a more complete review, see [7].

called micelles. Further increasing the concentration of lipids causes the phase transition: micelles clustering togheter forming cubic (in the case of spherical micelles) or hexagonal (in the case of cylindrical micelles) arrays with water filling the intervening spaces. In some systems there is also a second transition to the lamellar phase: the lipids form sheets, two lipids thick, with the hydrophilic end pointing outwards, while the lipids themselves move freely within the sheet as in a 2-dimensional liquid. [20] [3]

In ionic solutions, interactions at criticiality are governed by the long-range Coulomb force, rather than the short-range interactions in fluids without free charges, making them qualitatively distinct systems. Even though the Ising model explicitly describes short-range interactions, i.e. only nearest neighbor interactions, there seems to be a small window where Ising-like behaviour is observed in Coulombic systems. In [18] the phase transition between two distinct liquid phases (a *polyamorphism*) in several ionic solutions is studied.

Finally, the theoretical estimates in [19] are made from 25th-order high temperature expansions within the 3-dimensional Ising model on a simple cubic lattice. The Hamiltonian considered is the simplistic

$$\mathcal{H} = -\beta \sum_{\langle i,j \rangle} \phi_i \phi_j + \sum_i V(\phi_i^2), \tag{2.5.1}$$

with $\beta = 1/T$, $\langle i, j \rangle$ denoting a sum over nearest neighbors, ϕ_i being real variables. Estimates are made with V being both the ϕ^4:

$$V(\phi^2) = \phi^2 + \lambda_4(\phi^2 - 1)^2, \tag{2.5.2}$$

and the ϕ^6:

$$V(\phi^2) = \phi^2 + \lambda_4(\phi^2 - 1)^2 + \lambda_6(\phi^2 - 1)^3, \tag{2.5.3}$$

lattice model potentials, λ_4 and λ_6 being constants. Impressively accurate results are also obtained using the shockingly simple simplified version of a spin-1 Ising Hamiltonian:

$$\mathcal{H} = -\beta \sum_{\langle i,j \rangle} s_i s_j + D \sum_i s_i^2, \tag{2.5.4}$$

with $s_i \in \{-1, 0, 1\}$ and D a constant, yielding $\gamma \approx 1.237$, $\alpha = 0.112(5)$ and $\nu \approx 0.63$. The reader is referred to [19] for further details.

The compliance of experimental data across significantly dissimilar physical systems is simply astonishing. Even though Nature pertains to more than a single universality class, the power of universality should be clear: *There exists large classes of systems where each class may be qualitatively described by a single, often simple, Hamiltonian near criticality.*

2.6 Renormalization

Even with the universality class concept firmly in place, the understanding of critical phenomena is still not complete. This section introduces the renormalization group, the key to a modern understanding of critical phenomena.

2.6.1 Real Space Scaling

A more general theory of critical phenomena was first attempted by Kadanoff in 1966 [21] and later developed by Wilson in 1971 [22] into the theory of the *renormalization group* (RG).[4] The core idea is simple enough: at criticality, the

[4]See [23] for a thorough introduction in Wilson's own words.

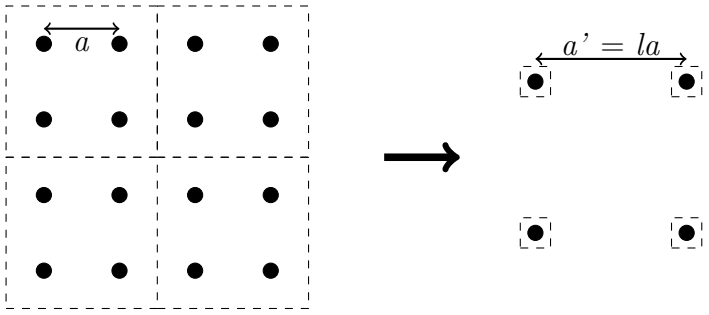

Figure 2.6.1: Illustration of the renormalization procedure in the 2-dimensional Ising model with $l = 2$.

correlation length diverges, and as such, the short-range fluctuations of the system should be irrelevant. The object is to find a consistent algorithm to trace out these fluctuations and study the effect of this action on the remaining degrees of freedom. To do this, one investigates how the system behaves under a *change of scale*, e.g., in the Ising model, a change in lattice constant (the fundamental shortest distance in a lattice based system) $a \mapsto a' = la$ with $l > 1$. In doing this one obtains a new, *renormalized* system with a new fundamental shortest distance, a'. In this case, the procedure amounts to clustering groups of l^d spins together and assigning them a single spin value (± 1), see Fig. 2.6.1.

Performing this in a statistical mechanics model, a fundamental criterion is that the partition function is left unchanged. In the case of the 1-dimensional Ising model, which may be solved exactly, one finds that the new partition function can be written in a similar form as the original, with new values of the parameters of the Hamiltonian (see e.g. [3] or [2]).

That is to say, when the correlation length of the system is large enough so that we may disregard short-length fluctuations, this can be achieved by regarding a new, transformed system with a reduced number of degrees of freedom and rescaled values of temperature, coupling constants, applied magnetic field etc. The next subsection formalizes this notion.

2.6.2 The Renormalization Group

This subsection follows [2] and [3]. First, note that a scale transformation of the form $a \mapsto a' = la$ implies that all lengths, measured in units of lattice spacings, must scale as

$$r \mapsto r' = l^{-1} r. \tag{2.6.1}$$

Equivalently, through dimensional analysis, all momenta must scale as

$$k \mapsto k' = lk. \tag{2.6.2}$$

Thus, the new system is one of shorter distances and larger momenta.

More generally, we define the scaling factor l in terms of the reduction of the number of degrees of freedom $N \mapsto N'$:

$$l^d = N/N', \tag{2.6.3}$$

and a renormalization group (RG) transformation as one that scales lengths, momenta and number of degrees of freedom according to (2.6.1), (2.6.2) and (2.6.3).

Assume we are working with a Hamiltonian, $\mathcal{H}(\{g_i\})$, dependent on any number of parameters g_i, $i = 1, 2, \ldots$ For convenience, collect the parameters in a vector: $\boldsymbol{g}$. A scale transformation will also transform these parameters:

$$\boldsymbol{g} \mapsto \boldsymbol{g}' = R_l(\boldsymbol{g}), \tag{2.6.4}$$

with R_l the appropriate renormalization group operator corresponding to the scale factor l. Successive use of this transformation will lead to a sequence of vectors $\boldsymbol{g}', \boldsymbol{g}'', ..., \boldsymbol{g}^{(n)}$, related by $\boldsymbol{g}^{(n)} = R_l(\boldsymbol{g}^{(n-1)}) = \cdots = R_l^n(\boldsymbol{g})$.

This sequence of vectors may be regarded as a "flow" in the vector space spanned by the g_i's; each transformation yields a new point in this space and successive transformations trace out a path taken by the Hamiltonian in the process of scaling. This vector space will hence be termed the *parameter space* of the system.

Now, the transformation R_l may have a *fixed point* $\boldsymbol{g}^*$, such that

$$R_l(\boldsymbol{g}^*) = \boldsymbol{g}^*. \tag{2.6.5}$$

As any length in the problem, the correlation length scales as $\xi' = b^{-1}\xi$, however, at the fixed point, this implies $\xi(\boldsymbol{g}^*) = l^{-1}\xi(\boldsymbol{g}^*)$, and as such, $\xi = \infty$ or $\xi = 0$. The latter case is of no interest here, but the former is exactly what is expected at a critical point. Note, however, that as the RG transformation only decreases ξ, and ξ is infinite at the critical point $\boldsymbol{g}^*$, it must also be infinite at every preceding point $\boldsymbol{g}^{(n)}$ of the RG flow. The locus of points flowing into a critical point constitutes a surface of system configurations with infinite correlation length in parameter space and is thus termed a critical surface.

To further study the critical behaviour of the system, assume we are in the vicinity of a fixed point:

$$\boldsymbol{g} = \boldsymbol{g}^* + \boldsymbol{\epsilon}.$$

The effect of an RG transformation will then be

$$g_i \mapsto g_i^* + \epsilon_i' = [R_l(\boldsymbol{g}^* + \boldsymbol{\epsilon})]_i = [R_l(\boldsymbol{g}^*)]_i + \left.\frac{\partial [R_l(\boldsymbol{g})]_i}{\partial g_j}\right|_{\boldsymbol{g}=\boldsymbol{g}^*} \epsilon_j + ..., \tag{2.6.6}$$

so to first order

$$\boldsymbol{\epsilon}' = R_l'(\boldsymbol{g}^*)\boldsymbol{\epsilon}, \tag{2.6.7}$$

$R_l'(\boldsymbol{g}^*)$ denoting the Jacobian matrix of R_l evaluated at $\boldsymbol{g}^*$. Now, given that the eigenvectors, $\{\boldsymbol{v}_i\}$, of $R_l'(\boldsymbol{g}^*)$ is a complete set[5], $\boldsymbol{\epsilon}$ may be expanded in their basis:

$$\boldsymbol{\epsilon} = \sum_i u_i \boldsymbol{v}_i.$$

Applying eq. (2.6.7) then yields

$$u_i' = \lambda_i u_i, \tag{2.6.8}$$

λ_i being the eigenvalue corresponding to $\boldsymbol{v}_i$. The u_i's are linear combinations of the original g_i's, and may be interpreted as the generalized coordinates of the system, analagous to those encountered in analytical mehcanics [2]. Furthermore, if two successive RG transformations of scale factors l_1 and l_2 are performed, the eigenvalues must obey $\lambda(l_1)\lambda(l_2) = \lambda(l_1 l_2)$, constraining them to be of the form

$$\lambda_i = l^{y_i},$$

for some exponent y_i independent of l [3]. The RG transformation (2.6.8) then reads

$$u_i' = l^{y_i} u_i.$$

The exponent y_i is called the *scaling dimension* of the *scaling field* u_i.

The scaling fields u_i are naturally classified into three categories depending on the value of λ_i (y_i):

$\lambda_i > 1$: ($y_i > 0$) In this case, u_i grows larger for each iteration of the RG flow. Such a scaling field is termed a *relevant variable*. As u_i can only grow through the RG process, the system is driven away from the critical point. Unless all relevant variables are zero at the beginning of the RG flow, the system must eventually flow away from the critical point.

$\lambda_i < 1$: ($y_i < 0$) With each iteration u_i decreases, approaches zero, and the system flows towards the fixed point. The scaling field is then called an *irrelevant variable*. If all the scaling fields of the problem are irrelevant, the system will flow to the critical point, and the RG trajectory must lay on the critical surface. As such, the critical surface must be the subspace spanned by the irrelevant variables.

$\lambda_i = 1$: ($y_i = 0$) u_i will not grow nor decrease, and will not contribute to the flow of the system. Hence it is termed a *marginal variable*. In such cases, higher-order terms may be included in the expansion (2.6.6) to study their effect.

[5]Which is equivalent to the function R_l having a non-zero Jacobian determinant, a pretty common thing for well-behaved, physical functions.

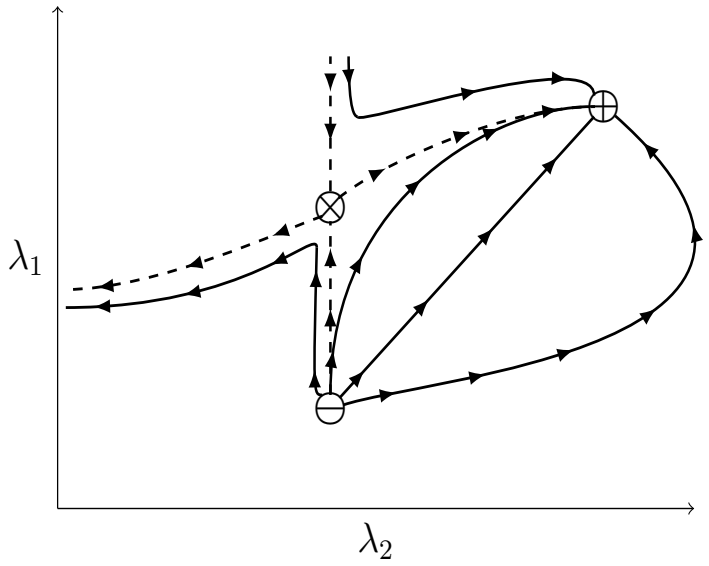

Figure 2.6.2: Illustration of a generic RG flow diagram for a system with two parameters. The repulsive, attractive and saddle points are marked as $\ominus$, $\oplus$ and $\otimes$ respectively. The dashed lines are the coordinate axes of the relevant and irrelevant variables with respect to the saddle point. These mark the separatrices of the system.

The critical points themselves, in turn, may be categorized by the relevant and irrelevant scaling fields of the problem: If, in the vicinity of a given critical point...

1. ... all scaling fields are relevant, the system will always flow away from the critical point, making it an unstable, *repulsive* fixed point. Such a point acts as a source of the flow and is denoted $\ominus$.

2. ... all scaling fields are irrelevant, the system always flows towards the critical point, making it a stable, *attractive* fixed point. Such a point acts as a sink in the flow and is denoted $\oplus$.

3. ... some scaling fields are relevant and some are irrelevant, the system will flow towards the critical point along the axes of the irrelevant variables and away from the critical point along the axes of the relevant ones. The system will only reach the critical point if all relevant variables are zero. Thus, it is *saddle point* of the flow. These points will be denoted $\otimes$.

The generic situation is illustrated in Fig. 2.6.2. Given an initial parameter configuration of the system, repeated RG operations will cause the system to flow in parameter space. The separatrices are the flowlines which exactly flows into or out of a saddle point. The separatrix corresponding to an irrelevant variable (the vertical in Fig. 2.6.2) partitions the space of parameter configuations into two classes: each class flowing towards a different attractive fixed point. The

separatrix corresponding to a relevant variable partitions the system into classes where within each class, every flow line originates from a given repulsive fixed point.

When a system exhibits more than one critical point, the critical exponents of each point may differ, and as such, the system may belong to several universality classes. If, however, all the critical points belong to the same universality class, such that the set of critical exponents are equal at every critical point, the system is said to *super-universal*. In such cases, the flow diagram itself will be termed a super-universality class, describing the parameter flow of every system belonging to it.

2.7 Classical to Quantum: Fields and Path Integrals

Any condensed matter system consists of rigid atoms, often in some kind of (dirty) lattice structure. Then, it is natural to think of such systems as having discrete positions with a fundamental shortest distance, the lattice constant. However, when one studies large-scale phenomena, the system is perceived as continuous and it makes sense to treat it that way. Then, the discrete variables should be replaced by *fields*. This was done with the spin-field in the Gibbs free energy in section 2.4.

In taking the continuum limit of a single discrete variable, s, one replaces sums with integrals:

$$\sum_s e^{\beta\mathcal{H}(s)} \to \int ds e^{-\beta\mathcal{H}(s)},$$

taking $ds = 1$, as it is discrete. We now want to take the continuum limit of a configuration of discrete variables, $\{s_i\}$, replacing them all with a local field $s(\mathbf{x})$. This means we need to sum over every possible, continuously varying field configuration. The natural extension is then the *path integral*:

$$Z = \sum_{\{s\}} e^{-\beta\mathcal{H}(\{s\})} \to \left[\prod_{s(\mathbf{x})} \int ds(\mathbf{x})\right] e^{-\beta\mathcal{H}[s(\mathbf{x})]} := \int \mathcal{D}s(\mathbf{x}) e^{-\beta\mathcal{H}[s(\mathbf{x})]}, \qquad (2.7.1)$$

with the Hamiltonian now a functional of the field $s(\mathbf{x})$. The product is (symbolically) taken to be over every possible field $s(\mathbf{x})$, such that $\int \mathcal{D}s(\mathbf{x})$ is viewed as an integral over all possible field configurations on the space of continuous functions where $s(\mathbf{x})$ lives.

This is a good place to give a reminder of the shape of the correlation function $\Gamma(\mathbf{x}, \mathbf{y})$ at the beginning of eq. (2.4.18). Readers familiar with quantum field

theory will recognize the *Feynman propagator* for the field s with mass $m = \sqrt{2bt}$ in a Euclidean scalar (ϕ^4) QFT. The validity of this mass would mean $m \sim \xi^{-1}$, and thus the field s is massless at criticality. This means [24], that the universal properties near the critical point are fully determined by the symmetries of terms in the Hamiltonian relevant at long distances. This is the idea of universality paraphrased in terms of field theory.

It is, in fact, possible to describe a magnet with a scalar field theory,[6] and in the context of quantum field theory, the partition function is replaced by its quantum analogue, the generating functional:

$$Z = \int \mathcal{D}\phi e^{iS[\phi]/\hbar},$$

where the measure $\mathcal{D}\phi$ can be taken to be over functions of coordinate space, as in eq. (2.7.1), or their Fourier components: $\mathcal{D}\phi = \prod_{\mathbf{k}} d\phi(\mathbf{k})$.[7] ϕ is now a scalar field, $S[\phi] = \int dt d^3x \mathcal{L}[\phi]$ is the action, where $\mathcal{L}$ is the Lagrangian density.

The role of the generating functional is parallell to the role of the moment generating function used in statistics, see e.g. [25]. If one considers a continuous random variable X, subject to a probability density function $f(x)$ such that $P(X = x) = f(x)dx$, the moment generating function is defined as

$$M_X(y) = \int e^{yx} f(x) dx.$$

All moments are found by differentiation of $M_X(y)$ as $\langle X^n \rangle = \frac{d^n}{dy^n} M_X(y)|_{y=0}$.

Analogously, with the introduction of a source field $J(x)$ taking the place of the variable y, the generating functional takes the form

$$Z[J] = \int \mathcal{D}\phi e^{i \int dt d^3x [\mathcal{L}[\phi] + J(x)\phi(x)]},$$

and all correlation functions are generated through functional derivatives (compare eqs. (2.2.4)):

$$\langle \phi(x_1) ... \phi(x_n) \rangle = \left(-i \frac{\delta}{\delta J(x_1)} \right) ... \left(-i \frac{\delta}{\delta J(x_n)} \right) \frac{Z[J]}{Z_0} \bigg|_{J=0},$$

$\langle \dots \rangle$ in this case denoting the time ordered vacuum expectation value $\langle 0 | T\{\dots\} | 0 \rangle$, T being the time ordering operator, and $Z_0 = Z[J = 0]$ is the normalization factor. This is, again, completely analagous to the statistical mechanics case of a partition function (eq. 2.2.4). For the details regarding the formalities of field quantization in the functional formalism, se e.g. chapter 9 of [4].

[6] This was hinted at in subsection 2.5.1, where the theoretical computations of the critical exponents were made using the ϕ^4 and ϕ^6 potentials.

[7] Recall that Fourier transformation $\phi(\mathbf{k})$ of the field $\phi(\mathbf{x})$ is defined as $\phi(\mathbf{k}) = \int d^3x \phi(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}$, such that $\phi(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \phi(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}$.

2.8 Quantum Renormalization

Critical behaviour still exists in QFTs, and all the qualitative features of an RG flow derived in subsection 2.6.2 are still valid. The quantum Hall effect (QHE) is, however, a quantum phenomenon and its natural language is thereby that of QFT. In this section the process of renormalization in QFT will be reviewed. This section follows [4], and in their spirit the work done here on out is in God's own units: $\hbar = c = 1$.

2.8.1 Momentum Space Scaling

Recall the definition of an RG transformation in subsection 2.6.2. There, the emphasis was on a change of length scale and $x \mapsto l^{-1}x$, in this section, we will focus on a change of *energy scale* and $k \mapsto lk$. To all systems exhibiting a shortest length scale, e.g. a lattice constant a, one can heuristically assign a highest energy scale:

$$\Lambda \sim \frac{1}{a} \, .$$

As spatial fluctuations only takes place over distances larger than a, so should only the Fourier modes with momenta $k < \Lambda$ contribute to the fields themselves, and thus, we impose a cut-off in the domain of integration:

$$[\mathcal{D}\phi]_\Lambda = \prod_{k<\Lambda} d\phi(k),$$

with $k = |\mathbf{k}|$, the Euclidean momentum. It is important to note that the cut-off is imposed on the Euclidean momentum, rather than the relativistic four-momentum in Minkowski space used in relativistic QFT. This is because the four-momentum satisfies $k^2 = E^2 - |\mathbf{k}|^2$, so the condition $k^2 < \Lambda^2$ would not effectively constrain the energy in light-like directions. The formal way to go from Minkowski to Euclidean space is through a Wick rotation: $t \mapsto -ix_0$, introducing a new spatial coordinate and changing the exponent of the generating functional: $i \int dt d^3x \mathcal{L} \to - \int d^4x \mathcal{L}$, now integrating over four spatial dimensions. This correspondence between relativistic QFTs of d spatial and one temporal dimension to Euclidean QFTs of $(d+1)$ spatial dimensions is indeed an important observation. For this reason, the following analysis will take place in arbitrary d dimensions.

An RG transformation as described above taking $a \mapsto a' = la$ would similarly take

$$\Lambda \mapsto \Lambda' \sim \frac{1}{a'} = \frac{1}{la} \sim l^{-1}\Lambda,$$

with $l > 1$, and is thus a transformation to a lower highest energy scale. Performing this operation in the functional formalism would then amount to integrating out the Fourier modes in the region $l^{-1}\Lambda < k < \Lambda$, and study the effect this has on the parameters $\{g\}$ of $\mathcal{L}[\phi, \{g\}]$. [8]

Formally, this is done by grouping the integration variables $\phi(k)$ into two groups, the low-momentum variables:

$$\tilde{\phi}(k) = \begin{cases} 0 & \Lambda/l < k < \Lambda \\ \phi(k) & k < \Lambda/l \end{cases},$$

which we want to keep, and the high-momentum variables

$$\hat{\phi}(k) = \begin{cases} \phi(k) & \Lambda/l < k < \Lambda \\ 0 & k < \Lambda/l \end{cases},$$

which are to be integrated out. The measure then splits as $[\mathcal{D}\phi]_\Lambda = \mathcal{D}\tilde{\phi}\mathcal{D}\hat{\phi}$, and the resulting generating functional (with $J = 0$ for simplicity) takes the form:

$$\begin{aligned} Z &= \int \mathcal{D}\tilde{\phi}\mathcal{D}\hat{\phi} \exp\left(-\int d^d x \mathcal{L}[\tilde{\phi} + \hat{\phi}, \{g\}]\right) \\ &\vdots \\ &= \int [\mathcal{D}\phi]_{b\Lambda} \exp\left(-\int d^d x \mathcal{L}_{\text{eff}}[\phi', \{g'\}]\right), \end{aligned}$$

where the dots represent a lengthy calculation, the details of which, in the case of a scalar field theory, can be found in section 12.1 of [4]. The main point is that it is possible to integrate out the high-momenta degrees of freedom in a QFT and rewrite the generating functional in a form similar to the original one. The Lagrangian is then replaced with an *effective* Lagrangian, $\mathcal{L}_{\text{eff}}$, with rescaled parameters and field strength. The field strength scales similarly to the parameters of the theory in subsection 2.6.2: close to a fixed point

$$\phi_i' = l^{\Delta_i}\phi_i,$$

where Δ_i is called the scaling dimension of the field ϕ_i. As an example, the ϕ^4-theory with Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) + \frac{1}{2}m^2\phi^2 + \frac{1}{4!}\lambda\phi^4, \tag{2.8.1}$$

[8] $\mathcal{L}$ is now explicitly written as a function of the parameters $\{g\}$ of the theory, e.g. masses, coupling constants, etc., to emphasize that these are indeed not constant.

close to the trivial fixed point $m = \lambda = 0$, scales as[9] (see section 12.1 of [4])

$$\phi' = l^{d/2-1}\phi, \quad m'^2 = l^2 m^2, \quad \lambda' = l^{4-d}\lambda.$$

An RG transformation can thus also be viewed as a change in energy scale, and an obvious scaling parameter in an experiment is then the temperature.

2.8.2 Parameter Classification and Mass Dimension

In units $\hbar = c = 1$, mass is the unit of energy, and hence the inverse unit of length. In the field theoretic formulation outlined above, the action must be dimensionless. As the measure, $d^d x$, has units $[\text{mass}]^{-d}$, the Lagrangian then must have units $[\text{mass}]^d$. We say that the Lagrangian has mass dimension d. We will denote this as $[\mathcal{L}] = d$.

This, in turn, implies that every term of the Lagrangian must have mass dimension d, and a mass dimension may be assigned to every field appearing in the theory. From the specific way terms are built from fields, one may find the mass dimension of the parameters of the theory. As an example, take QED:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - ie\overline{\psi}\gamma^\mu \psi A_\mu, \tag{2.8.2}$$

where γ^μ are unitless matrices, ψ is a spinor, $\overline{\psi} = \psi^\dagger \gamma^0$, A^μ the electromagnetic vector potential, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. One sees immediately that $[\psi] = (d-1)/2$ from the first term. The second term yields the mass dimension of A_μ: $[A_\mu] = d/2 - 1$. Finally, we get the mass dimension of the coupling constant, $[e] = 2 - d/2$ from the last term.

Now, qualitatively, in a renormalization procedure the parameters always flow in the direction $\Lambda \to 0$. As this limit is approached, parameters proportional to the energy, i.e. with positive mass dimension, should approach fixed value. Likewise, parameters with zero mass dimension should be left unchanged, and parameters with negative mass dimension should diverge.

These three possible behaviours of parameters looks familiar, and indeed, these three classes of possibilities are actually just a new way of identifying irrelevant, marginal and relevant parameters, respectively. However, note that a relevant parameter does not truly need to diverge, just flow to a different fixed point than the one the system currently is in the vicinity of. Also, marginal parameters

[9] In the process of renormalization, higher-order terms will appear in $\mathcal{L}$, such as $C(\partial_\mu \phi)^4$ and $D\phi^6$. They scale as $C' = Cl^{-d}$ and $D' = Dl^{6-2d}$, so, as $l > 1$, the C-term will die off quickly. The D-term, however, will grow in $d = 2$. This justifies the usage of the ϕ^6-potential for theoretical calculations of critical exponents in subsection 2.5.1.

may well change in a group flow, but one needs to include higher-order terms in the expansion, eq. (2.6.6), to find their behavior. For instance, in $d = 4$ (three spatial, one temporal dimension), it is a well known fact that the electron charge changes with energy scale, even though it has mass dimension 0.

The argument here is purely qualitative, but this classification does indeed work in practice. See [4], specifically sections 10.1 and 12.1, for the formalities regarding the analysis.

As another example, we get the correct behavior for the ϕ^4 theory (eq. 2.8.1): naturally, $[m^2] = 2$, so from the second term, $[\phi] = d/2 - 1$, which again yields $[\lambda] = 4 - d$ from the last term. Exactly the scaling exponents quoted in the previous subsection. The same analysis also applies to the $(\partial_\mu \phi)^4$-and ϕ^6-terms of footnote 9.

The classification of parameters in terms of mass dimension is particularly powerful in constructing possible Lagrangians for a system. If we want to construct an *effective* theory, where renormalization alone can not drastically change it and the system is in the vicinity of a fixed point, only terms with marginal or irrelevant parameters are allowed. Depending on which fields the theory consists of, this puts enormous constraints on the possible terms allowed in the Lagrangian.

This will be made use of later, in constructing an effective field theory for the quantum Hall effect.

2.9 Conformal Field Theory

Exactly at a critical point, a theory is invariant under a change of scale. If the theory also is invariant under the larger group of conformal transformations it is termed a *conformal field theory* (CFT). CFTs are particularly interesting in 2-dimensional critical phenomena, where they provide a classification scheme for the universality classes. CFT is a well studied and rich topic, and only a fragment of its lore is presented here for the purposes of completing the understanding of Nature's segregation of 2d critical phenomena into universality classes and certain implications on the 2d RG flow. The discussion here follows that of [26], specifically the entries of P. Ginsparg and J. L. Cardy. Certain points are also borrowed from [27]. Some readers may benefit from consulting Appendix A, on the group theory, and Appendix B, on vectors, one-forms and tensors before continuing with this section.

2.9.1 First Definitions and the Conformal Group

Firstly, define the conformal group as the group of coordinate transformations leaving the metric tensor (see Appendix B) locally invariant up to a change of scale:

$$G_{ij}(x) \mapsto G'_{ij}(x') = \frac{\partial x^k}{\partial x'^i}\frac{\partial x^l}{\partial x'^j}G_{kl}(x) \stackrel{!}{=} \Omega(x)G_{ij}(x), \tag{2.9.1}$$

where Ω is strictly positive. This definition has two immediate important consequences. First, the conformal transformations are the transformations that preserves angles between curves, that is

$$\cos\theta = \frac{G_{ij}V^iU^j}{\sqrt{(G_{mn}V^mV^n)(G_{lk}U^lU^k)}},$$

is conserved for any two vectors V and U. Secondly, the entire Poincaré group, consisting of the Lorentz transformations and translations, is contained in the conformal group.

To find the complete set of conformal transformations, study the effect of an infinitesimal transformation $x^i \to x^i + \epsilon^i$ on the line element:

$$\begin{aligned} ds^2 = G^{ij}dx_idx_j \to \frac{\partial(x^i+\epsilon^i)}{\partial x^k}\frac{\partial(x^j+\epsilon^i)}{\partial x^l}G^{kl}d(x_i+\epsilon_i)d(x_j+\epsilon_j) \\ = G^{ij}dx_idx_j + (\partial^i\epsilon^j + \partial^j\epsilon^i)dx_idx_j. \end{aligned}$$

To satisfy eq. (2.9.1), we thus need $\partial_i\epsilon_j + \partial_j\epsilon_i = a(x)G_{ij}$, for some positive $a(x)$. Tracing both sides with G^{ij}, we find

$$a(x) = \frac{2}{\mathrm{tr}(G)}\partial_i\epsilon^i,$$

with $\mathrm{tr}(G) = G^i_i = G_{ij}G^{ij}$. The general constraint on an infinitesimal conformal transformation is hence

$$\partial_i\epsilon_j + \partial_j\epsilon_i = \frac{2}{\mathrm{tr}(G)}(\partial\cdot\epsilon)G_{ij}. \tag{2.9.2}$$

Now specifying to 2-dimensional Euclidean space ($G_{ij} = \delta_{ij}$ and $x = (x^1, x^2)$), eq. (2.9.2) yields two copies of the *Cauchy-Riemann equations*:

$$\partial_1\epsilon_1 = \partial_2\epsilon_2, \qquad \partial_1\epsilon_2 = -\partial_2\epsilon_1.$$

Hence, the transformation ϵ is holomorphic[10] in the complex coordinate $z = x^1 + ix^2$, and we may write it as

$$\epsilon(z) = \epsilon^1(x, y) + i\epsilon^2(x, y).$$

The finite transformations are then the holomorphic transformations:

$$z \to f(z),$$

with $f(z)$ a holomorphic function on $\mathbb{C}$. As there are infinitely many linearly independent holomorphic functions, there is *an infinite number of constraints on 2-dimensional conformal field theories*, and therein lies their power.

Note that only the fractional linear transformations (FLTs)

$$z \mapsto \frac{az + b}{cz + d}, \tag{2.9.3}$$

with $a, b, c, d \in \mathbb{C}$ have inverses on the entire Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. So the (global) conformal group in two dimensions is isomorphic to

$$PSL(2, \mathbb{C}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; a, b, c, d \in \mathbb{C},\ ac - bd = 1 \right\} / \mathbb{Z}_2, \tag{2.9.4}$$

each matrix representing a transformation of the form (2.9.3). The constraint on the determinant $ac - bd = 1$ is enforced so as not to overcount each FLT. The $\mathbb{Z}_2$ action identifies matrices that are equal up to an overall sign change, as they would give rise to the same FLT. Lastly, it is easy to see that function composition of two FLTs, $f(z) = (a_1 z + b_1)/(c_1 z + d_1)$ and $g(z) = (a_2 z + b_2)/(c_2 z + d_2)$:

$$f(g(z)) = \frac{(a_1 a_2 + b_1 c_2)z + (a_1 b_2 + b_1 d_2)}{(c_1 a_2 + c_2 d_1)z + (c_1 b_2 + d_1 d_2)},$$

yields the same coefficients as matrix multiplication of the corresponding matrices would.

It is understood that under a transformation of the form (2.9.3), maps $-d/c \mapsto \infty$ and $\infty \mapsto a/c$ if $c \neq 0$, and $\infty \mapsto \infty$ if $c = 0$.

The fundamental conformal transformations are

[10]Recall that "holomorphic" means "complex analytic", or "complex differentiable". Complex analyticity of a function f is a *very* stringent condition as it demands that the function may be written as $f(z)$ with no dependence on the complex conjugate $\bar{z}$, and that the limit $\lim_{h\to 0}[f(z+h) - f(z)]/h$ must exist and be equal for $h \in \mathbb{C}$ approaching zero along any curve in $\mathbb{C}$. However, the ability to write the function as $f(z) = f^1(x, y) + if^2(x, y)$ with f^1 and f^2 obeying the Cauchy-Riemann equations is a sufficient condition. See [28] for a brilliant introduction to complex analysis.

- Translations: $z \to z + w$, with matrix $\begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix}$.
- Rotations: $z \to e^{i\theta} z$, $\lambda \in \mathbb{R}$, with matrix $\begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}$.
- Dilatations: $z \to \lambda^2 z$, $\lambda \in \mathbb{R}$, with matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$.
- Special conformal: $z \to \frac{z}{cz+1}$, with matrix $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$.

The latter may also be viewed as the composition of inversion ($z \to -1/z$), followed by a translation and yet another inversion.

In summation, the conformal *algebra* in two dimensions is infinite-dimensional (spanned by the z^n for $n \in \mathbb{N}$), whereas the conformal group in two dimensions is $PSL(2, \mathbb{C})$. As such one must distinguish between local (infinitesimal) conformal transformations and global ones.

Following Ginsparg [26], define a 2d conformal field theory to satisfy the following:

1. The theory consists of a set of fields $\{A_i\}$, in general infinitely large and containing all derivatives of the fields.
2. There is a subset of fields $\{\phi_j\} \subset \{A_i\}$ called *quasi-primary*, transforming as
$$\phi_j(z) \to \left| \frac{df(z)}{dz} \right|^{\Delta_j/2} \phi_j(f(z)),$$
with Δ_j the scaling dimension of ϕ_j, under the conformal transformation $z \to f(z)$.
3. The rest of the fields $\{A_i\}$ may be described as linear combinations of $\{\phi_j\}$ and their derivatives.
4. The vacuum $|0\rangle$ is invariant under the global conformal group.

One note before continuing; one usually takes z and $\bar{z}$ to be independent variables in CFT calculations, and impose that they are complex conjugates only at the end of a calculation. This formalism yields two copies of the conformal algebra: one holomorphic and one anti-holomorphic (holomorphic in $\bar{z}$), and all fields are written as explicitly dependent on z and $\bar{z}$. This is done in the following.

2.9.2 Radial Quantization

Following Ginsparg in [26], we need to introduce the quantization procedure for conformal QFTs. One of the strengths of a conformally invariant theory, is that

it allows us to change coordinates to any locally equivalent coordinate system, which may simplify a problem greatly. The process of *radial quantization* is one such application. First, consider the coordinates $\zeta = \sigma^0 + i\sigma^1$, letting σ^0 denote the "time" and σ^1 the "space" dimension. Compactify the spatial coordinate σ^1 to eliminate infrared divergences: $\sigma^1 \equiv \sigma^1 + 2\pi$, thus defining a cylinder in the σ^0, σ^1 coordinates. These coordinates may be mapped back to the complex plane via the conformal map

$$\zeta \to z = e^{\zeta} = e^{\sigma^0 + i\sigma^1}. \tag{2.9.5}$$

Now, the infinite past and future $\sigma^0 = \pm\infty$ are mapped to $z = 0, \infty$, respectively, and "equal time surfaces", $\sigma^0 =$ constant, are circles centered at the origin of $\mathbb{C}$. To define a QFT on $\mathbb{C}$, we need to find the operators that implement the conformal mappings. This is the procedure of radial quantization. As an example, the Hamiltonian can be viewed as the dilatation generator, because a dilatation on $\mathbb{C}$, $z \to e^a z$ is equivalent to a time translation $\sigma^0 \to \sigma^0 + a$ by eq. (2.9.5).

2.9.3 The Stress-Energy Tensor

Generally in field theory, exact symmetries are equivalent to conserved currents and charges by Noether's theorem. Conformal transformations corresponds to a rescaling of the metric tensor, the stress-energy (or momentum-energy) tensor is the response of the action to a variation of the metric (see Appendix B), and a symmetry should leave the action invariant. As such, we should study the stress-energy tensor.

For an infinitesimal conformal transformation (in the real coordinates) $x^i \to x^i + \epsilon^i$, we have

$$\delta S \sim \int d^2x T_{ij} \delta G^{ij} = \int d^2x T_{ij} (\delta \cdot \epsilon) \delta^{ij} = \int d^2x T^i_i (\delta \cdot \epsilon),$$

using the result from the previous subsection. Demanding a stationary action then amounts to demanding $T^i_i = 0$, that is, T_{ij} is traceless. As always, the stress-energy tensor is also symmetric: $T_{ij} = T_{ji}$, and divergence-free: $\partial^i T_{ij} = 0$. These relations all hold in the real coordinates.

Now, in the complex frame $(z, \bar{z})$, we see from the line element $ds^2 = dx^2 + dy^2 = dz d\bar{z}$, that the metric tensor takes the form $G_{zz} = G_{\bar{z}\bar{z}} = 0$ and $G_{z\bar{z}} = G_{\bar{z}z} = 1/2$. The components of the stress-energy tensor in this frame is found by the transformation properties for tensors under coordinate change from Appendix B

and employing the complex differentiation operators[11]

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \frac{1}{i}\frac{\partial}{\partial y}\right), \qquad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \frac{1}{i}\frac{\partial}{\partial y}\right).$$

This gives $T_{zz} = \frac{1}{4}(T_{00} - 2iT_{01} - T_{11})$, $T_{\bar{z}\bar{z}} = \frac{1}{4}(T_{00} + 2iT_{00} - T_{11})$, and $T_{z\bar{z}} = T_{\bar{z}z} = \frac{1}{4}(T_{00} + T_{11}) = 0$. Invoking the conservation of the stress-energy tensor yields

$$0 = \bar{\partial} T_{zz} + \partial T_{\bar{z}z} = \bar{\partial} T_{zz},$$
$$0 = \partial T_{\bar{z}\bar{z}} + \partial T_{z\bar{z}} = \partial T_{\bar{z}\bar{z}},$$

because $T_{\bar{z}z} = T_{z\bar{z}} = 0$. Finally, we see that the non-vanishing components of the stress-energy tensor may be written as

$$T_{zz} \equiv T(z), \qquad T_{\bar{z}\bar{z}} \equiv \overline{T}(\bar{z}),$$

where ($\overline{T}(\bar{z})$) $T(z)$ is (anti-)holomorphic.

A current may be defined in real coordinates by [27]

$$J_i = T_{ij}\epsilon^j.$$

This current is conserved:

$$\partial^i J_i = (\partial^i T_{ij})\epsilon^j + T_{ij}\partial^i \epsilon^j = 0,$$

as $\partial^i T_{ij} = 0$, and $T_{ij}\partial^i \epsilon^j = \frac{1}{2} T_{ij}(\partial^i \epsilon^j + \partial^j \epsilon^i) = \frac{1}{2}(\partial \cdot \epsilon) T_i^i = 0$ where the first equality follows from symmetry of T_{ij}. In the complex frame, this amounts to

$$J_z = T(z)\epsilon(z), \qquad J_{\bar{z}} = \overline{T}(\bar{z})\bar{\epsilon}(\bar{z}).$$

The conserved charge is found by integrating along a constant-time circle in $\mathbb{C}$:

$$Q = \frac{1}{2\pi i}\oint \left(dz T(z)\epsilon(z) + d\bar{z}\overline{T}(\bar{z})\bar{\epsilon}(\bar{z})\right). \tag{2.9.6}$$

According to Ginsparg [26], the infinitesimal symmetry variation (with the infinitesimal parameter ϵ) of a field A, may be found from the equal time commutator as $\delta_\epsilon A = \epsilon[Q, A]$, however, the commutator needs to be defined. This is naturally done in terms of time-ordering, which in $\mathbb{C}$ now amounts to *radial* ordering. Define the radial ordering operator by its action on two operators A and B as

$$R\left(A(z)B(w)\right) = \begin{cases} A(z)B(w) & |z| > |w| \\ B(w)A(z) & |z| < |w| \end{cases},$$

completely analogous to the usual time ordering operator of QFT. The commutator now, must always be calculated in terms of the radially ordered product. What this means is exemplified below.

[11] See e.g. [28]. Another sufficient condition for holomorphicity of a function f is $\partial f/\partial \bar{z} = 0$, as this only holds if the Cauchy-Riemann equations are satisfied.

2.9.4 Primary Fields and the Central Charge

A consequence of the definition of a conformal field theory is that the correlation functions of quasi-primary fields determine the rest of the correlation functions in the theory. However, we may go one step further and define a *primary field* Φ, of weight $(h, \bar{h})$, to be a field transforming as

$$\Phi(z,\bar{z}) \to \left(\frac{df}{dz}\right)^{h} \left(\frac{d\bar{f}}{d\bar{z}}\right)^{\bar{h}} \Phi(f(z),\bar{f}(\bar{z})),$$

where $h, \bar{h} \in \mathbb{R}$ are not complex conjugates. A primary field is automatically quasi-primary, and it is possible to show that $h + \bar{h} = \Delta$ is the scaling dimension of the field. Under an infinitesimal transformation $z \to z + \epsilon(z)$, $\bar{z} \to \bar{z} + \bar{\epsilon}(\bar{z})$, let $\Phi(z,\bar{z}) \to \Phi(z,\bar{z}) + \delta_{\epsilon,\bar{\epsilon}}\Phi(z,\bar{z})$. We then get

$$\delta_{\epsilon,\bar{\epsilon}}\Phi = \left(h\partial\epsilon + \epsilon\partial + \bar{h}\bar{\partial}\bar{\epsilon} + \bar{\epsilon}\bar{\partial}\right)\Phi, \tag{2.9.7}$$

where now $\partial = \frac{\partial}{\partial z}$ and $\bar{\partial} = \frac{\partial}{\partial \bar{z}}$.

However, from the previous subsection, the transformation $z \to z + \epsilon(z)$ for a primary field Φ yields

$$\begin{aligned}
\delta_\epsilon\Phi(w,\overline{w}) &= [Q, \Phi(w,\overline{w})] \\
&= \frac{1}{2\pi i}\oint [dzT(z)\epsilon(z), \Phi(w,\overline{w})] \\
&= \frac{1}{2\pi i}\oint dz\epsilon(z)\Big(T(z)\Phi(w,\overline{w}) - \Phi(w,\overline{w})T(z)\Big) \\
&= \frac{1}{2\pi i}\left(\oint_{|z|>|w|} - \oint_{|z|<|w|}\right) dz\epsilon(z)R(T(z)\Phi(w,\overline{w})) \\
&= \frac{1}{2\pi i}\oint_{C[w]} dz\epsilon(z)R(T(z)\Phi(w,\overline{w})),
\end{aligned}$$

where $C[w]$ is some curve around w not containing the origin. We also get an analagous term with the barred quantites. Demanding equality with the transformation properties for primary fields already defined in eq. (2.9.7) amounts to demanding that *the stress-energy tensor generates the conformal transformations*. This is indeed allowed, and we find the following short-distance behavior of the product operator expansion of T with any primary field Φ:[12]

$$R(T(z)\Phi(w,\overline{w})) = \frac{h}{(z-w)^2}\Phi(w,\overline{w}) + \frac{1}{(z-w)}\partial_w\Phi(w,\overline{w}) + ... \tag{2.9.8}$$

[12]Recall the residue formula for complex integration: $\oint_C \frac{dz}{2\pi i} f(z) = \sum_k \mathrm{res}_{z_k} f$, where the residues are given in terms of the singularities of f contained in the contour C. For a pole of order n at z_k: $\mathrm{res}_{z_k} f = \lim_{z\to z_k} \frac{1}{(n-1)!}\left(\frac{d}{dz}\right)^{n-1}(z-z_k)^n f(z)$. See e.g. [28] for further details.

where only the singular terms are displayed. The term involving $\overline{T}(\bar{z})$ is completely analagous. These short-distance properties may, in fact, be taken to *define* the stress-energy tensor for a conformal QFT [26].

The stress-energy tensor itself is not a primary field. It can be shown that its operator product expansion with itself yields

$$T(z)T(w) = \frac{c/2}{(z-w)^4} + \frac{2}{(z-w)^2}T(w) + \frac{1}{z-w}\partial_w T(w) + ...$$

only displaying the singular terms. The constant c is called the *central charge* and cannot be solely determined by requiring T to generate conformal transformations. It will in general depend on the specific CFT under consideration. In fact, it will be used to *classify* the possible allowed CFTs in two dimensions.

2.9.5 The Virasoro Algebra

One last bit of conformal formality must be introduced before we get to the magic. Define the Laurent expansion[13] of the stress-energy tensor:

$$T(z) = \sum_{n\in\mathbb{Z}} z^{-n-2}L_n, \qquad \overline{T}(\bar{z}) = \sum_{n\in\mathbb{Z}} \bar{z}^{-n-2}\overline{L}_n,$$

where the coefficients are found by

$$L_n = \frac{1}{2\pi i}\oint dz z^{n+1}T(z), \qquad \overline{L}_n = \frac{1}{2\pi i}\oint d\bar{z}\bar{z}^{n+1}\overline{T}(\bar{z}).$$

The L_n and $\overline{L}_n$ are operators and their commutation relations defines an algebra. The commutator is calculated as before:

$$\begin{aligned}
[L_m, L_n] &= \left(\oint \frac{dz}{2\pi i}\oint \frac{dw}{2\pi i} - \oint \frac{dw}{2\pi i}\oint \frac{dz}{2\pi i}\right) z^{n+1}w^{n+1}R(T(z)T(w)) \\
&= \oint \frac{dz}{2\pi i}\oint \frac{dw}{2\pi i} z^{n+1}w^{n+1}\left(\frac{c/2}{(z-w)^4} + \frac{2T(w)}{(z-w)^2} + \frac{\partial_w T(w)}{z-w} + ...\right) \\
&= (n-m)L_{m+n} + \frac{c}{12}(n^3-n)\delta_{m+n,0}\ ,
\end{aligned}$$

where the last line follows from application of the residue theorem and an integration by parts of the $\partial_w T$-term.

The L_n's spans an infinite-dimensional algebra, called the Virasoro algebra, and these may be viewed as the generators of the conformal symmetries. Observe that the commutation relation determining the algebra is uniquely defined in terms of the central charge.

[13] The generalization of the Taylor expansion to a series expansion that is absolutely convergent in an annulus in the complex plane where the function is holomorphic. See e.g. [29].

2.9.6 The Conformal Ward Identity

In general, Ward identities are constraints on correlation functions reflecting some symmetry of the system. See e.g. section 9.6 of [4] for a general discussion of Ward identities originating from Noether currents in QFT. An important result of CFT is the conformal Ward identity, originating from the conformal symmetry, satisfied by the primary fields.

The starting point is looking at an n-point correlation function of primary fields under the infinitesimal transformation $z_i \to z_i + \epsilon(z_i)$. In the following, we suppress the field's dependence of the barred variables ($\bar{z}_i$), as we consider only the real transformation. The variation is calculated as in the previous subsection:

$$\begin{aligned}
\delta_{n-\text{point}} &= \left\langle \oint_C \frac{dz}{2\pi i}\epsilon(z) R\left(T(z)\Phi_1(z_1)...\Phi_n(z_n)\right)\right\rangle \\
&= \sum_{i=1}^{n}\left\langle \Phi_1(z_1)...\left(\oint_{C[z_i]}\frac{dz}{2\pi i}\epsilon(z)R(T(z)\Phi_i(z_i))\right)...\Phi_n(z_n)\right\rangle \\
&= \sum_{i=1}^{n}\left\langle \Phi_1(z_1)...\left(\oint_{C[z_i]}\frac{dz}{2\pi i}\epsilon(z)\left[\frac{h}{(z-z_i)^2}+\frac{1}{z-z_i}\partial_{z_i}\right]\Phi_i(z_i)\right)...\Phi_n(z_n)\right\rangle,
\end{aligned}$$

where C is a contour containing all z_i and $C[z_i]$ only contains the specific z_i. This validity of splitting contour integrals this way is an immediate consequence of the residue theorem. In the third line, the expansion in eq. (2.9.8) has been written out. As the relation above is valid for arbitrary ϵ, there is an equality of integrands and we get:

$$\langle T(z)\Phi_1(z_1)...\Phi_n(z_n)\rangle = \sum_{i=1}^{n}\left(\frac{h_i}{(z-z_i)^2}+\frac{1}{z-z_i}\partial_{z_i}\right)\langle\Phi_1(z_1)...\Phi_n(z_n)\rangle, \quad (2.9.9)$$

the conformal Ward identity.

2.9.7 Descendant Fields

The strength of the primary fields defined above is that they are sufficient to represent all fields of the CFT. Generally, one may partition the fields into families denoted $[\Phi_n]$, where each family contains a single primary field (Φ_n), and an infinite number of non-primary fields called descendant fields. These are the irreducible representations of the conformal group, and between them, they contain all the fields of the CFT: $\{A_i\} = \cup_n[\Phi_n]$.

The descendant fields for a primary field Φ are extracted from the product operator expansion of the stress-energy tensor with Φ:

$$\begin{aligned} R(T(z)\Phi(w)) &= \sum_{n\geq 0}(z-w)^{n-2}L_{-n}\Phi(w) \\ &= \frac{1}{(z-w)^2}L_0\Phi + \frac{1}{z-w}L_{-1}\Phi + L_{-2}\Phi + (z-w)L_{-3}\Phi + ... \end{aligned}$$

as the Laurent expansion coefficients:

$$\Phi^{-n}(w) := L_{-n}\Phi(w) = \oint \frac{dz}{2\pi i}\frac{1}{(z-w)^{n-1}}T(z)\Phi(w). \tag{2.9.10}$$

We are now ready for the first piece of magic: correlation functions involving descendant fields may be written entirely in terms of correlation functions of primary fields. To illustrate this, consider the correlation function of $n-1$ primary fields and one descendant field Φ^{-m}. Following [27]:

$$\begin{aligned} &\langle\Phi_1(z_1)...\Phi_{n-1}(z_{n-1})\Phi^{-m}(z)\rangle \\ &= \oint \frac{dw}{2\pi i}\frac{1}{(w-z)^{m-1}}\langle\Phi_1(z_1)...\Phi_{n-1}(z_n)T(w)\Phi(z)\rangle \\ &= \oint \frac{dw}{2\pi i}\frac{1}{(w-z)^{m-1}}\sum_{i=1}^{n-1}\left(\frac{h_j}{(w-z_i)^2} + \frac{1}{w-z_i}\partial_{z_i}\right)\langle\Phi_1(z_1)...\Phi_{n-1}(z_{n-1})\Phi(z)\rangle \\ &= \mathcal{L}_{-m}\langle\Phi_1(z_1)...\Phi_{n-1}(z_{n-1})\Phi(z)\rangle, \end{aligned}$$

using the definition of the descendant fields and the conformal Ward identity (eq. 2.9.9). The differential operator $\mathcal{L}_{-m}$ is defined as

$$\mathcal{L}_{-m} = -\sum_{i=1}^{n-1}\left(\frac{(1-m)h_i}{(z_i-z)^m} + \frac{1}{(z_i-z)^{m-1}}\partial_{z_i}\right).$$

It is in general possible to write the correlation functions of arbitrary non-primary fields in terms of those of primary fields [26].

This is an amazing fact of 2-dimensional CFTs: all the dynamics of the theory is determined solely by the primary fields and the stress-energy tensor.

2.9.8 Universality Classes

As previously stated, 2d critical phenomena are scale invariant. In fact, at a critical point, 2-dimensional systems can usually be modelled by CFTs with central charge $c < 1$.[14]

[14]Nobody seems to *prove* this, though several authors states that it is usually the case.[3], [26] As c may be found by trial and error for any given theory [3], it may well be that all fathomable 2d statistical mechanics models with critical behaviour have been categorized.

Another amazing fact about 2d CFT is that it is possible to fully classify all possible theories with $c < 1$. This is done by demanding unitary representations of the Virasoro algebra. The only values for c where unitarity is not excluded are given by the formula:

$$c = 1 - \frac{6}{m(m+1)}, \qquad m = 3, 4, ... \tag{2.9.11}$$

Furthermore, for each such c the allowed values for the conformal weight is

$$h_{p,q} = \frac{[(m+1)p - mq]^2 - 1}{4m(m+1)}, \tag{2.9.12}$$

with $p,\ q$ integers constrained by $p \in [1, m-1],\ q \in [1, p]$. For the details of this derivation see Ginsparg in [26] and the references therein.

The powerful implication of formulas (2.9.11) and (2.9.12) is that for each allowed CFT there is a *finite* amount of primary fields, which again dictate the dynamics of the theory. Now, it is possible to tie specific statistical mechanics models to CFTs specified solely by a value of c. In Cardy [26], this is done for the Ising and 3-state Potts models by studying the simplest possible CFTs, $c = 1/2$ and $c = 4/5$, and demanding consistency with their order-parameter symmetries, $\mathbb{Z}_2$ and $\mathbb{Z}_3$, respectively.[15]

There is thus a correspondence between the symmetry group of the order parameter of a system, and the central charge of the CFT describing the system at a critical point. This is a beautiful result of the CFT formalism in two dimensions: *each universality class is given almost uniquely by the value of the central charge.*[16] We may conclude that the unversality class of a system is determined, first by its dimensionality, and if it is 2-dimensional, its central charge. The problem of determining critical exponents from a given CFT is adressed in the original seminal paper of Belavin, Polyakov and Zamolodchikov in 1984 [30].

2.9.9 Zamolodchikov's C-Theorem

The final important result of CFT we quote here is a deep theorem regarding the RG flow of 2-dimensional systems. We cite it in Cardy's [26] wording:

There exists a function C of the coupling constants which is non-increasing along renormalization group trajectories, is stationary only at fixed points, and at a fixed point, is equal to the value of the central charge for the corresponding theory.

[15] The $c = 7/10$ model is actually in between the two of them. This corresponds to the *tricritical* Ising model, that is, an Ising model where three phases coexist at the critical point. Likewise, the next allowed central charge after $4/5$ is $6/7$, corresponding to the tricritical 3-state Potts model.

[16] The precise statement is the ADE classification of conformal field theories.[26]

The reader is referred to the original paper by Zamolodchikov [31] for the details of the proof.

An immediately interesting feature of the theorem is the statement that C is non-decreasing along RG trajectories and takes the value of the central charge, c, at critical points. This means that if the system is perturbed away from a critical point and sets out to float towards a different fixed point, the resulting CFT is either the same one or *simpler*. "Simpler" here means that fewer primary fields will be allowed by a smaller value of c, thus limiting the complexity of the dynamics of the theory. This fits in beautifully with our understanding of the action of an RG transformation as removing degrees of freedom - such an operation *should* lead to a less complicated theory.

Zamolodchikov's c-theorem also has important implications on the β-functions of 2-dimensional systems, which is the subject of the next section.

We round off this section by mentioning that an analogue of the C-theorem, called the A-theorem, has been conjectured for 4-dimensional systems [32]. The existence of a 4-dimensional RG potential could have very interesting consequences, however, that subject will not be pursued here. The interested reader is referred to the review article [33] for the current status of the A-theorem (2015).

2.10 The β-functions and Duality

The idea that the parameters of a theory will change with a scale transformation should now be properly implanted, however, we are interested in knowing exactly how the parameters of a Lagrangian flows in parameter space. This is determined by the β-functions of the theory.

The β-function for a parameter g^i is defined as[17]

$$\beta^i(\{g\}) = \frac{\partial g^i}{\partial t}, \tag{2.10.1}$$

with $t = \ln(\mu/\mu_0)$, where μ is the energy scale and μ_0 a reference scale where the parameters of the theory have been determined experimentally. In QFT's they may be found through perturbation theory using the Callan-Symanzik equation (see e.g. section 12.2 of [4]). They may also be inferred from the shape of an experimentally generated phase diagram. This method is particularly powerful if the phase diagram exhibits a duality. In general, one may write $\beta = (\beta^1, \beta^2, ...)$

[17] From now on, upper indices will be used to denote physical β-functions for reasons which will become apparent. For consistency upper indices will also be used for the parameters g^i.

defining a vector field on parameter space. The flow lines of this vector field[18] are then the flow lines of the RG-flow as portrayed in Fig. 2.6.2.

2.10.1 Dualities of Parameter Space

A *duality symmetry* is, similar to an RG transformation, a transformation on parameter space, leaving the generating functional (partition function) and correlation functions of the theory invariant. Duality symmetries are, however, discrete and (mathematically) reversible, whereas an RG transformation is continuous and (mathematically) irreversible.[19]

Consider then, a theory in which there exists a duality symmetry, D, of the parameter g^i in parameter space:

$$D(g^i) = \tilde{g}^i(g), \tag{2.10.2}$$

g now denoting the whole set of parameters, i.e. $g = \{g\}$. Under a change of scale, g^i is subject to a flow determined by eq. (2.10.1). As the system must behave according to both of these equations, they should commute. Following [34], the effect of a duality transformation on a general function $F(g)$ on parameter space is

$$D[F(g)] = F(\tilde{g}).$$

On the same function, consider an infinitesimal RG-transformation R:

$$R[F(g)] = F(g) + \frac{\partial F}{\partial g^i}\beta^i(g)dt.$$

For consistency we must then demand

$$\begin{aligned} DR[F] &= RD[F] \\ D\left[F(g) + \frac{\partial F}{\partial g^i}\beta^i(g)dt\right] &= F(\tilde{g}) + \frac{dF(\tilde{g}(g))}{dt}dt \\ F(\tilde{g}) + \frac{\partial F}{\partial \tilde{g}^i}\beta^i(\tilde{g})dt &= F(\tilde{g}) + \frac{\partial F}{\partial \tilde{g}^i}\frac{\partial \tilde{g}^i}{\partial g^j}\frac{\partial g^j}{\partial t}dt \\ \beta^i(\tilde{g}) &= \frac{\partial \tilde{g}^i}{\partial g^j}\beta^j(g). \end{aligned} \tag{2.10.3}$$

So the β-functions must transform as contravariant vectors (see Appendix B) under the duality transformation. This is the reason for writing upper indices

[18] The lines to which the field vectors are tangent.

[19] At this point it is more than appropriate to point out that the renormalization group is not a mathematical group on the grounds that integrating out degrees of freedom is an irreversible process and hence inverses of the transformations do not exist.

on the β-functions, they are true geometric quantities and one must distinguish between contravariant (upper) and covariant (lower) indices, related by an appropriate metric G_{ij} on the parameter space

$$\beta_i = G_{ij}\beta^j. \tag{2.10.4}$$

If a duality symmetry is experimentally observed in a system, one can thereby put constraints on the possible β-functions that may describe parameter flow. Depending on the type of duality, i.e., the form of $d\tilde{g}/dg$, constraints may be very stringent, leading to β-functions uniquely determined by symmetry alone.

2.10.2 Implications of the C-Theorem

As the β-functions indeed define a vector field on the parameter space, a natural question to ask is whether the vector field is conservative. That is - may β be expressed as the gradient of a scalar field? In the case of a 2-dimensional system, that is exactly the question Zamolodchikov's C-theorem answers.

The function C decreases along RG trajectories which is exactly the direction β points. From the interpretation of the gradient this yields $\beta \propto -\nabla C$, with the gradient taken in parameter space. The exact relation is found by Zamolodchikov [31], valid in the vicinity of a fixed point:

$$\beta^i = -\frac{1}{12}G^{ij}\frac{\partial}{\partial g^j}C,$$

where G_{ij} is a metric and C is a specific combination of the components of the stress-energy tensor. As such, the C-function can truly be though of as the *renormalization group potential*.

For 3-dimensional parameter spaces, this has the interesting consequence that the β-function is curl free:

$$\nabla \times \beta = 0.$$

For a 2-dimensional parameter space, $g = (g^1, g^2)$, the relation simplifies to

$$\partial_1\beta^2 - \partial_2\beta^1 = 0. \tag{2.10.5}$$

Now, assume the β-function is defined close to a saddle point of the flow. Assume next that the C-function is symmetric with with respect to this point, in the sense that the velocity of the RG flow into the saddle point (along the irrelevant parameter) is equal to the velocity of the RG flow out of the saddle point (along the relevant parameter). This situation would mean that the curvature of the surface determined by the C-function is equal along the two principal directions

(with opposite sign). In this *very* special case, we may also take the β-function to be divergence free:

$$\nabla \cdot \beta = \partial_1 \beta^1 + \partial_2 \beta^2 = 0. \tag{2.10.6}$$

When this is the case, we define the complexified parameter $g = g^1 + ig^2$, and β-function $\beta = \beta^1 + i\beta^2$, and witness that equations (2.10.5) and (2.10.6) together imply

$$\frac{\partial}{\partial \bar{g}} \beta = \left(\frac{\partial}{\partial g^1} + \frac{1}{i} \frac{\partial}{\partial g^2} \right) \beta = \frac{\partial \beta^1}{\partial g^1} + \frac{\partial \beta^2}{\partial g^2} + i \left(\frac{\partial \beta^2}{\partial g^1} - \frac{\partial \beta^1}{\partial g^2} \right) = 0. \tag{2.10.7}$$

So β is holomorphic in $\bar{g}$, and anti-holomorphic in g. In such cases, the β-function will be taken to be a function of g and $\bar{g}$, rather than g^1 and g^2: $\beta = \beta(g, \bar{g})$, similarly to the analysis in the discussion of CFT.

Chapter 3

Quantum Hall Physics

3.1 Introduction

The quantum Hall effect (QHE) is one of the most studied condensed matter systems. The integer effect (IQHE) was discovered in 1980 by Klaus von Klitzing [35], showing that "... the Hall resistance [...] has fixed values which depend only on the fine structure constant and the speed of light, and is insensitive to the geometry of the device", giving him the 1985 Nobel Prize in Physics. The paper shows a resistance quantized as $R = R_K/n$ with n an integer, and

$$R_K = h/e^2 = 25\ 812.807\Omega, \tag{3.1.1}$$

h being the Planck constant and e the elementary charge. R_K is now named the von Klitzing constant. In 1981, Robert B. Laughlin argued that the exact quantization discovered by von Klitzing was a consequence of gauge invariance [36].

The fractional effect (FQHE) was discovered by Daniel C. Tsui and Horst L. Störmer in 1982 [37], and later explained by Laughlin as an incompressible quantum fluid with fractionally charged quasiparticles in 1983 [38]. This awarded the three of them jointly with the 1998 Nobel Prize.

More recently, the 2010 Nobel Prize was awarded to Andre Geim and Konstantin Novoselov "for groundbreaking experiments regarding the 2-dimensional material graphene". These experiments was, of course, measuring the QHE in graphene [39]. Graphene is a particularly exotic material and the topic of QHE in graphene will be revisited.

Most recently, the 2016 Nobel Prize was awarded to David J. Thouless, Duncan M. Haldane and J. Michael Kosterlitz "for theoretical discoveries of topological

phase transitions and topological phases of matter". This led to a deeper understanding of the Hall the response in crystalline materials than the previous gauge-invariance argument perpetrated by Laughlin, showing that the quantized resistivities (conductivities), follow from a topological invariant, the first *Chern number*, which must be an integer [40]. This would explain why the QHE is so robust, appearing in a host of different materials and geometries. Furthermore, in 2013, the first *topological insulator*, a topological phase of matter, was experimentally discovered as a QHE. It seems fair to associate also this Nobel Prize with the QHE in general.

So, since its first discovery in 1980 the QHE has sparked four Nobel Prizes. If that was not enough, as the resistance is measured as quantized integer steps of the Klitzing constant as accurately as about one part in a billion [41], the effect is now used to *define* the Ohm, by defining the von Klitzing constant according to eq. (3.1.1) [42]. Appropriately, the popularity of the QHE has brought about a host of books, papers, reviews, master theses and PhD dissertations; theoretical as well as experimental and numerical, regarding the QHE as a general phenomenon, or focusing on specific and exotic new materials.

It is futile to try to give a complete review of this rich topic, so here only the basics needed to understand the QHE and the work done in this thesis will be provided. A good source for further information is [43], which the discussion here follows to a certain extent.

3.2 The Classical Hall Effect

Discovered by Edwin Hall in 1879, the effect would carry his name into the quantum world a hundred years later. One should always start with a quick recapitulation of analagous classical physics when this is possible, and in this case it also helps to build some intuition and introduce notation and terminology. Furthermore, resistivities and conductivites measured in the QHE are macroscopic quantities, obeying relations imposed on them from classical physics.

3.2.1 The Experiment

Hall's original experiment was performed in a thin sheet of gold leaf, 2cm wide and 9cm long. It was placed in a strong perpendicular magnetic field with current running through in the longitudinal direction, while measuring the current in the transverse direction [44]. The result was a measured current flowing in the transverse direction. Figure 3.2.1 shows the general experimental setup.

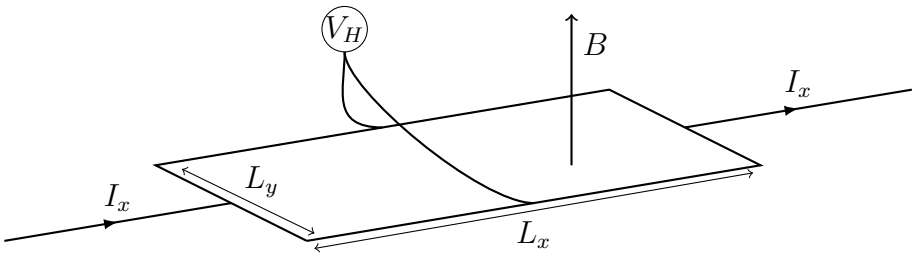

Figure 3.2.1: Experimental setup for the Hall effect with the "Hall bar" geometry. A current is applied in the longitudinal direction to a thin slab of conducting material in the presence of a perpendicular magnetic field. A voltage drop $V_H = V_y$ is induced and measured in the transverse direction.

The intuitive explanation for the measured transverse *Hall voltage*, V_H, is that magnetic fields change the direction of motion of charged particles, according to the Lorentz force law

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B},$$

q being the charge of the particle, $\mathbf{v}$ its velocity and $\mathbf{B}$ the magnetic field. This, of course leads to an acceleration:

$$\mathbf{a} = \frac{q\mathbf{v} \times \mathbf{B}}{m},$$

of constant magnitude and direction always perpendicular to the direction of motion, constituting a circular motion. If now the transverse dimension of the Hall bar is small enough, the electrons of the current will be dragged by the magnetic force towards one side of the bar and charge will accumulate until the resulting electric field, E_y, matches the magnetic force. In the case of a perpendicular magnetic field, this happens when $E_y q = q v_d B$, or

$$V_H = \frac{v_d B}{L_y}, \tag{3.2.1}$$

v_d being the electron drift velocity and L_y the transverse dimension. We used here that the electric field strength and voltage drop between two points in a material relates as $V = Ed$, with d the distance between the points. The result is Hall voltage that is linear in the applied magnetic field strength, precisely what is measured in experiment. The Hall resistance R_H, is defined as the quotient of Hall voltage and longitudinal current:

$$R_H = \frac{V_H}{I_x}. \tag{3.2.2}$$

The accurate way of analysing the situation is boosting to the reference system of the electrons in the bar with drift velocity v_d in the x-direction, subject to

an electric field in the x-direction and a magnetic field in the z-direction. The electromagnetic field tensor then takes the shape ($c = 1$) $F^{10} = -F^{01} = E$, $F^{21} = -F^{12} = B$ and all other $F^{\mu\nu}$ zero. A boost in the x-direction then yields a transverse electric field:

$$E'_y = F'^{20} = \Lambda^2_\sigma \Lambda^0_\rho F^{\sigma\rho} = -\gamma v_d B,$$

amounting to the same result as (3.2.1), given that the drift velocity is small compared to the speed of light, so $\gamma = 1/\sqrt{1 - v_d^2} \approx 1$.

For future reference, we also define the *carrier mobility* μ by the equation

$$v_d = \mu E, \tag{3.2.3}$$

as a measure of how easily an electron maneuvers its way through the Hall bar.

3.2.2 Resistivity v. Resistance

To rid ourselves of geometric considerations, resistivites and conductivities are used in place of resistances. The resistivity of a material carrying a current density J, driven by the electric field E is defined as

$$\rho := \frac{E}{J} = \frac{AV}{dI} = \frac{A}{d}R,$$

using that $J = I/A$, A being the cross-sectional area of the conductor and d its length. The latter form may be used to relate the resistance with the resistivity in a material of constant cross-sectional area. In an effectively 2-dimensional[1] material, the current density is defined as $J_x = I_x/L_y$ (compare Fig. 3.2.1) and the resulting resistivity is

$$\rho_{xx} = \frac{L_y}{L_x} R_{xx}. \tag{3.2.4}$$

When current can flow in more than one direction in a conductor, the electric field and current density are vector quantities and the resistivity must include shear terms to accurately describe their relation. ρ then becomes a matrix and is defined by the equation

$$\mathbf{E} = \rho \mathbf{J}. \tag{3.2.5}$$

[1]This is a classical effect, and by no means truly 2-dimensional. The Hall bar has a well defined thickness, but the effect is only apparent when the material is sufficiently thin. In this case, it is appropriate to mathematically restrict the current to only flowing in two dimensions. This restriction also translates well to the QHE regime.

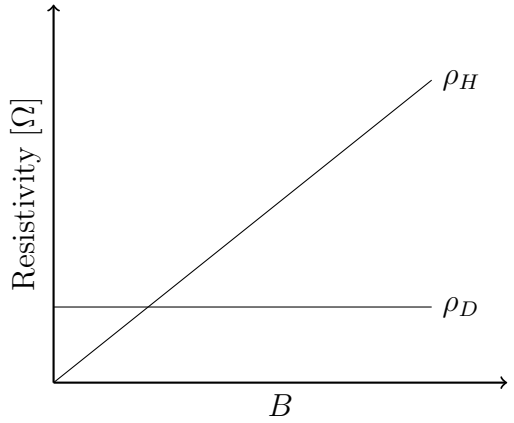

Figure 3.2.2: Illustration of measured resistivities as magnetic field strength is increased in the classical Hall effect. See the text for explanation.

The elements of the resistivity matrix in the steady state with current only flowing in *one* direction may then be defined in terms of resistances as in eq. (3.2.4):

$$\begin{aligned} \rho_{yy} &= \frac{E_y}{J_y} = \frac{L_y}{L_x} R_{yy} \\ \rho_{xy} &= \frac{E_x}{J_y} = \frac{L_y}{L_y} R_{xy} = R_{xy} \\ \rho_{yx} &= \frac{E_y}{J_x} = \frac{L_y}{L_y} R_{yx} = R_{yx} = R_H. \end{aligned} \tag{3.2.6}$$

Combining these equations with eqs. (3.2.2) and (3.2.1), one gets the situation depicted in Fig.+ 3.2.2. Note also that the Hall resistivity may be written

$$\rho_{yx} = \frac{E_y}{J_x} = \frac{v_d B}{n e v_d} = \frac{B}{ne}, \tag{3.2.7}$$

where n is the *carrier density*, i.e. the density of conducting electrons and e the elementary charge.

3.2.3 Conductivities and Transport Matrices

The measured quantities are resistivities (resistances), but mathematically, equations are usually phrased in terms of conductivities. The conductivity is the inverse of the resistivity:

$$\sigma = \rho^{-1}. \tag{3.2.8}$$

In this type of 2-dimensional transport phenomena the transport matrices take the form

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix} \qquad \rho = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{xx} \end{pmatrix}, \tag{3.2.9}$$

as is derived from the Onsager relations and symmetry considerations in Appendix C. This symmetric form makes it possible to express ρ and σ as complex numbers:

$$\sigma = \sigma_{xy} + i\sigma_{xx} \qquad \rho = \rho_{xy} + i\rho_{xx}, \tag{3.2.10}$$

where they still will be related through $\sigma = -\rho^{-1} = -1/\rho$.[2] This notation has powerful implications which will be made use of later. Notationally we will also sometimes denote the direct (longitudinal) and Hall (transverse) resistivities $\rho_{xx} = \rho_D$ and $\rho_{xy} = \rho_H$, and similarly for conductivities $\sigma_{xx} = \sigma_D$, $\sigma_{xy} = \sigma_H$. They are then related as[3]

$$\sigma_H = \frac{-\rho_H}{\rho_H^2 + \rho_D^2} \qquad \sigma_D = \frac{\rho_D}{\rho_H^2 + \rho_D^2}\ , \tag{3.2.11}$$

and

$$\rho_H = \frac{-\sigma_H}{\sigma_H^2 + \sigma_D^2} \qquad \rho_D = \frac{\sigma_D}{\sigma_H^2 + \sigma_D^2}\ . \tag{3.2.12}$$

3.3 From Classical to Quantum Hall Effects

This section will introduce further relevant terminology regarding the QHE, before concluding with Laughlin's theory for the integer effect, as it provides the most intuitive[4] way of thinking of the system.

We now move on to consider the same general experimental setup, but in a very cold environment (a few mK in many cases) and in two dimensions. In this situation, *two dimensions* is meant to be truly two dimensions in the physical sense: current flows in materials so thin that the charge carriers (electrons) physically can not move in the third dimension. This can be achieved by engineering current conducting structures of thickness less than the de Broglie wavelength of the carrier electrons.

These conditions form a system of highly correlated, strongly coupled charge carriers, termed a *2-dimensional degenerate electron gas* (2DEG).

[2] We define the conductivity and resistivity to be related through inversion with an additional minus sign, so as to keep the sign of the longitudinal resistivity and conductivity positive. This allows us restrict ourselves to examining the upper half complex plane.

[3] Often, one redefines the Hall resistivity and Hall conductivity to absorb the annoying factor of -1.

[4] Compared to Chern-Simons theory.

Examples of structures which achieve this Flatland Utopia for exotic quantum effects are

- MOSFETs (metal-oxide-semiconductor field-effect transistors): this was the system where von Klitzing originally discovered the integer QHE. [35] The *inversion layer*, the interface between the semiconductor and the insulator (oxide, usually SiO_2), is an effectively 2-dimensional area.
- Semi-conducting heterojunctions: on the interface between two dissimilar crystalline semiconductors, such as GaAs-GaAlAs and InGaAs-InP. These types of structures are collectively termed *quantum wells*.
- Topological insulators: although they are characterized as insulators for their disability to conduct currents, the *surface* of a bulk topological insulator may still allow for charge transport, creating a 2-dimensional QH environment.
- "Naturally" 2-dimensional materials: where layers a single atom thick can be manufactured. Examples are graphene and black phosphorus.

It is the fact that the QHE is observed in such a multitude of different materials, with virtually nothing in common but their dimensionality, that makes it interesting for this thesis. Some of these materials are particularly interesting and all of them will be revisited.

3.3.1 Landau Levels

From the experimental conditions outlined above, one can make some predictions of the IQHE from quantum mechanics. We will assume a system of non-interacting electrons without any sort of background potential. Consider first the Hamiltonian of an electron in a magnetic field:

$$H = \frac{1}{2m}\left(-i\hbar\nabla + e\mathbf{A}\right)^2. \tag{3.3.1}$$

Let the magnetic field point in the positive z-direction and adopt the *Landau gauge*: $\mathbf{A} = (0, Bx, 0)$ such that $\mathbf{B} = \nabla \times \mathbf{A}$. This way, the Hamiltonian respects translational symmetry in the y-direction and we may assume plane-wave behaviour in the wave functions: $\psi(x, y) = \exp[ik_y y]\phi(x)$, yielding

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{1}{2}m\omega_c\left(l_B^2 k_y + x\right)^2,$$

with $\omega_c = eB/m$, the cyclotron frequency and $l_B = \sqrt{\hbar/eB}$ is the magnetic length. This is just the Hamiltonian for a harmonic oscillator shifted along the

x-axis. Energy levels are then given by the standard

$$E_n = \hbar\omega_c \left(n + \frac{1}{2}\right),$$

resulting in a huge degeneracy of one state for each value of k_y belonging to the same energy level E_n. These energy levels are referred to as *Landau levels*, after Lev Landau who first solved the problem.

Following [43], the degeneracy can be estimated. The wave functions for the harmonic oscillator have a gaussian profile localized about $x = 0$, so with a shift we expect them to be localized at $x = -l_B^2 k_y$. Now, $k_y = 2\pi m/L_y$ with $m \in \mathbb{Z}$, and $x \in [0, L_x]$, so the number of degenerate states in each Landau level is given by

$$N_L = \frac{L_y}{2\pi} \int_{-L_x/l_B^2}^{0} dk_y = \frac{L_x L_y}{2\pi l_B^2} = L_x L_y \frac{B}{\Phi_0}, \tag{3.3.2}$$

defining the *magnetic flux quantum* $\Phi_0 = h/e$. This means there is one state for each Landau level for each flux quantum through the system. We define the Landau filling factor

$$\nu = \frac{N}{N_L}, \tag{3.3.3}$$

N being the total number of filled states.

Dropping a voltage in the x-direction amounts to to adding a term $eE_x x$, E_x being the corresponding electric field, to the Hamiltonian. We may complete the square to see that the Hamiltonian once again takes the shape of a shifted harmonic oscillator, however, also with a shift in ground state energy, yielding (see e.g. section 1.4.2 of [43]):

$$E_{n,k_y} = \hbar\omega_c \left(n + \frac{1}{2}\right) + eE \left(k_y l_B^2 - \frac{eE_x}{m\omega_c^2}\right) + \frac{1}{2} m \frac{E_x^2}{B^2}.$$

The energy now depends on k_y, lifting the previous degeneracy and introducing an electron drift in the y-direction. The drift velocity is given by the dispersion relation:

$$v_d = \frac{1}{\hbar} \frac{\partial E_{n,k_y}}{\partial k_y} = \frac{E_x}{B}.$$

The corresponding current density is $j_y = nev_d$, and at ν filled Landau levels the electron density is $n = B\nu/\Phi_0$, yielding

$$j_y = \frac{e^2}{h} \nu E_x,$$

corresponding to

$$\sigma_{xy} = \frac{e^2}{h}\nu. \tag{3.3.4}$$

This seems to be the correct result, and indeed it is. The problem, however, is that there is no prediction of plateaux invariant to variations in B, because there are partially filled Landau levels for general values of B. In general, we want eq. (3.3.4), but only valid for $\nu \in \mathbb{N}$ (IQHE) or $\nu \in \mathbb{Q}$ (FQHE).

Also, this calculation did not take *degeneracy* into account. Two types of degeneracy contributes to the QHE, both of which will be introduced now.

3.3.2 Zeeman Splitting and Spin Degeneracy

There is an additional term in the Hamiltonian arising from the electron spin coupling to the magnetic field (see e.g. [45]):

$$2\hbar\mu_B \mathbf{S} \cdot \mathbf{B},$$

$\mathbf{S}$ denoting the spin, and $\mu_B = e\hbar/2m$ the Bohr magneton. This famously leads to an energy splitting between electrons with spin parallell (spin up) and antiparallell (spin down) to the magnetic field, known as Zeeman splitting. The energy difference between the two spin states is

$$\Delta E_Z = \mu_B g B,$$

where the Landé g-factor $g = 2$ in the low energy regime. Seemingly, this is the same as the separation between two Landau levels, i.e. spin up in Landau level n has the same energy as spin down in Landau level $n + 1$.

What actually happens in systems of strongly coupled charge carriers in dirty systems is a renormalization of the mass and Landé factor: $m \mapsto m^*$ and $g \mapsto g^*$. The effective mass, m^*, should then replace the masses in kinetic terms of the Hamiltonian. Often m^* is only a few precent of the original m, leading to a massive increase in Landau energy. The g-factor is also increased, but usually not a much as m is decreased, and hence the Zeeman splitting is in most cases much smaller than the Landau splitting [43]. The energy spectrum is then given as in Fig. 3.3.1, where the spin degeneracy has been lifted.

In real materials, the energy levels depicted in Fig. 3.3.1 are broadened into bands (compare Fig. 3.3.5), and depending on the conditions of the experiment, a spin degeneracy may yet be observed. This is achieveable with the right amount of disorder in the system, or by effectively increasing the disorder by applying a *back gate voltage*, increasing the electron density in the system, see e.g. [46] for an example. Such systems will be termed *spin-degenerate*, whilst systems in which the spin degeneracy is lifted are called *spin-polarized*.

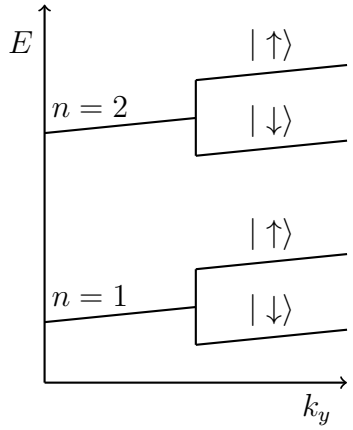

Figure 3.3.1: Energy spectrum of the simplified QH model. $|\uparrow\rangle$ and $|\downarrow\rangle$ denote spin up and spin down states, respectively.

3.3.3 The Brillouin Zone and Valley Degeneracy

A common way of describing single-electron dynamics in a solid is by modelling the atoms as stationary points on a lattice, constituting sources of a periodic potential. Consider non-interacting electrons moving in a rectangular lattice with lattice constants a and b in the x - and y-directions, subject to a potential $V(\mathbf{r}) = V(\mathbf{r}+(na, mb))$ with n, m integers. The wavefunctions are then restricted by *Bloch's theorem* (see e.g. chapter 10 of [47]):

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \qquad u_{\mathbf{k}}(\mathbf{r} + (na, mb)) = u_{\mathbf{k}}(\mathbf{r}), \tag{3.3.5}$$

where $\mathbf{k} = (k_x, k_y)$ and $\mathbf{r} = (x, y)$. It is well known that this kind of periodic potential automatically yields the band structure of solids and $\mathbf{k}$ can be used to label distinct states within each band. Assuming N_x and N_y lattice points in the x - and y-directions and enforcing periodic boundary conditions, one gets the further restrictions:

$$e^{i\mathbf{k}\cdot\mathbf{r}+ik_xN_xa} = e^{i\mathbf{k}\cdot\mathbf{r}+ik_yN_yb} = e^{i\mathbf{k}\cdot\mathbf{r}},$$

yielding $k_x = n_x 2\pi/N_x a$ and $k_y = n_y 2\pi/N_y b$ with $n_x \in \{1, ..., N_x\}$ and $n_y \in \{1, ..., N_y\}$. As such, one only needs to consider the region of reciprocal space spanned by $0 \leq k_x \leq 2\pi/a,\ 0 \leq k_y \leq 2\pi/b$, or equivalently

$$-\frac{\pi}{a} \leq k_x \leq \frac{\pi}{a}\ , \qquad -\frac{\pi}{b} \leq k_y \leq \frac{\pi}{b},$$

as any point outside this region is equivalent to a point inside it. The section of reciprocal space which spans all distinct configurations of the system within a given band is called the *Brillouin zone*, $\mathcal{B}$. It is the unit cell of the reciprocal space. The centre of the Brillouin zone is termed the Γ-point.

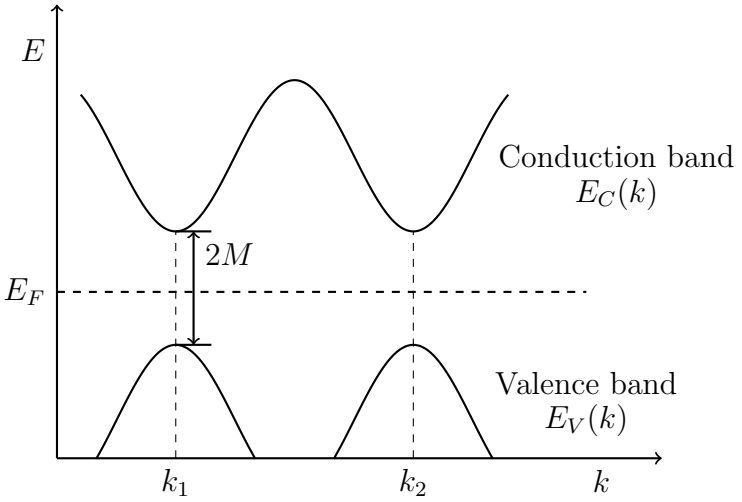

Figure 3.3.2: Illustration of valley degeneracy in a 1-dimensional semiconductor. The conduction band has two equivalent minima close to the Fermi level E_F, making k_1 and k_2 equivalent points in the interior of the Brillouin zone. They will then contribute an equal amount to conducting currents, yielding a degeneracy of 2. The mass gap parameter, M, interpreted as the band gap between the top valence and bottom conducting energies is shown.

It is common to plot the band structure of a system only over the Brillouin zone, as the pattern will be repetitive. In some materials, the energy has minima of equal magnitude, at different values of $\mathbf{k} \in \mathcal{B}$, with corresponding maxima in the valence band energy. Those points are termed equivalent points in the interior of the Brillouin zone and the system exhibits a corresponding *valley degeneracy* equal to the number of minima contributing to charge transport. The situation in a 1-dimensional solid is depicted in Fig. 3.3.2.

Note that if the equivalent points are situated at the corners of $\mathcal{B}$, then they are, in fact, the same point. This can be seen from a displacement of the origin of $\mathcal{B}$, as any region of the same dimensions is an equivalent representation. In this case, a valley degeneracy is not expected.

3.3.4 Topology of the Brillouin Zone

An interesting consequence of the double-periodicity of the wave vector is that $\mathcal{B}$ in the 2-dimensional case is a *torus*. This is seen by gluing the opposite sides of $\mathcal{B}$ together in an orientation-preserving way, see Fig. 3.3.3. In the prescence of a magnetic field, the situation is altered, but the Brillouin zone can still be identified with a torus, as will be shown here. This subsection follows the argument found in [48].

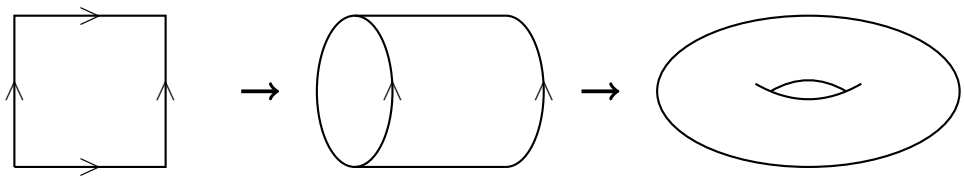

Figure 3.3.3: The process of gluing opposite sides of a rectangle together to create a torus.

Starting again with a single particle on a lattice, let $\mathbf{a}$ and $\mathbf{b}$ denote the two lattice vectors, taken to be along the x - and y-directions for simplicity, and $\mathbf{R} = n\mathbf{a} + m\mathbf{b}$, for some $n, m \in \mathbb{Z}$. We now turn on a uniform, perpendicular magnetic field. The immidiate problem is that the Hamiltonian (eq. (3.3.1)) is not invariant under translations as a translation works on the vector potential as $\mathbf{A}(\mathbf{r}) \to \mathbf{A}(\mathbf{r} + \mathbf{R})$, which is generally not the same. This means that the translation operator and the Hamiltonian do not commute, and hence do not have simultaneous eigenstates. Translation by lattice vectors are then not a symmetry of the system.

To remedy this problem, first recall that a translation operator which translates the system by $\mathbf{R}$ may be written as

$$T_{\mathbf{R}} = e^{(i/\hbar)\mathbf{R}\cdot\mathbf{p}},$$

$\mathbf{p}$ being the momentum operator. Then recall that for an electron in an electromagnetic system, the conjugate momentum is given as $\mathbf{p} = m\dot{\mathbf{x}} - e\mathbf{A}$, and the introduction of a magnetic field to a mechanical system manifests as the substitution $m\dot{\mathbf{x}} \to \mathbf{p} + e\mathbf{A}$ in the Hamiltonian.[5] This leads to a gauge-dependent interpretation of momentum. In that spirit, introduce the *magnetic translation operators*:

$$T^B_{\mathbf{R}} = \exp\left[\frac{i}{\hbar}\mathbf{R}\cdot(\mathbf{p} - e\mathbf{A})\right],$$

which in essence keeps the exponent equal to $i\mathbf{R}\cdot m\dot{\mathbf{x}}$. Now, T^B_R leaves the Hamiltonian invariant, and $[T^B_{\mathbf{R}}, H] = 0$ (see [49] for the full argument). The new issue is that translation by single lattice vectors $\mathbf{a}$ and $\mathbf{b}$ does not commute:

$$T^B_{\mathbf{a}}T^B_{\mathbf{b}} = e^{i2\pi\phi}T^B_{\mathbf{b}}T^B_{\mathbf{a}},$$

where $\phi = abB/\Phi_0$ is the number of magnetic flux quanta through one unit cell.

[5] Compare this with the quantization procedure in QFTs where the replacement $p_\mu \to -i\partial_\mu$ is made in the Lagrangian, followed by $-i\partial_\mu \to -iD_\mu = -i\partial_\mu + eA_\mu$ in the prescence of a gauge field.

If, however, ϕ is a rational number, $\phi = p/q$ with p, q relatively prime, a subset of the translations commute:

$$T^B_{q\mathbf{a}} T^B_{\mathbf{b}} = T^B_{\mathbf{b}} T^B_{q\mathbf{a}}.$$

We may then define a new effective lattice spanned by $q\mathbf{a}$ and $\mathbf{b}$, such that the system is invariant to translations by lattice vectors of the form $\mathbf{R}' = n(q\mathbf{a}) + m\mathbf{b}$, $n, m \in \mathbb{Z}$.

Now simultaneous eigenfunctions of H and $T^B_{\mathbf{R}'}$ can be found: if ψ is such a function, the eigenvalues of unit translations are given as:

$$T^B_{\mathbf{q}a}\psi = e^{ik_x qa}\psi$$
$$T^B_{\mathbf{b}}\psi = e^{ik_y b}\psi,$$

with k_x and k_y generalized crystal momenta restricted by $k_x \in [0, 2\pi/qa]$, $k_y \in [0, 2\pi/b]$. This once again yields the Brillouin zone torus, with one dimension reduced by a factor of q.

States within a single band can again be labeled by the crystal momentum $\mathbf{k} = (k_x, k_y)$, and written in Bloch form:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}).$$

The periodicity properties of the $u_{\mathbf{k}}$ are given by the action of the translation operator:

$$u_{\mathbf{k}}(\mathbf{r} + q\mathbf{a}) = e^{-i\pi py/b} u_{\mathbf{k}}(\mathbf{r})$$
$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{b}) = e^{i\pi px/qa} u_{\mathbf{k}}(\mathbf{r}),$$

known as the generalized Bloch conditions.

3.3.5 Dispersion and Effective Mass

Close to the minima of the conduction band, where charge transport may occur in a semiconductor, the energy can be approximated by a parabola. If we take a minimum to lie at the Γ-point $(k_x, k_y) = (0, 0)$:

$$E_C(k) \approx E_C(0) + \frac{1}{2}\left(\frac{\partial^2 E_C}{\partial k_i \partial k_j}\right)_{k=0} k_j k_i,$$

leading to an *effective mass matrix*

$$m^*_{ij} = \left(\frac{\partial^2 E_C}{\partial k_i \partial k_j}\right)^{-1}_{k=0}.$$

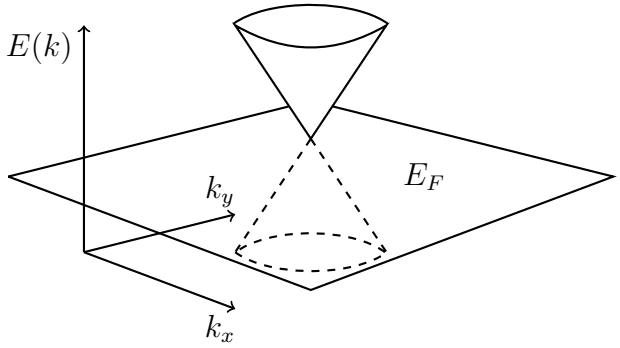

Figure 3.3.4: Band gap closure at the Fermi level surface, resulting in a Dirac cone and massless particles.

We may further approximate the situation by constraining all electrons to belong either to the conducting or valence band. Then, from the Fermi-Dirac distribution:

$$1 = \frac{1}{\exp[(E_C - E_F)/kT] + 1} + \frac{1}{\exp[(E_F - E_V)/kT] + 1},$$

which gives

$$E_C + E_V = 2E_F,$$

so the Fermi level is exactly halfway in between the conducting and valence bands. This means that

$$m^*_{ij} = \left(\frac{\partial^2(E_C - E_F)}{\partial k_i \partial k_j}\right)^{-1}_{k=0} = -\left(\frac{\partial^2(E_F - E_V)}{\partial k_i \partial k_j}\right)^{-1}_{k=0} = \frac{1}{2}\left(\frac{\partial^2(E_C - E_V)}{\partial k_i \partial k_j}\right)^{-1}_{k=0},$$

when the Fermi level is taken to have approximately zero curvature around the Γ-point. Now, using the Fermi level as a reference energy, around the Γ-point we have

$$E_C(k) - E_F \approx M + \frac{1}{2}\frac{k_i k_j}{m^*_{ij}},$$

where

$$M = E_C(0) - E_F(0) = \frac{1}{2}(E_C(0) - E_V(0)),$$

is termed the *mass gap parameter* or *Dirac mass parameter*, as it is the one appearing in the Dirac Hamiltonian (see e.g. [50]).

Hence, there are two interpretations of mass in this problem and they may in principle differ. However, the aspect of both these mass interpretations relevant here is that if the band gap closes, they both disappear. Closing a band gap means, of course, that the valence and conduction bands touch, so automatically $M = 0$. At this point, we may approximate the energy by a linear function. Linear functions are odd, and so the quadratic terms in the energy expansion

vanishes and the effective mass diverges: $m^*_{ij} \to \infty$. The interpretation of the effective mass as the reciprocal of the curvature of $E_C(k)$ loses its meaning in that context. However, the dispersion relation takes a familiar form. In an isotropic medium:

$$E(k) - E_F \approx v_F k,$$

where $v_F = \left(\frac{\partial E}{\partial k}\right)_{k=0}$ is the Fermi velocity. We hence have no mass term and *relativistic dispersion*, compare $E = p$ for particles moving at the speed of light. This means that particles in a closed-gap system are effectively massless.

In materials where this happens, the conduction and valence bands collectively give the dispersion relation the shape of a cone, called a *Dirac cone*, see Fig. 3.3.4.

3.3.6 Laughlin's Argument for the Integer Effect

The first widely accepted theory of the IQHE is Laughlin's theory of gauge invariance and localization. Even though this method of understanding the IQHE has given way to topological theories, it is included here as its resemblance to band theory provides an intuitive picture of the IQHE. In this picture [36] one views the quantized conductivities as a consequence of the disorder (e.g. crystalline impurities) apparent in any real material. The effect of this disorder is a broadening of the Landau levels into bands, depicted in Fig. 3.3.5. At a plateau, all states are localized at impurities and can not contribute to charge transport. This is illustrated by the Fermi level being in a mobility gap.

Laughlin considered the "ribbon geometry", obtained by gluing the current-injecting edges of the Hall bar together, making a ribbon of circumference L_x, carrying a current I_x, everywhere penetrated by a perpendicular magnetic field B, and subject to a voltage drop V_y across it. The Landau gauge $\mathbf{A} = By\mathbf{e}_x$ yields a vector potential everywere pointing around the ribbon. Laughlin argued that the current relates to the voltage drop through an adiabatic derivative of the total electonic energy U with respect to the magnetic flux through the surface of the ribbon, which again relates to a uniform change in the vector potential:

$$I_x = \frac{\partial U}{\partial \Phi} = \frac{1}{L}\frac{\partial U}{\partial A},$$

where the change in A is viewed as a gauge transformation. The consequence of this gauge transformation on the wavefunctions is the usual multiplication by phase:

$$\psi \to \exp[ieAx/\hbar]\psi.$$

Now, for a delocalized state, this transformation is illegal unless

$$A = n\frac{h}{eL} = n\frac{\Phi_0}{L},$$

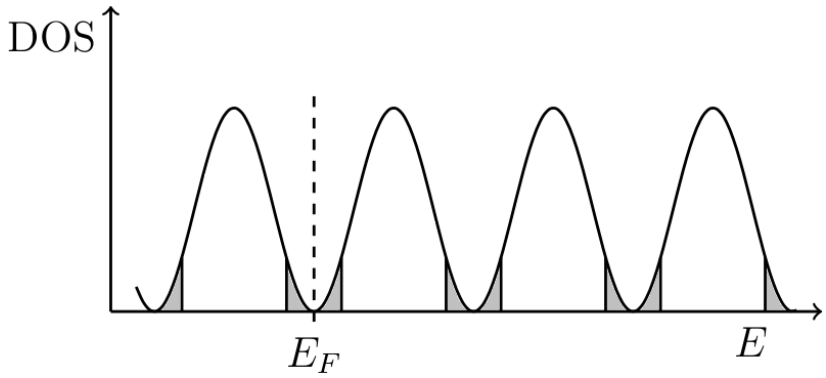

Figure 3.3.5: Illustration of the density of states as a function of energy in the QHE in a disordered material. The shaded regions mark localized states, incapable of contributing to charge transport. On a plateau, the Fermi energi E_F lies in a mobility gap consisting of localized states.

to get the correct periodicity around the loop. By this gauge invariance, adding one flux quantum through the ribbon maps the system back onto itself. In that case, the only increase in energy possible must arise from the transport of ν electrons from one side of the ribbon to the other. This yields

$$I_x = \frac{\Delta U}{\Delta \phi} = \frac{\nu e V_y}{\phi_0} = \nu \frac{e^2}{h} V_y,$$

so the Hall conductance appears in quanta of

$$G_K = \frac{e^2}{h}.$$

A more detailed presentation is given in [51], where it is argued that it is natural to identify the integer ν with the number of filled Landau levels below the Fermi level. The derivation above did not take spin or valley degeneracy into consideration, so the correct expression would be

$$\sigma_{xy} = g_s g_v \nu \frac{e^2}{h}, \tag{3.3.6}$$

$g_s \in \{1, 2\}$ and $g_v \in \mathbb{N}$ denoting spin and valley degeneracies respectively. As we will see shortly, at a plateau, the diagional components of conductivity and resisitivity vanishes: $\rho_D = \sigma_D = 0$, so an equivalent formulation is

$$\rho_{xy} = g_s g_v \frac{h}{e^2} \frac{1}{\nu}. \tag{3.3.7}$$

We now move on to see what is actually observed in experiments.

3.4 Experimental Integer Quantum Hall Effect

What is found in experiments resembles what is depicted in Figs. 3.4.1, 3.4.2 and 3.4.3. The following will highlight characteristics of the different systems.

3.4.1 Spin Polarization and Degeneracy

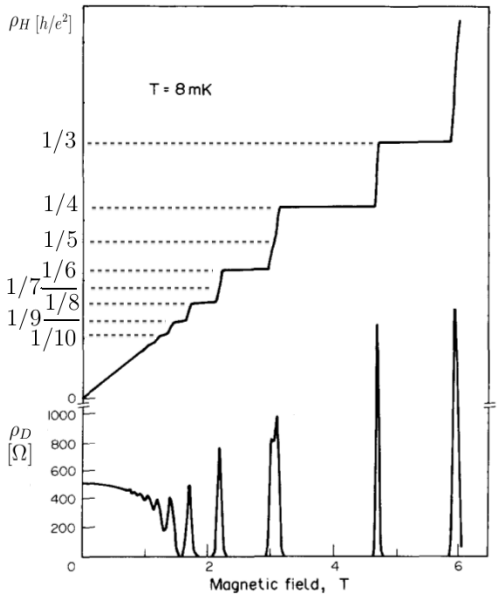

Figure 3.4.1: The spin-polarized integer quantum Hall effect in a GaAs-AlGaAs heterojunction, showing the sequence $\nu = 3, 4, 6, 8, 9, 10$, and a peak in ρ_D about $B = 6T$ signalling the transition to the $\nu = 2$ plateau. Modified from [52].

The longitudinal resistance, ρ_D, vanishes at most values of B while the Hall resistance, ρ_H, is quantized to plateaux, impervious to small variations in B, situated exactly on integer values of R_K. When B is sufficiently tuned, ρ_H jumps to a new plateau, accompanied by a sharp spike in ρ_D.

Shown in Fig. 3.4.1 is a subset of the sequence $\nu = 0, \pm 1, \pm 2, ...$, the 0-plateau is recognized in experiment by a diverging resistivity $\rho_H \to \infty$, leading to the insulator plateau. This would have appeared in Fig. 3.4.1, had the magnetic field been increased further and the measuring equipment not reached saturation. The absence of the $\nu = 5, 7$ plateaux may be attributed to experimental error.

Fig. 3.4.2 shows another oberved sequence in the IQHE. Depicted here is a subsequence of the even numbered plateaux $\nu = 0, \pm 2, \pm 4, ...$, namely $\nu = 2, 4, 6$. Interestingly, both of the experiments in Figs. 3.4.1 [52] and 3.4.2 [46] were done

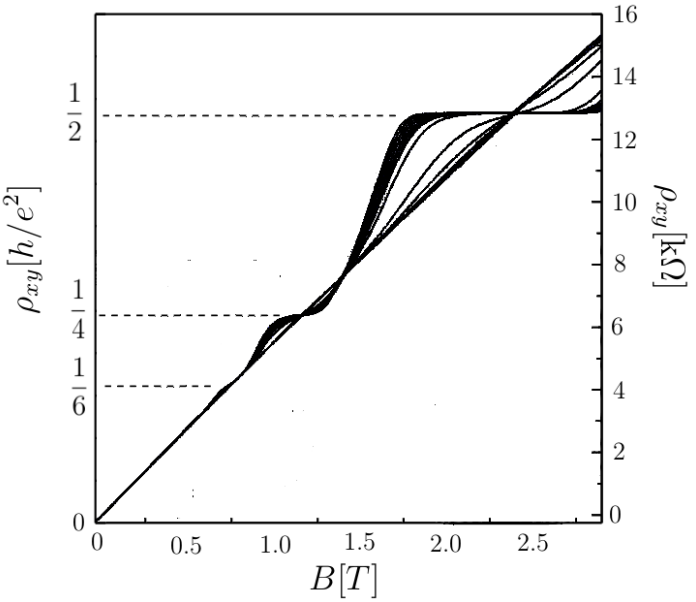

Figure 3.4.2: The spin-degenerate IQHE in a GaAs-AlGaAs heterojunction, showing the sequence $\nu = 2, 4, 6$. The experiment was done in the range 0.3-16K at a constant back-gate voltage of $V_g = -0.145$V. Modified from [46].

in GaAs-AlGaAs heterojunctions, each of similar characteristics. [52] uses layers of 1µm GaAs, 10-15nm $Al_{0.3}Ga_{0.7}As$, 140nm Si-doped $Al_{0.3}Ga_{0.7}As$ and a 20nm GaAs cap, grown on a semi-insulating GaAs-substrate. [46] uses the same, except with doping concentration $x = 0.33$ in $Al_xGa_{1-x}As$ and layer thicknesses 1µm, 200nm, 40nm and 10nm.

Changing the thickness of layers should not so drastically change the band structure of the heterojunction that it yields a new equivalent point in the Brillouin zone. The degeneracy observed in Figure 3.4.2 is then due to spin, and attributable to the applied gate voltage.

3.4.2 Spin and Valley Degeneracy

Valley degeneracies are rarely observed in the quantum Hall effects. Fig. 3.4.3, however, shows a sequence corresponding to a total degeneration of $g = g_v g_s = 4$. As spin degeneracy can only account for a factor of 2, there must be at least two equivalent points in the interior of the Brillouin zone.

Fig. 3.4.3 exhibits the sequence $\nu = 2, 6, 10, 14, 18$, considered a subsequence of the general $\pm 2, \pm 6, \pm 10, ...$ The extraordinary characteristic of this sequence is the fact that there is no 0-plateau. The consequence is that Laughlin's explanation cannot be the whole truth as eqs. (3.3.6) and (3.3.7) clearly predict

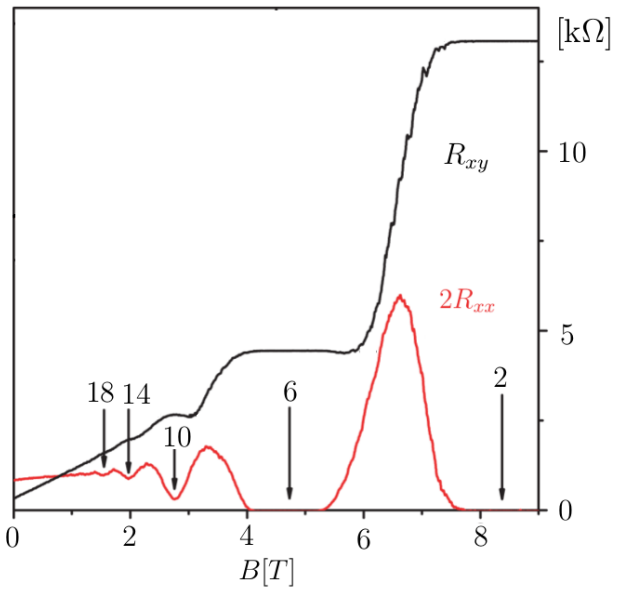

Figure 3.4.3: The spin and valley degenerate IQHE in graphene, showing the sequence $\nu = 2, 6, 10, 14, 18$. The experiment was done at 30mK and a back-gate voltage of 15V. Modified from [53].

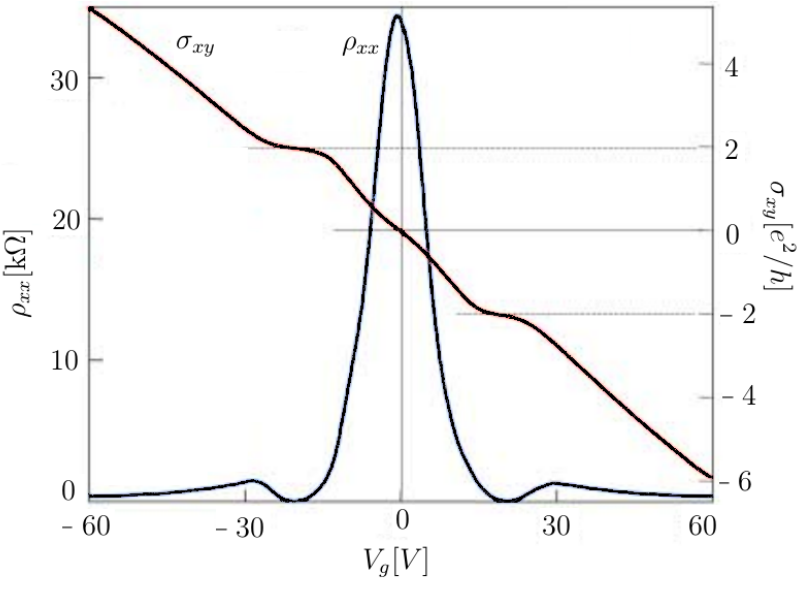

Figure 3.4.4: Room temperature IQHE in graphene, showing the plateaux $\nu = \pm 2$. Note the missing 0-plateau. The experiment vas done at 300K at a strong magnetic field of 29T. Modified from [54].

insulating behaviour for $\nu = 0$.

The lack of a zero plateau is confirmed experimentally, portrayed in Fig. 3.4.4. That experiment is also particularly interesting as it was performed at *room temperature*, that is, $T = 300\text{K}$.

The lack of a zero-plateau, quadruple degeneracy and room temperature QHE are all novel features of *graphene*, a very exotic material which deserves special mention.

3.4.3 Graphene

Graphene is a relatively new material with tons of possible applications. During the past decade a morass of research has been dedicated to this material and its (hypotethical) cousins stanene, silicene and germanene. Whole books have been written solely about the physics of graphene[6] and it is, again, an impossible task to give a complete review. Presented here are some key aspects of graphene in relation to the QHE.

Graphene consists of carbon atoms structured in a hexagonal lattice. What makes it so special is that the lattice can be manufactured a single atom thick. This automatically yields an effectively 2-dimensional system. In addition, the hexagonal lattice must be viewed as two inequivalent triangular sublattices in order to define lattice vectors for which the system is translationally invariant, and which spans the entire system. The two sublattices may then share the set of lattice vectors, but two distinct origins is required. The situation is depicted in Fig. 3.4.5.

The Brillouin zone, $\mathcal{B}$, of graphene is also a hexagon, and the band structure of graphene has *six* equivalent points, situated at the corners of $\mathcal{B}$ [55]. Generally when equivalent points are situated at the corners, no valley degeneracy is expected because the points would, topologically, be the same point (see subsection 3.3.3). Graphene is different. Because of the subdivision of the lattice, $\mathcal{B}$ too is subdivided into two inequivalent subtriangles, each with three equivalent points at their corners. There are then two *distinct* groups of equivalent points at the corners, and the situation is as if one has two equivalent points in the interior of $\mathcal{B}$: a valley degeneracy of $g_v = 2$.

This fact is also seen by a simple translation of the unit cell $\mathcal{B}$ in reciprocal space of any magnitude less than the dimensions of $\mathcal{B}$. This action will leave exactly two equivalent points in the interior. Whichever points they are, they are usually named K and K' and called the *Dirac points*, for reasons explained shortly.

[6] See [55] for a thorough study of experimental and theoretical aspects of graphene

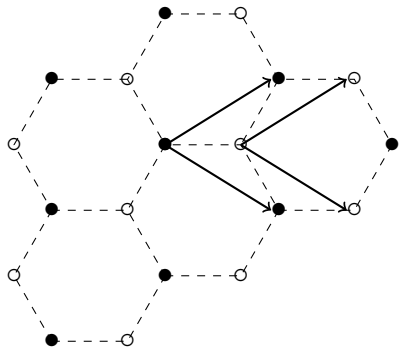

Figure 3.4.5: Honeycomb lattice structure of graphene and its division into two sublattices.

Another novel feature of graphene is its relativistic dispersion relation. As we have seen (subsection 3.3.5), a relativistic dispersion goes hand in hand with a closed band gap and linearized energy, and this is exactly what happens at the Dirac points: Dirac cones form. This means that the mobile electrons in graphene move at the Fermi velocity and are effectively massless.

Indeed, in a massless Dirac theory, the Hall conductance in graphene takes on the observed value [56]:

$$\sigma_{xy} = g_s g_v \left(\nu + \frac{1}{2}\right) \frac{e^2}{h} = 2(2n+1)\frac{e^2}{h},$$

where the extra term of $1/2$ can be interpreted as a consequence of the lowest Landau level being half filled at charge neutrality [55]. This implies that the $\nu = 0$ level has exactly half the degeneracy of the $\nu > 0$ levels.

3.5 Fractional Quantum Hall Effect

The FQHE will not be covered at length here. What is important to this thesis is that the experimental results coincide with the predictions of the theory to be proposed in the next chapter. The general situation in the FQHE is that the measured Hall conductance is a *rational number* in units of e^2/h:

$$\sigma_{xy} = \frac{e^2}{h}\frac{p}{q}, \tag{3.5.1}$$

with $p, q \in \mathbb{N}$ relatively prime.

The original discovery of the FQHE by Tsui and Stormer [37] found the $1/3$ plateau in a GaAs-AlGaAs heterojunction. Illustration from an experiment in a semiconductor heterojunction is shown in Figure 3.5.1.

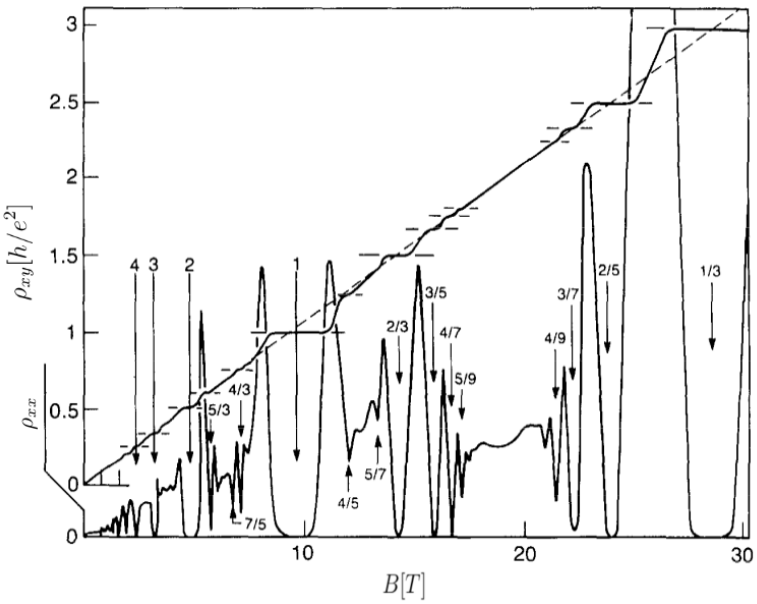

Figure 3.5.1: Fractional quantum Hall effect in a GaAs-AlGaAs structure with high carrier mobility. ρ_{xy} and ρ_{xx} are plotted together, not to scale. Modified from [57].

A striking feature of the fractional plateau structure depicted in Fig. 3.5.1 is that they seem to be constrained to filling factors, ν, with odd-number denominators. This is indeed what is normally observed.[7] Note that the material used in this experiment is a GaAs-AlGaAs heterojunction, the same type of structure used as an example of IQHE in the preceding section.

The experimental condition for observing FQHE, rather than IQHE seems to rest entirely on the electron mobility: FQHE is observed in very clean samples with little disorder. This means a higher mobility. In [57] (Fig. 3.5.1) the carrier mobility was $\mu \approx 1.3 \cdot 10^6 \text{cm}^2/\text{Vs}$, compare with $\mu = 4.1 \cdot 10^4 \text{cm}^2/\text{Vs}$ and $\mu = 3.79 \cdot 10^4 \text{cm}^2/\text{Vs}$ in [52] (Fig. 3.4.1) and [46] (Fig. 3.4.2), respectively.

Graphene displays a peculiar behaviour in the FQHE as well. FQHE was discovered in graphene in 2009 [58] [59], in suspended graphene devices: free of any substrate that leads to additional disorder. A recent study [60] found the sequence

$$\nu = \frac{1}{3}, \frac{2}{3}, \frac{2}{5}, \frac{3}{5}, \frac{3}{7}, \frac{4}{7}, \frac{4}{9}, \tag{3.5.2}$$

for $\nu < 1$, and

$$\nu = \frac{4}{3}, \frac{8}{5}, \frac{10}{7}, \frac{14}{9}, \tag{3.5.3}$$

[7] Even-numbered denominators (half-integer) effects have been confirmed at present.

for $\nu > 1$. The special feature is that for $\nu > 1$, the plateaux seems to be constrained to fractions with even numerator.

3.5.1 Laughlin's Argument for the Fractional Effect

As already mentioned, the FQHE is described by Laughlin as an incompressible quantum fluid with fractionally charged particles. This description suggests that the FQHE is due to a new kind of correlated many-electron ground state. Laughlin argues this wavefunction for this state of N non-interacting electrons takes the form [38]

$$\psi(z_1, ..., z_N) = \left(\prod_{j<k} f(z_j - z_k) \right) \exp\left(-\frac{1}{4} \sum_{i=1}^{N} |z_i|^2 \right), \tag{3.5.4}$$

in complex coordinates $z_i = x_i + y_i$ and units $l_B = 1$. He further argues that the condition for the system to be in the lowest Landau level is that $f(z)$ is a polynomial and that conservation of angular momentum means that the order of the product in eq. (3.5.4) is L, L being the total angular momentum. Furthermore, the antisymmetry of the fermionic wavefunction ψ restrains f to be odd. The result is that f is of odd order m, with m the angular momentum of a single particle in the composite multiparticle ground state.

Laughlin goes on to show that this yields an effective fluid of quasiparticles of charge e/m. By gauge invariance arguments similar to those discussed in subsection 3.3.6, this gives the conductance $e^2/(mh)$, explaining the experimental value $e^2/(3h)$ found by Tsui and Stormer [37].

Laughlin's original procedure does, however, only predict plateaus of the form $\nu = 1/m$ with m odd and not the more general $\nu = p/q$. Over the years, certain modifications to Laughlin's model have been proposed in an attempt to explain the complete effect. An intuitive explanation arises from a proposed degeneracy of the ground state: if the degeneracy level of the ground state is p, then p flux quanta would have to be added for the system to return to its original state by the Laughlin argument. This would yield the Hall conductance $pe^2/(mh)$. The introduction of this degeneracy may, however, seem quite suspicious, taking into consideration that Laughlin's trial wavefunction (3.5.4) seems non-degenerate, and that non-degeneracy in the case of electrons on a sphere has been proved. See [61] and the references therein for a more complete discussion.

3.6 Topological Arguments

The explanations for the QHEs up until this point constitute a patchwork of ad-hoc theories and modifications, usually taken in constrained geometries. One would expect much more general arguments for such a universal phenomenon. This was achieved with the development of *topological theories*. Readers whose topology and differential geometry are a little rusty are advised to look through Appendices A and B for a recapitulation of the most important concepts.

3.6.1 Berry's Phase and the Berry Connection

The cleanest way to introduce the QHE as a topological phenomenon is through the Berry connection, and the cleanest way to introduce the Berry connection is through the Berry phase. As such, these concepts are introduced here. This subsection mainly follows mainly [45] for the introduction of Berry's phase and [43] for the introduction of the Berry connection.

We start off by considering a general Hamiltonian as a function not only of coordinates and velocities, but also of the set of parameters $\{g_i\}$ defining the theory: $H = H(\{g_i\})$. In section 2.6 we saw that this is indeed the correct way to view a Hamiltonian. Now, imagine the system this Hamiltonian is describing being in the n'th energy state: $|\psi_n\rangle$. We will explore how the state evolves if the parameters of the Hamiltonian is varied.

The *adiabatic theorem* (see e.g. [45]) states that if the parameters are varied while the system is in the n'th energy state, the state will evolve into the n'th state of the new system with new parameter values, given that the variation is done adiabatically.

If then, the parameters are varied adiabatically from the initial point $\{g_i^0\}$ in parameter space, through some closed loop $\{g_i(t)\},\ t \in [0, T]$, and end up back again at $\{g_i^0\}$, the system should return to the original state up to a possible phase difference

$$|\tilde{\psi}_n(T)\rangle = U(T)|\psi_n\rangle,$$

with $U(T)$ some phase. To compute this phase, observe that in general, for any point $\{g_i(t)\}$ along the curve in parameter space, we would have

$$|\tilde{\psi}_n(t)\rangle = U(t)|\psi_n(\{g(t)\})\rangle.$$

Taking the time derivative:

$$\frac{\partial}{\partial t}|\tilde{\psi}_n\rangle = \dot{U}|\psi\rangle + U|\dot{\psi}\rangle,$$

and finally the inner product with $\langle\tilde{\psi}_n|$:

$$\begin{aligned}
\langle\tilde{\psi}_n|\frac{\partial}{\partial t}|\tilde{\psi}_n\rangle &= U^*\dot{U}\langle\psi_n|\psi_n\rangle + U^*U\langle\psi_n|\dot{\psi}_n\rangle \\
\frac{1}{i\hbar}\langle\tilde{\psi}_n|H(t)|\tilde{\psi}_n\rangle &= U^*\dot{U} + \langle\psi_n|\frac{\partial g_i}{\partial t}\frac{\partial}{\partial g_i}|\psi_n\rangle \\
\frac{\dot{U}}{U} &= -\frac{i}{\hbar}E_n(t) - \langle\psi_n|\nabla_g|\psi_n\rangle\cdot\dot{\mathbf{g}} \\
U(T) &= \exp\left[-\frac{i}{\hbar}\int_0^T E_n(t')dt' + i\oint \mathbf{A}\cdot d\mathbf{g}\right],
\end{aligned} \tag{3.6.1}$$

where in the second line, we invoked the Schrödinger equation, and in the third used that $U^* = U^{-1}$ for a phase and wrote the parameters $\{g_i\}$ as a vector $\mathbf{g}$. In the fourth line, we see that the terms contributing to the phase is the standard, dynamical contribution, and the closed loop integral of a vector field over the parameter space. This contribution is known as the *Berry phase*, and the vector field is the *Berry connection*, defined here as

$$\mathbf{A} = i\langle\psi_n|\nabla_g|\psi_n\rangle,$$

where the gradient is differentiating with respect to the parameters g_i.

Now, the Berry connection is constructed from the states $|\psi_n\rangle$, and the choice of these intermediate states is only unique up to a phase. Consider then, if a different set of states, $\exp[i\omega(\mathbf{g})]|\psi_n\rangle$, were chosen. The Berry connection is altered as

$$\mathbf{A}' = \langle\psi_n|e^{-i\omega(\mathbf{g})}\nabla_g e^{i\omega(\mathbf{g})}|\psi_n\rangle = \mathbf{A} + i\nabla_g\omega,$$

which is reminiscent of a $U(1)$ gauge transformation, and does not alter the integral in eq. (3.6.1). Indeed, the Berry connection is a proper one-form connection as discussed in Appendix B. We may then compute the *Berry curvature* (or Berry field strength) as

$$F = \mathrm{d}A = \partial_i A_j \mathrm{d}g^i \wedge \mathrm{d}g^j.$$

The curvature is then independent of the states used to define the connection as per its gauge invariance. The curvature F, that is, *the curvature of the vector bundle with base space the parameter space of the system,* is then to be considered a topological artifact of the system.[8]

Finally observe that through Stokes' theorem, the integral in eq. (3.6.1) may now be written as

$$\int F,$$

which is manifestly independent of the specific construction of A.

[8]The term 'vector bundle' is not defined in Appendix B, but think of it as a fiber bundle where the fiber is a vector space.

3.6.2 Hall Conductivity as the First Chern Number

We are now ready to show that the exact quantization of the IQHE is a consequence of a topological invariant. This subsection follows mainly [48]. Elements from [43] are also used, and that reference is recommended for an introduction to the Chern form and number.

First, recall the results of subsection 3.3.4. If one models the QHE as non-interacting electrons in a 2-dimensional lattice, subject to a perpendicular magnetic field, the Brillouin zone is a torus and the wavefunctions for single particles may be written in Bloch form. Letting the index α denote a single band, write

$$\psi^{\alpha}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u^{\alpha}_{\mathbf{k}}(\mathbf{r}),$$

with $\mathbf{k}$ labeling the different states within the band α. The Schrödinger equation for a single particle may then be rewritten as

$$H(\mathbf{k}) u^{\alpha}_{\mathbf{k}} = E^{\alpha} u^{\alpha}_{\mathbf{k}},$$

with $H(\mathbf{k}) = \frac{1}{2m}(-i\hbar\nabla + \hbar\mathbf{k} + e\mathbf{A})^2 + V(\mathbf{r})$, $V(\mathbf{r})$ denoting the periodic lattice potential.

Our starting point is the linear response formula for charge transport in the quantum Hall effect: the Kubo formula, which takes the form

$$\sigma_{xy} = -ie^2\hbar \sum_{E^{\alpha}<E_F<E^{\beta}} \frac{(v_y)_{\alpha\beta}(v_x)_{\beta\alpha} - (v_x)_{\alpha\beta}(v_y)_{\beta\alpha}}{(E^{\alpha} - E^{\beta})^2},$$

where the sum is over all bands of energy less than or greater than the Fermi level. The v's denote the matrix elements of the velocity operator $\mathbf{v} = (-i\hbar\nabla + \mathbf{A})/m$, and may be found from the dispersion relation of the Hamiltonian:

$$\begin{aligned}
(v_i)_{\alpha\beta} &= \frac{1}{\hbar}\left\langle u^{\alpha}_{\mathbf{k}} \middle| \frac{\partial H}{\partial k_i} \middle| u^{\beta}_{\mathbf{k}} \right\rangle \\
&= \frac{1}{\hbar}(E^{\beta} - E^{\alpha})\left\langle u^{\alpha}_{\mathbf{k}} \middle| \frac{\partial u^{\beta}_{\mathbf{k}}}{\partial k_i} \right\rangle \\
&= \frac{1}{\hbar}(E^{\alpha} - E^{\beta})\left\langle \frac{\partial u^{\alpha}_{\mathbf{k}}}{\partial k_i} \middle| u^{\beta}_{\mathbf{k}} \right\rangle,
\end{aligned}$$

with $i = x, y$, and the inner product is taken over real and reciprocal space.

Inserting this into the Kubo formula yields

$$\sigma_{xy} = -i\frac{e^2}{\hbar} \sum_{E^{\alpha}<E_F<E^{\beta}} \left(\langle \partial_1 u^{\alpha}_{\mathbf{k}} | u^{\beta}_{\mathbf{k}} \rangle \langle u^{\beta}_{\mathbf{k}} | \partial_2 u^{\alpha}_{\mathbf{k}} \rangle - \langle \partial_2 u^{\alpha}_{\mathbf{k}} | u^{\beta}_{\mathbf{k}} \rangle \langle u^{\beta}_{\mathbf{k}} | \partial_1 u^{\alpha}_{\mathbf{k}} \rangle \right),$$

where $\partial_i = \partial/(\partial k_i)$. Inserting the identity $\sum_{E^\alpha < E_F < E^b} (|u^\alpha_{\mathbf{k}}\rangle\langle u^\alpha_{\mathbf{k}}| + |u^\beta_{\mathbf{k}}\rangle\langle u^\beta_{\mathbf{k}}|) = \mathbf{1}$, yields for one filled band, α:

$$\sigma^\alpha_{xy} = \frac{e^2}{h}\frac{1}{2\pi i}\int d^2k \left(\langle\partial_2 u^\alpha_{\mathbf{k}}|\partial_1 u^\alpha_{\mathbf{k}}\rangle - \langle\partial_1 u^\alpha_{\mathbf{k}}|\partial_2 u^\alpha_{\mathbf{k}}\rangle\right), \tag{3.6.2}$$

with the inner product taken only over real space. However, observe the similarity of this form of the conductivity with the form of the Berry curvature of section 3.6.1. Indeed, if we define the Berry connection as before:

$$A_i = i\langle u^\alpha_{\mathbf{k}}|\partial_i|u^\alpha_{\mathbf{k}}\rangle,$$

we see that the curvature

$$F_{ij} = \partial_i A_j - \partial_j A_i = i\left(\langle\partial_i u^\alpha_{\mathbf{k}}|\partial_j u^\alpha_{\mathbf{k}}\rangle - \langle\partial_j u^\alpha_{\mathbf{k}}|\partial_i u^\alpha_{\mathbf{k}}\rangle\right),$$

or

$$F = \partial_i A_j \mathrm{d}k^i \wedge \mathrm{d}k^j.$$

Then, eq. (3.6.2) may be written as

$$\sigma^\alpha_{xy} = -\frac{e^2}{h}\frac{1}{2\pi}\int_{T^2} F, \tag{3.6.3}$$

where the integral is over the entire Brillouin torus (recall subsection 3.3.4).

The quantity $\frac{1}{2\pi}F$ is known as the *first Chern form*, and its integral over a torus

$$C_1 = \frac{1}{2\pi}\int_{T^2} F,$$

is the *first Chern number*. The amazing fact is that *the first Chern number is always an integer*, independent of the particular connection chosen to calculate it. It is given only in terms of the topology of the principal $U(1)$-bundle where the connection $\mathbf{A}$ lives, see e.g. the chapter on Chern-Simons theory in [62] for a formal proof. The total conductivity is finally obtained by summing over all contributing (filled) bands:

$$\sigma_{xy} = -\frac{e^2}{h}\sum_\alpha C_1^{(\alpha)}.$$

The transition into viewing the QHE in terms of topological features signifies an upheaval from the old Laughlin picture of localization. Impurities, where localization of charge carriers take place, do indeed seem *irrelevant* in this discussion. This is also the power of the topological understanding of the problem. The robustness of the QHE, i.e. the fact that it manifests in widely different materials and under different conditions, may stem from the fact that these differences do not alter the topology of the Brillouin zone.

3.6.3 Hofstadter: A Possible Fallacy

There is a certain point in the above argument that is immidiately suspicious. Recall, again, subsection 3.3.4, where it was shown that the Brillouin zone is a torus, even when magnetic fields are present. That argument rests entirely on the condition that the parameter $\phi = abB/\Phi_0$, with $\Phi_0 = h/e$, is a rational number. Now, in a world where everything is quantized at the most fundamental level, it is indeed reasonable to assume that $\phi = p/q$, with p and q two (possibly very large) integers. In any case, one certainly never *measures* irrational numbers, that is by default an impossibility.

It is, however, worth mentioning that the case of electrons moving in 2-dimensional lattices had already been studied quite a lot before the discovery of the QHE. One of the most notable theoretical studies was performed by Hofstadter in his 1976 paper "Energy Levels and Wave Functions of Bloch Electrons in Rational and Irrational Magnetic Fields" [63], where he derives his famous *butterfly*. The condition set by Hofstadter for wether a magnetic field is "rational" or "irrational", is precisely that of ϕ being rational or irrational.

For $\phi \in \mathbb{Q}$, the butterfly spectrum for the energy, see Fig. 3.6.1, is obtained,

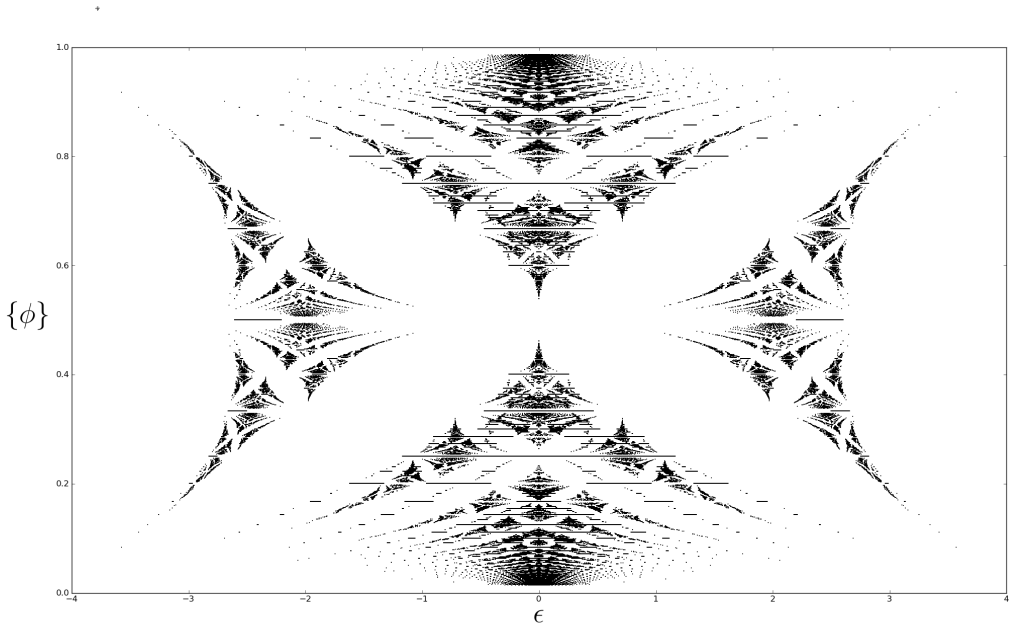

Figure 3.6.1: The energy spectrum of an electron on a lattice in a strong magnetic field: the Hofstadter butterfly. On the horizontal axis is the reduced energy, ϵ, and the vertical axis is the rational part of ϕ, denoted $\{\phi\}$. The spectrum has been generated by the procedure outlined in [63].

whereas when ϕ is irrational, the spectrum is homeomorphic to the *Cantor set*: an uncountably infinite set of measure zero.[9] The butterfly spectrum fits nicely into the QHE picture, with its discretized energy plateaus at fractional values. The odd-denominator fractions even plays a special role in Hofstadter's model as well (compare with odd-denominator FQHE). The reader is referred to the original paper [63] for the details.

Now, for an energy spectrum in the form of a Cantor set, the situation is significantly different. Recall Laughlin's argument for the integer effect (subsection 3.3.6). There, the conductivity was linked with the electronic energy. If the energy spectrum is a Cantor set, this necessarily means there will be irrational numbers lurking in it.[10] In that case, there will be *irrational* plateaux of the QHE. However, speaking of plateaux at all could be meaningless in that context, as they could in principle be infinitely close to each other, making the conductivity spectrum of the QHE a Cantor set of its own.

That situation is incompatible with the QHE. One may debate the realism of ϕ being an irrational number. The rationals are, after all, *dense* in the real numbers, meaning that for any $x \in \mathbb{R}$, rational or irrational, there are elements of $\mathbb{Q}$ infinitely close to it. It is, however, unsettling that measurable quantities should depend on the rationality or irrationality of a certain number, as arbitrarily small perturbations in that number would drastically affect what is measured. This is a point Hofstadter himself makes in [63].

3.6.4 Chern-Simons Theory

Now it is time to turn to the cutting edge forefront of modern condensed matter physics: topological effective field theories. The details of Chern-Simons theory in topological terms will not be reviewed here, as it is cumbersome and not essential to this thesis. However, take note that the exact form of the *Chern-Simons action* (see below) can be properly motivated from topological considerations alone (from the *second* Chern form $F \wedge F$). See e.g. [62] for a solely topological argument.

The Chern-Simons theory is an *effective* field theory in the sense that renormalization will not alter the parameters drastically, that is, it is a theory in the vicinity of an attractive fixed point. It is this point of view that we will adopt here, in constructing it. In the following, set $\hbar = c = 1$.

First, recall the discussion of subsection 2.8.2. In constructing the effective field theory, only terms with parameters of zero or positive mass dimensions are al-

[9] Basically meaning it is *there* and it is *huge*, but it has zero area.

[10] Recall that the set of all rationals, $\mathbb{Q}$, is countable, as it is in principle possible to systematically tabulate all rational numbers.

lowed in the Lagrangian. The other fundamental restriction we put on the theory is that of $U(1)$-gauge invariance. As we work in $d = 3$, that is, (2+1), all terms in the QED Lagrangian (eq. (2.8.2)) are allowed. However, none of them gives rise to the QHE in the following sense (see [43], or [64] for the entire argument).

In any action-based theory, the functional derivative of the action with respect to the vector potential yields a current. For instance, with QED:

$$j^\nu = \frac{\delta S}{\delta A_\nu} = \frac{\delta}{\delta A_\nu}\left[-e\int d^d x \bar{\psi}\gamma^\mu \psi A_\mu\right] = -e\bar{\psi}\gamma^\nu\psi,$$

the Dirac vector current. In the context of the QHE, the current needs to be related to the electric field, via the vector potential. The proportionality factor will then be the conductivity matrix. The QED Lagrangian lacks terms with this feature.

Indeed, any Lagrangian term constructed from $\mathbf{E}$ or $\mathbf{B}$ (such as the $F_{\mu\nu}F^{\mu\nu}$-term of QED) is immediately out of the question, as these are only proportional to the derivatives of A, not A itself. Any term including matter fields, such as ψ, are also out as there should be no matter fields in the expression for the current.

Now, by some kind of miracle, in (2+1) dimensions there is exactly one term obeying all these constraints [64]. The resulting action is termed the *Chern Simons action*:

$$S_{CS} = \frac{k}{4\pi}\int d^3x \epsilon^{\mu\nu\rho} A_\mu \partial_\nu A_\rho, \tag{3.6.4}$$

where $\epsilon^{\mu\nu\rho}$ is the totally antisymmetric symbol defined by $\epsilon^{012} = +1$. Recall that the mass dimension of e is $2 - d/2$, and that of A is $d/2 - 1$. The combination $A_\mu\partial_\nu A_\rho$ then has mass dimension 2 with $d = 3$, so automatically $[k] = 1 = [e^2]$, exactly what is expected from a conductivity.

Take note that the integrand, i.e. the Lagrangian term, is not gauge invariant. The transformation $A_\mu \to A_\mu + \partial_\mu\alpha$ yields

$$S_{CS} \to S_{CS} + \frac{k}{4\pi}\epsilon^{\mu\nu\rho}\int d^3x \partial_\mu(\alpha\partial_\nu A_\rho).$$

Usually, integrals of total derivatives may be taken to be zero by the divergence theorem, but this is not necessarily the case here. The problem is that α is not necessarily single valued [64], it is only defined up to a shift $n2\pi$, $n \in \mathbb{Z}$.

The reader is referred to chapter 5 of [43] and chapter 2 of [64] for the full argument and calculation of the transformation properties of S_{SC}. Here, the result is merely quoted:

$$S_{SC} \to S_{CS} + \frac{2\pi k}{e^2}. \tag{3.6.5}$$

This looks bad, but remember that the action only enters the QFT formalism through the generating functional

$$Z[A] = \int \mathcal{D}A e^{iS_{CS}},$$

and so the generating functional, and by extension the physics it describes, is indeed gauge invariant on condition that $\frac{k}{e^2} \in \mathbb{Z}$, that is, $k = e^2\nu$, with $\nu \in \mathbb{Z}$. Now, restoring SI units, we get

$$k = \frac{e^2}{\hbar}\nu,$$

with $\nu \in \mathbb{Z}$.

Finally, the current can be found:[11]

$$j^\sigma = \frac{\delta S}{\delta A_\sigma} = \frac{k}{2\pi}\epsilon^{\sigma\nu\rho}\partial_\nu A_\rho, \tag{3.6.6}$$

specifically

$$j_y = \frac{e^2}{h}\nu E_x,$$

gives the correct value for σ_{xy} in the IQHE.

The FQHE may also be explained with an effective action, by adding another $U(1)$ gauge field with a similar Lagrangian term to the action. See e.g. [64] for this construction.

3.6.5 The Effective Low Energy Lagrangian

Topological theories are powerful, and well accepted in condensed matter physics today. The fact that the conductivity quantization is protected by a topological invariant has deep consequences for our understanding of the QHE. It explains why relatively simple models (i.e. non-interacting electrons on a lattice) yields such accurate results about the QHE: the measured quantities are given in terms of the topology of the problem, and are invariant to "perturbations", such as material specifications and impurities.

The fact that one model, specified by the Chern-Simons action, may explain the QHE as a whole near the plateaux rings the familiar bell of universality. As is illustrated in Fig. 3.4.2, the temperature is a natural scaling parameter: it is in the limit $T \to 0$ one gets the flat plateaux and discontinuous jumps between

[11] See [43] for this derivation. There is a non-trivial factor of 2 arising from a counterintuitive integration-by-parts.

them. It is then natural to identify T with the scaling parameter and the plateaux $(\sigma_H, \sigma_D) = (e^2\nu/h, 0)$ with the attractive fixed points.

We would like to study the flow of the QHE in parameter space, as parametrized by T. The problem with doing this in the framework of Chern-Simons theory, however, is exactly the argument that allowed us to construct the Chern-Simons action in the first place: it is already completely renormalized. This means that it yields accurate results in the vicinity of any given fixed point (plateau), but can not be used to study transitions between attractive fixed points, or flows further away from fixed points.

Thus, Chern-Simons theory cannot be used in constructing β-functions for the QHE. However, one does not always need to know a specific action or Lagrangian in order to put constraints on β-functions if dualities are present, as discussed in section 2.10. The preceding discussion on topological effects does, however, motivate the following general structure of the low-energy effective, but not completely renormalized, Lagrangian (as suggested in [65]):

$$\mathcal{L} = \sigma_H \mathcal{L}_{\text{top}} + \sigma_D \mathcal{L}_{\text{kin}},$$

with $\mathcal{L}_{\text{top}}$ being a topological coupling term and $\mathcal{L}_{\text{kin}}$ a kinetic term. This means that the conductivities should be viewed as the parameters of the theory, and we should study RG flow of the parameter $\sigma = \sigma_H + i\sigma_D$, or equivalently $\rho = \rho_H + i\rho_D$.

3.7 Universality and Duality in the Quantum Hall Effects

To further motivate the use of the word "universality" in the context of the QHE, we must turn to experimentally observed values of critical exponents. Another experimental fact of the QHE is the so-called semi-circle law: it should also be taken into account in determining the structure of the parameter space. These are the themes reviewed in this section, before we round of this chapter with the dualities that are to be expected in the QHE parameter space, and the restrictions they put on the β-functions.

3.7.1 Critical Exponents

As was rewieved in chapter 2, equality of critical exponents classify systems into universality classes. Even though the many quantum Hall effects can be qualitatively described by a single model, equality of the critical exponents provides the true justification of an RG analysis of the QHEs.

The critical exponent most often quoted in QH experiments is called κ, which can be found in two ways [66]:

$$\left(\frac{\partial \rho_H}{\partial B}\right) \propto T^{-\kappa} \quad \text{or} \quad \Delta B \propto T^{\kappa}, \tag{3.7.1}$$

ΔB denoting the distance between two extrema in $(\partial \rho_D / \partial B)^{-1}$. Recalling the terminology of subsection 3.3.6, both $\left(\frac{\partial \rho_H}{\partial B}\right)$ and ΔB represent measures for the range of delocalized states at temperature T [67]. Another critical exponent sometimes qouted in the QHE is ν, the critical exponent for the correlation length, defined in section 2.3. In these systems, the exponents ν and κ are also related through $\kappa = p/2\nu$, p being the critical exponent for inelastic scattering length, or quantum coherence length: $L_\Phi \propto T^{-p/2}$. ν is the exponent usually calculated in numerical approximations, and p is generally accepted to be 2, although some discrepancies exists [68]. In any case, κ is by far the most used critical exponent experimentally, and as such it is the one quoted here.

Table 3.1 shows the experimentally obtained critical exponent κ for a variety of transitions, including two fractional transitions, in different materials. The most often qouted value of $\kappa = 0.42$ is within the experimental error of all the reported values included here. Although the universal nature of κ has been brought into question by several authors (see e.g. [66], [69] and [75]), table 3.1 indicates that a large subset of the QHEs should belong to a single universality class. Furthermore, each transition has a similar critical exponent, hence dualities should exist between each plateau, fractional and integer. That is, all plateaux of the QHE should be in the same universality class.

3.7.2 The Semi-Circle Law

Another interesting experimental feature of the QHEs is the semi-circle law. It states, quite simply, that plotting the measured data points (σ_H, σ_D), or equivalently (ρ_H, ρ_D), as parametrized by the applied magnetic field or back-gate voltage, traces out semi-circles in parameter space. The situation is depicted in Fig. 3.7.1. The semi-circles in the IQHE of that experiment are centered on the horizontal σ_H-axis at half-integer values of e^2/h, with radii $e^2/2h$, such that the endpoints are on plateau values.

Even though Fig. 3.7.1 depicts the situation for the full integer sequence, semi-circles with endpoints at the critical points emerge in every IQHE, with increased precision for $T \to 0$. As such, the dualities of the QHE should respect semi-circles in general.

System [References]	$1 \to 0$	$2 \to 1$	$3 \to 2$	$4 \to 3$	$5 \to 4$	$6 \to 5$	$\pm 6 \to \pm 2$	$\frac{2}{5} \to \frac{1}{3}$	$\frac{2}{3} \to \frac{3}{5}$
InGaAs/InP [67]	0.42(4)	-	-	-	-	-	-	-	-
$Al_xGa_{1-x}As$/ $Al_{0.33}Ga_{0.67}As$* [69]	-	-	0.42(1)	0.42(1)	0.42(1)	0.42(1)	-	-	-
$Al_xGa_{1-x}As$/ $Al_{0.33}Ga_{0.67}As$** [70]	-	-	-	0.42(1)	-	-	-	-	-
HgTe Quantum Well [71]	-	0.45(4)	0.40(2)	-	-	-	-	-	-
Graphene*** [72]	-	-	-	-	-	-	0.41(4)/ 0.37(5)	-	-
AlGaAs-GaAs [73][74]	-	-	-	-	-	-	-	0.43(2)	0.38(4)

Table 3.1: Experimentally estimated values for critical exponent κ in the QHE. Columns correspond to the respective transition the exponent was obtained for.
*x ranges from 0.65% to 1.6%. It is worth noting that in this experiment, values for κ up to 0.57 was obtained for doping concentrations outside this range.
**$x = 0.85\%$.
*** Different values using $\partial\rho_H/\partial B$ and ΔB respectively.

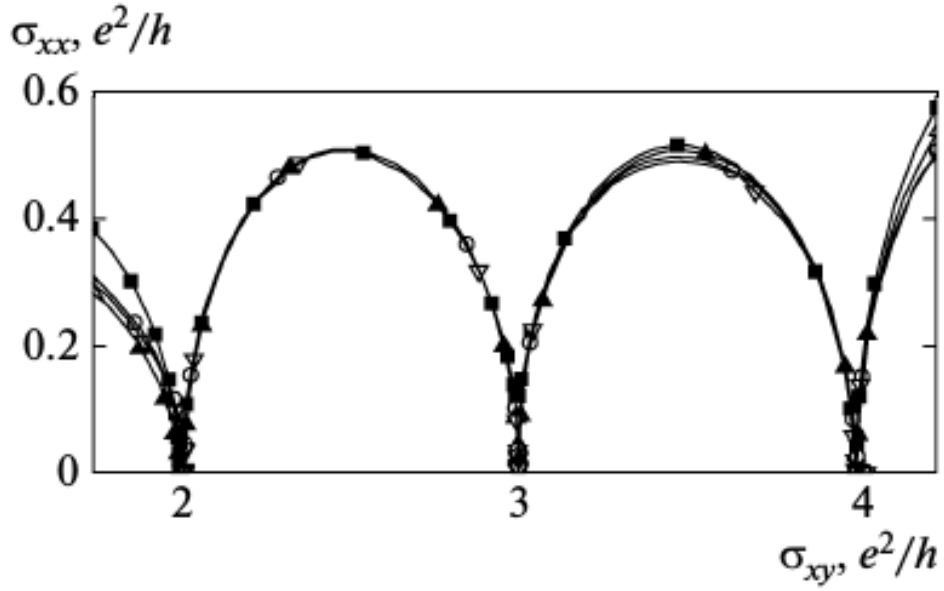

Figure 3.7.1: Example of the observed semi-circle law in the IQHE in an InGaAs/GaAs structure. The curve $(\sigma_H(B), \sigma_D(B))$ traces out a semi-circle of radius $e^2/2h$. Modified from [76].

3.7.3 Dualities for the Quantum Hall Effect

We now have sufficient information about the QHE to make a prediction about the β-functions. Here, we start with the "simplest" case of dualities. Later, it will become apparent how they must be altered. From now on, we will work in units $e^2/h = 1$, for simplicity. The parameter space with coordinates (σ_H, σ_D) will be termed $\mathbb{H}$, for reasons which will soon be revealed. We will model $\mathbb{H}$ in complex coordinates, as explained in subsection 3.2.3, and write $\sigma = \sigma_H + i\sigma_D$.

Firstly, the simplest integer effect: a completely non-degenerate QH system, exhibiting the full integer sequence. Then, the fixed points in $\mathbb{H}$ are located at $\sigma = \nu$, where $\nu \in \mathbb{Z}$. Since every fixed point belongs to the same universality class, one duality should be a simple translation in σ_H:

$$T : \sigma \mapsto \sigma + 1. \tag{3.7.2}$$

The form of the translation generalizes naturally to the other integer effects. Even in the fractional case, as depicted in Fig. 3.5.1, a translation of 1 should be included, e.g. in mapping the plateaux $1/3 \mapsto 4/3$ and $2/3 \mapsto 5/3$. Demanding translational invariance, we get the following restriction on the β-function, $\beta(\sigma) = \beta^1(\sigma) + i\beta^2(\sigma)$:

$$\beta(\sigma + 1) = \beta(\sigma). \tag{3.7.3}$$

Periodicity alone is not a strong condition to put on the β-function, but combined with another duality, namely *inversion*, it produces very strict constraints. The existence of an inversion duality can be motivated from the fact that the system is equally well described in the conductivities σ and the resistivities ρ, related by $\rho = -\sigma^{-1}$.[12] That is, we study the β-functions transformation properties under

$$S : \sigma \mapsto -\sigma^{-1}. \tag{3.7.4}$$

From eqs. (2.10.3) and (3.2.12), we get the following constraint on β:

$$\begin{aligned}
\beta(-1/\sigma) &= \left(\frac{\partial \rho_H}{\partial \sigma_H}\beta^1 + \frac{\partial \rho_H}{\partial \sigma_D}\beta^2\right) + i\left(\frac{\partial \rho_D}{\partial \sigma_H}\beta^1 + \frac{\partial \rho_D}{\partial \sigma_D}\beta^2\right) \\
&= \frac{(\sigma_H^2 - \sigma_D^2)\beta^1 + 2\sigma_H\sigma_D\beta^2}{(\sigma_H^2 + \sigma_D^2)^2} + i\frac{-2\sigma_D\sigma_H\beta^1 + (\sigma_H^2 - \sigma_D^2)\beta^2}{(\sigma_H^2 + \sigma_D^2)^2} \\
&= \frac{1}{|\sigma|^4}(\sigma_H^2 - \sigma_D^2 - i2\sigma_H\sigma_D)(\beta^1 + i\beta^2) \\
&= \sigma^{-2}\beta(\sigma).
\end{aligned} \tag{3.7.5}$$

This is really a relation between the β-functions of ρ and σ:

$$\beta(\rho) = \frac{d\rho}{dt} = \frac{\partial \rho}{\partial \sigma^j}\frac{d\sigma^j}{dt} = \sigma^{-2}\beta(\sigma),$$

however, we will hypothesize that the relation (3.7.5) is valid for the β-function of the σ-flow on its own. That is, in units $e^2/h = 1$, $\beta(-1/\sigma) = \sigma^{-2}\beta(\sigma)$ with $-1/\sigma$ being a new value of conductivity.[13]

Now, conditions (3.7.3) and (3.7.5) *together* are very strong. They, together with the condition that β is holomorphic, imply that β is a *modular form* of weight -2.

The condition that β is divergence free, and thus holomorphic, is an *ansatz*. The next chapter will explore the geometric constraints that should be put on the QHE parameter space in order to justify this assumption. Then, the theory of modular forms will be briefly reviewed to show how the β-function may be uniquely determined (up to an overall constant factor) from this condition.

[12] Compare this with the Wannier-Kramers duality of the 2-dimensional Ising model, which maps a low temperature model to a high temperature model.

[13] This assumption is not very well motivated, and will indeed not hold. It is, however, a first step towards studying modular mathematics. Further motivation for doing so will be presented in subsection 4.4.1.

Chapter 4

Geometry of the Quantum Hall Parameter Space

4.1 Introduction

This chapter will review mathematics that is relevant to this thesis, but in no way included in a standard physics education. We start by introducing the *Poincaré half plane*, $\mathbb{H}$, a hyperbolic geometric space which will be identified with the parameter space of the quantum Hall effect. From there, we move on to the *modular group*, the group generated by the transformations T and S of the last section, and incidentally, a subgroup of the symmetry group of $\mathbb{H}$. We will be particularly interested in the maximal subgroups of the modular group, which are examples of *congruence groups*, and holomorphic functions transforming covariantly under their group action. Finally, these results are combined with those of Chapters 2 and 3 to make predictions about the renormalization group flow of the quantum Hall effect.

The first connection between physics and the modular group seems to have been drawn by Cardy in 1982, in studying phase diagrams for Abelian lattice gauge theories with a "θ-parameter" [77]. In 1989, a connection with the QHE was made by Shapere and Wilczek [78].

The idea of rigorously including *emergent modular symmetries* in the QHE was orginally conjectured by Lütken and Ross in 1992 [65], and further developed by, among others, Lütken, Ross, Burgess, Dolan and Nissinen (see e.g. [79], [80], [81], [82], [83], [84]). The aim of this chapter is to properly motivate all assumptions that must be made for including modular mathematics in the QHE, and to arrive at many of the same results as first presented by the authors mentioned.

4.2 The Poincaré Half Plane

The discussion of this section mostly follows [85] and [86]. Certain aspects are also found in [28], which provides a rather dense introduction.

The Poincaré half plane, $\mathbb{H}$, is a model of 2-dimensional hyperbolic geometry. For a definition of the term "hyperbolic geometry", and the pure mathematical reason for studying it, see Appendix D. This section reviews the aspects of the half plane model important for computation. The upper half plane is defined as the set of complex numbers above the real axis:

$$\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}, \tag{4.2.1}$$

with the line element

$$ds^2 = \frac{dx^2 + dy^2}{y^2}, \tag{4.2.2}$$

and hence metric tensor

$$G_{ij} = \frac{1}{y^2}\delta_{ij}. \tag{4.2.3}$$

Note that in this metric, the entire real line is infinitely far away from any point in $\mathbb{H}$. The real line thus constitutes a "line of infinity".

4.2.1 Lines in $\mathbb{H}$

Now, straight lines (geodesics) of $\mathbb{H}$ may be found from the geodesic equation (see Appendix B). The connections of $\mathbb{H}$ are found from the metric tensor to be $\Gamma^2_{22} = \Gamma^1_{12} = \Gamma^1_{21} = -\Gamma^2_{11} = -\frac{1}{y}$, and the rest zero. For a curve $\gamma(t) = (x(t), y(t))$, the geodesic equation then reads

$$\ddot{x} - \frac{2\dot{x}\dot{y}}{y} = 0 \tag{4.2.4}$$

$$\ddot{y} + \frac{\dot{x}^2 - \dot{y}^2}{y} = 0. \tag{4.2.5}$$

Putting x =constant, the first equation is satisfied trivially, and second is separable:

$$\frac{\ddot{y}}{\dot{y}} = \frac{\dot{y}}{y},$$

yielding

$$\dot{y} \propto y,$$

so that

$$\gamma(t) = (a, be^{ct}),$$

for $a, b, c \in \mathbb{R}$ are geodesics in $\mathbb{H}$. That is, straight lines parallell to the imaginary axis. The second set of solutions is found by first noting that eq. (4.2.4) is separable:

$$\begin{aligned} \frac{1}{\dot{x}}\frac{d\dot{x}}{dt} &= \frac{2}{y}\frac{dy}{dt} \\ \dot{x} &= cy^2. \end{aligned} \tag{4.2.6}$$

Substituting into eq. (4.2.5) yields

$$\ddot{y} + c^2y^3 - \dot{y}^2/y = 0 \tag{4.2.7}$$

Next, assume the geodesics are of unit speed, i.e.

$$G_{ij}\dot{\gamma}^i\dot{\gamma}^j = \frac{1}{y^2}(\dot{x}^2 + \dot{y}^2) = 1. \tag{4.2.8}$$

This is done withour loss of generality as geodesics of other speeds may be found from reparametrization (see e.g. chapter 9, problem 27 of [87]). Substituting for $\dot{x}$ in eq. (4.2.5) using eq. (4.2.8) yields

$$\ddot{y} + y - \frac{2\dot{y}^2}{y} = 0. \tag{4.2.9}$$

Equating eqs. (4.2.7) and (4.2.9) finally gives

$$\begin{aligned} y &= c^2y^3 - \dot{y}^2/y, \\ \dot{y} &= y\sqrt{1 - c^2y^2}, \end{aligned}$$

a separable equation with solution

$$\ln\left(\frac{yb}{\sqrt{1 - c^2y^2} + 1}\right) = t,$$

for some positive constant b. Solving for y gives

$$y = \frac{2b}{b^2e^{-t} + c^2e^t}.$$

As b, c must be positive, they may be expressed as $b = e^B$, $c = e^C$ for some $B, C \in \mathbb{R}$. Multiplying numerator and denominator by $e^{-(B+C)}$ then allow us to write

$$y(t) = e^{-C}\text{sech}(t), \tag{4.2.10}$$

after a constant shift in t. Integrating eq. (4.2.6) then yields

$$x(t) = e^{-C}\tanh(t) + A, \tag{4.2.11}$$

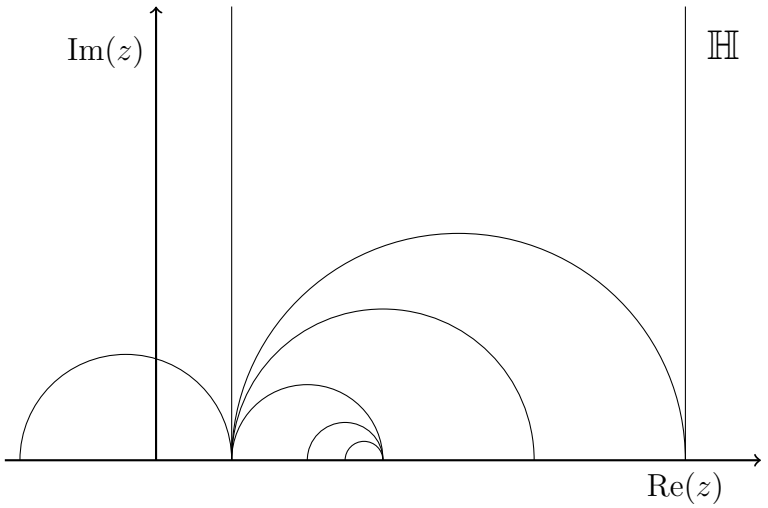

Figure 4.2.1: A selection of straight lines in $\mathbb{H}$. Every line is infinitely long. All lines displayed are parallell to one another as they at most have points in common at infinity (the real line).

for some $A \in \mathbb{R}$. The other family of geodesics, $(x(t), y(t))$ for $t \in (-\infty, \infty)$, are thus semi-circles with origin on the x-axis. Finding two classes of solutions to the system of second order differential equations spans the solution space for this problem: the semi-circles centered on the real line and lines parallell to the imaginary axis from the real line to imaginary infinity are all the lines in $\mathbb{H}$. Fig. 4.2.1 shows a selection of straight lines in $\mathbb{H}$.

Note that given a $\mathbb{H}$-line l, and a point $z \in \mathbb{H}$ not contained in l, there exists infinitely many lines through z not intersecting l. This is in sharp contrast to Euclidean geometry where only one such line (the line through z parallell to l) may be found.

4.2.2 Congruence in $\mathbb{H}$

For further insight into the workings of $\mathbb{H}$, we need to know the symmetry group of the space. This is the group of *isometries*: distance and angle preserving transformations of the space. Under its group action, geometric objects mapping onto each other will be called *congruent*. For example, in Euclidean space, the symmetry group consists of translations and rotations.

Naturally, the group of isometries of $\mathbb{H}$ will be a subgroup of the group of automorphisms of the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, that is, the conformal group

introduced in section 2.9:

$$PSL(2,\mathbb{C}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a,b,c,d \in \mathbb{C}, ad-bc=1 \right\} /\mathbb{Z}_2,$$

where each matrix corresponds to a unique fractional linear transformation:

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \leftrightarrow \gamma(z) = \frac{az+b}{cz+d},$$

which acts on a set of complex numbers.

The problem with the group $PSL(2,\mathbb{C})$ is that it contains elements that do not conserve $\mathbb{H}$. For instance, the transformation $z \to i/(iz) = \bar{z}/|z|^2$ would take an element of $\mathbb{H}$ outside of $\mathbb{H}$. To constrain this group to only map $z \in \mathbb{H}$ as $z \mapsto \frac{az+b}{cz+d} \in \mathbb{H}$, amounts to restraining the coefficients in the FLT to be real.[1] The symmetry group of $\mathbb{H}$ is then

$$PSL(2,\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a,b,c,d \in \mathbb{R}, ad-bc=1 \right\} /\mathbb{Z}_2. \tag{4.2.12}$$

For FLTs given by $PSL(2,\mathbb{R})$, observe that

$$\mathrm{Im}(z) \mapsto \frac{\mathrm{Im}(z)}{|cz+d|^2}, \tag{4.2.13}$$

hence, the sign of $\mathrm{Im}(z)$ is conserved, so any element of $\mathbb{H}$ is mapped back to $\mathbb{H}$.

As already mentioned in section 2.9, conformal transformations preserve angles. Furthermore, the conformal transformations map circles on $\overline{\mathbb{C}}$ to circles on $\overline{\mathbb{C}}$ [85]. The fact that the elements of $PSL(2,\mathbb{R})$ maps $\mathbb{H}$-lines to $\mathbb{H}$-lines then follows from the fact that the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ itself is a $\overline{\mathbb{C}}$-circle and all $\mathbb{H}$-lines are subsets of $\overline{\mathbb{C}}$-circles orthogonal to $\overline{\mathbb{R}}$. Proving that the length of line segments is invariant under $PSL(2,\mathbb{R})$-transformations is now an algebraic exercise. We will here merely state that this is true.[2]

4.3 The Quantum Hall Symmetry Groups

This section marks the starting point for our exploration of modular mathematics - a field not commonly known to physicists. Good primary sources are [88], [89], [90], and the classic [91], which may be the one most relevant to this discussion.

[1] See e.g. [85] for a formal proof.

[2] The metric function of $\mathbb{H}$ may indeed by *defined* by demanding that lengths are measured along $\mathbb{H}$-lines and that two segments are of equal length if and only if an element of $PSL(2,\mathbb{C})$ maps them onto eachother. See [85] for this construction.

4.3.1 The Modular Group

As shown in the previous chapter, we are looking for β-functions which transforms in a specific manner under

$$T : z \mapsto z + 1, \qquad S : z \mapsto -\frac{1}{z}.$$

The transformations T and S do not commute:

$$TS(z) = T\left(-\frac{1}{z}\right) = \frac{z-1}{z} \tag{4.3.1}$$

$$ST(z) = S(z+1) = -\frac{1}{z+1}, \tag{4.3.2}$$

and as such, they generate an *infinite discrete group*. As previously done with general FLTs, we may make the identifications

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{4.3.3}$$

in order to work with the matrix representation of the group.

The group generated by T and S is called *the modular group* and may be expressed as (see e.g. [90]):

$$SL(2,\mathbb{Z}) = \langle T, S\rangle = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a,b,c,d \in \mathbb{Z},\ ad - bc = 1 \right\}, \tag{4.3.4}$$

as usual, to get rid of sign degeneration, we define $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\mathbb{Z}_2$. Also observe that as elements of $PSL(2,\mathbb{Z})$,

$$S^2 = (ST)^3 = I, \tag{4.3.5}$$

I denoting the identity. These relations are in fact the *defining relations* for $PSL(2,\mathbb{Z})$, in the sense that any two generators, T, S, obeying eq. (4.3.5) must generate a group isomorphic to $PSL(2,\mathbb{Z})$ [90].

Similarly to the group action of $PSL(2,\mathbb{C})$ on the Riemann sphere, the modular group also acts on the compactified Poincaré half plane $\overline{\mathbb{H}}$. A transformation $\gamma(z) = \frac{az+b}{cz+d}$ with $c \neq 0$ takes

$$\gamma : -\frac{d}{c} \mapsto \infty, \qquad \gamma : \infty \mapsto \frac{a}{c}, \tag{4.3.6}$$

so every real fraction is equivalent to the point ∞ under the group action of $PSL(2,\mathbb{Z})$. This motivates the following form of the compactification of $\mathbb{H}$ (see section 1.2 of [89]):

$$\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}. \tag{4.3.7}$$

This form of $\overline{\mathbb{H}}$ will be referred to as the *physical compactification* of $\mathbb{H}$.[3]

4.3.2 Congruence Groups

As will be shown, the entire modular group is too large, in the sense that demanding the β-functions to transform under the entire $SL(2,\mathbb{Z})$ (or $PSL(2,\mathbb{Z})$) is too restrictive. We will then turn our attention to the maximal subgroups of $SL(2,\mathbb{Z})$, called the *congruence groups*. As these groups are contained within the modular group, they impose softer conditions on the β-functions.

Following chapter 1 of [89], define the *principal congruence group of level N* to be

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \;\middle|\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\mathrm{mod} N) \right\},$$

where $(\mathrm{mod}N)$ works elementwise on the matrix. Recall that $x \equiv y\ (\mathrm{mod}N)$ if $x = aN + y$ with $a \in \mathbb{Z}$ and $y < N$ (see e.g. [92]). To save space on notation, writing $x \equiv_N y$ will be taken to mean $x \equiv y\ (\mathrm{mod}N)$. Note the special case $\Gamma(1) = SL(2,\mathbb{Z})$.

Now, a subgroup $\Gamma \subset SL(2,\mathbb{Z})$ is called a *congruence group at level N*, if $\Gamma(N) \subset \Gamma$ for an $N \in \mathbb{N}$. Following [91], there are exactly four groups between $SL(2,\mathbb{Z})$ and $\Gamma(2)$. Defining

$$P = ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \qquad R = TST = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

they are:

$$\Gamma_T = \{\gamma \in SL(2,\mathbb{Z}) \mid \gamma \equiv_2 I \text{ or } \gamma \equiv_2 T\} \tag{4.3.8}$$
$$\Gamma_S = \{\gamma \in SL(2,\mathbb{Z}) \mid \gamma \equiv_2 I \text{ or } \gamma \equiv_2 S\} \tag{4.3.9}$$
$$\Gamma_R = \{\gamma \in SL(2,\mathbb{Z}) \mid \gamma \equiv_2 I \text{ or } \gamma \equiv_2 R\} \tag{4.3.10}$$
$$\Gamma_P = \{\gamma \in SL(2,\mathbb{Z}) \mid \gamma \equiv_2 I \text{ or } \gamma \equiv_2 P \text{ or } \gamma \equiv_2 P^2\}. \tag{4.3.11}$$

Of these groups, Γ_T, Γ_R and Γ_S have index 3 in $SL(2,\mathbb{Z})$, while Γ_P has index 2.[4] Indeed, Γ_T, Γ_S and Γ_R, are furher related by conjugation:

$$\Gamma_S = P^{-1}\Gamma_T P, \qquad \Gamma_R = P^{-2}\Gamma_T P^2, \tag{4.3.12}$$

[3]This terminology will be justified shortly. Note, however, that this version of $\overline{\mathbb{H}}$ is qualitatively different from the one obtained by considering the continuation of all possible $\mathbb{H}$-lines, and subsequently uniting $\mathbb{R}$, instead of $\mathbb{Q}$, to $\mathbb{H} \cup \{\infty\}$.

[4]See Appendix A for the definition of the term "index" in group theory. What this essentially means is that Γ_P is larger than Γ_X for $X = T, S, R$.

where the notation means $x^{-1}Ax = \{x^{-1}ax \mid a \in A\}$. This can be seen from the fact that

$$P^{-1}TP = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \equiv S \text{ (mod2)}$$
$$P^{-2}TP^2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \equiv R \text{ (mod2)}.$$

The groups may also be expressed in terms of their generators as:

$$\Gamma_T = \langle T, R^2 \rangle \tag{4.3.13}$$
$$\Gamma_S = \langle S, T^2 \rangle \tag{4.3.14}$$
$$\Gamma_R = \langle R, T^2 \rangle \tag{4.3.15}$$
$$\Gamma_P = \langle P, S^{-1}PS \rangle, \tag{4.3.16}$$

these results are summarized in table 3 of [91].[5] The formal proofs of these expressions are rather complicated, but note that they can be easily motivated from the original definitions of groups. Take the definition (4.3.8) of Γ_T: T itself is obviously a generator, and the simplest non-trivial element of $SL(2,\mathbb{Z})$ which is $\equiv_2 I$, and constructed from T and S, is $R^2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. The generators for Γ_R and Γ_S may then be found from conjugation.

For future reference, $\Gamma(2)$ may also be expressed in terms of its generators T^2 and R^2:

$$\Gamma(2) = \langle T^2, R^2 \rangle.$$

4.3.3 Modular Correspondance

There is another group, not included in $SL(2,\mathbb{Z})$, which will be of interest. The motivation for including this group will be presented later, but it is included in the discussion here for completeness. Note first that Γ_T is related to Γ_R by scaling z by a factor of 2,[6] and let

$$G = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

[5] Note that [91] uses a different notational convention than here. Also, there the second generator for Γ_T is quoted as PSP^{-1}. However, $TPSP^{-1} = R^2$, which is easier to work with.

[6] This observation is most easily made in the graphic representations of the phase diagrams presented below.

Conjugating the generators of Γ_T by G yields

$$GTG^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = T^2$$
$$GR^2G^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = R,$$

the generators for Γ_R. We thus have

$$\Gamma_R = G\Gamma_T G^{-1}. \tag{4.3.17}$$

This relation is termed a *modular correspondance*. We get a similar relation conjugating the generators of Γ_S:

$$GT^2G^{-1} = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} = T^4$$
$$GSG^{-1} = \begin{pmatrix} 0 & -2 \\ 1/2 & 0 \end{pmatrix} = Q,$$

where the last equality defines Q. Note that $Q \notin SL(2,\mathbb{Z})$, but the matrix group generated by Q and T^4 will still be intimately related to Γ_S. We term this group, obtained from the modular correspondance with Γ_S, Γ_Q:

$$\Gamma_Q = G\Gamma_S G^{-1} = \langle Q, T^4 \rangle. \tag{4.3.18}$$

4.4 Construction of the QHE Parameter Space

4.4.1 The Motivation of Shapere and Wilczek

As mentioned, Shapere and Wilczek came up with a motivation for studying the modular group in the context of the QHE in their 1989 paper [78]. The choice was made to include this motivation here, after the introduction of the modular group and its subgroups. This is because prior knowledge of these groups will help the reader understand why the results presented here are interesting.

Recall first Laughlin's argument for the FQHE (subsection 3.5.1). It predicts plateaux for $\nu = 1/m$ with m an odd number. Following [78], given Laughlin's trial wavefunction valid at some filling factor ν, there are two canonical ways of constructing an approximate ground state wavefunctions at other filling factors. The first, we will not go into detail on, but the result yields a state with filling factor $\nu' = 1 - \nu$. Note that this method is only valid in the limit of infinite magnetic field.[7]

[7]See the original paper [78] and the references therein for the details.

The other way of obtaining a new wavefunction is to multiply the old wavefunction by a factor $\prod_{i<j}(z_i - z_j)^{2p}$, with p an integer. This yields a state with filling factor $\nu' = (\nu^{-1} + 2p)^{-1}$. There are then two generators of symmetry transformations of the filling factor:

$$\nu \to 1 - \nu,$$
$$\nu^{-1} \to \nu^{-1} + 2.$$

It is not obvious that these are related to $SL(2,\mathbb{Z})$, but introducing the variable $\alpha = \nu^{-1} - 1$, we get

$$\alpha \to 1/\alpha,$$
$$\alpha \to \alpha + 2,$$

the generators of Γ_S, up to a sign change in the inversion. This is the motivation described in [78].

The problematic first transformation, $\nu \to 1 - \nu$, is not contained in $SL(2,\mathbb{Z})$. However, as mentioned, that transformation is only valid in the limit $B \to \infty$. Focusing instead on the second transformation, $\nu^{-1} \to \nu^{-1} + 2$, observe that it is equivalent to

$$\nu \to \frac{\nu}{2\nu + 1},$$

which is the transformation R^2 described in the previous section. Now, experimentally, the FQHE is observed in conjunction with the integer sequence $\nu = 1, 2, 3, 4...$ (compare Fig. 3.5.1). In that spirit, we may hypothesize a translational symmetry $\nu \to \nu + 1$. As discussed in subsection 3.7.3, all integer plateaux belong to the same universality class, so translational symmetry makes sense. In that case, the symmetry group acting on ν is $\langle T, R^2 \rangle = \Gamma_T$.

Laughlin's theory of the FQHE is approximative in nature, so this discussion does by no means present evidence that modular symmetries should be apparent in the QH parameter space. It does, however, show some motivation for further studying the effect of enforcing *exact* symmetry under $SL(2,\mathbb{Z})$, or its subgroups.

4.4.2 Geometric Identification

Now, $PSL(2,\mathbb{Z}) \subset PSL(2,\mathbb{R})$, so every element of the modular group is also a symmetry transformation of $\mathbb{H}$. As we wish to study functions with special transformation properties under $PSL(2,\mathbb{Z})$, it makes sense to consider them defined on a space where the geometry is invariant under $PSL(2,\mathbb{Z})$.

That is, we identify the Quantum Hall parameter space of coordinates $\sigma = \sigma_H + i\sigma_D$ with $\mathbb{H}$.[8] This way, parameter flow driven by varying the magnetic field follows "straight lines" (compare the semi-circle law, Fig. 3.7.1). Also, the attractive fixed points of the RG flow are identified with the plateaux in the QHE, and as such, all attractive fixed points lie on the real line (Fig. 3.7.1).[9] Similarly, we assume that all repulsive fixed points of the flow lie on the real line. With the metric given on $\mathbb{H}$, the consequence is that every fixed point that is not a saddle point, is infinitely far away from any given point of the flow, and the assumption that β is divergence free may be taken seriously.

4.4.3 Fixed Point Structure

As all plateaux of the QHEs lie on fractions, the *physical compactification* of $\mathbb{H}$ defined above, $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$, is indeed the set of all physically relevant points. Furthermore, the point ∞ should be taken to be a repulsive (UV) fixed point of the parameter flow for σ [84]. This makes sense as the longitudinal conductivity always approaches a finite value in the low-temperature limit. Similarly, ∞ may be an attractive (IR) fixed point for ρ in systems exhibiting an *insulator phase*, the zero plateau: $\sigma_D = \sigma_H = 0$, by the duality transformation $\rho = -1/\sigma$.

The classification of ∞ as a repulsive fixed point in the σ-flow for all QHEs allows for some interesting analysis. By the relations (4.3.6), all fractions are equivalent to ∞ under $SL(2,\mathbb{Z})$. This means that if $SL(2,\mathbb{Z})$ is the symmetry group of the QHE, all fixed points (that are not saddle points) are repulsive. This makes RG-flow quite difficult to achieve, and is the first step towards considering subgroups of the modular group, rather than the full group. The nail in the coffin for $SL(2,\mathbb{Z})$ is presented in the next section.

Considering the congruence groups, recall first the definition (4.3.8) of Γ_T. For any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_T$, the definition implies $c \equiv_2 0$ and $a \equiv_2 1 \equiv_2 d$. So the fractions $-d/c,\ a/c$, equivalent to the repulsive ∞, always has the structure (odd integer)/(even integer). In turn, this means that all fractions of the form even/odd, and odd/odd are attractive fixed points. In particular, as 1 is an odd number, every integer is an attractive fixed point under Γ_T. This means that every integer plateau is classified as an attrictive fixed point by Γ_T, so Γ_T is a candidate symmetry group for the full (spin-polarized) integer sequence $\nu = 0, \pm 1, \pm 2, ...$

[8] If this assumption seems forced, keep in mind that there is no mathematical reason to assume Euclidean over hyperbolic geometry for an arbitrary abstract space. See Appendix D for a justification of this statement.

[9] Note that the semi-cirle law is a *well* established feature of the QHE [93].

Similarly, using the definition of Γ_R (4.3.10), $a \equiv_2 1 \equiv_2 d$ still, but $c \equiv_2 0$ or $c \equiv_2 1$. The repulsive fixed points for Γ_R is thus on the form odd/even and odd/odd, while the attractive are even/odd. This means that Γ_R includes the spin-degenerate sequence $\nu = 0, \pm 2, \pm 4, ...$

For Γ_S (4.3.9), we have $a \equiv_2 0 \equiv_2 d$ and $c \equiv_2 1$, or $a \equiv_2 1 \equiv_2 d$ and $c \equiv_2 0$. This means all fractions on the form even/odd and odd/even are repulsive, while the ones on the form odd/odd are attractive. This gives the integer sequence $\nu = \pm 1, \pm 3, \pm 5...$, which, to the author's knowledge, has not been observed. However, this implies for Γ_Q the integer sequence $\nu = \pm 2, \pm 6, \pm 10, ...$, exactly the spin and valley degenerate sequence observed for graphene.

Finally, by the same analysis for Γ_P, all fractions are again repulsive, and cannot account for any integer sequence by our assumptions. For this reason, Γ_P will not be considered any further.

4.5 Modular Forms

The interesting groups have been defined, and it is time to take a look at functions with special transformation properties under their action. Again, [88], [89], [90] and [91] are all great primary sources for the mathematical theory reviewed here. In the following $\gamma \in \Gamma$, for any subgroup $\Gamma \subseteq SL(2, \mathbb{Z})$, will always mean

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{or} \quad \gamma : z \to \frac{az + b}{cz + d}.$$

4.5.1 Modular Forms at Level 1

Let $k \in \mathbb{Z}$. A holomorphic function $f : \mathbb{H} \to \mathbb{C}$ is a *modular form of weight k* if it is holomorphic at ∞, and for all $z \in \mathbb{H}$, and every $\gamma \in SL(2, \mathbb{Z})$, we have

$$f(\gamma(z)) = (cz + d)^k f(z). \tag{4.5.1}$$

In fact [89], it is enough to check this criterion for the *generators* of $SL(2, \mathbb{Z})$, which then amounts to

$$f(z + 1) = f(z), \qquad f(-1/z) = z^k f(z). \tag{4.5.2}$$

Observe that these are exactly the conditions for the β-functions, eqs. (3.7.3) and (3.7.5), proposed in section 3.7.3, with $k = -2$. The requirement of analyticity at ∞ makes sure that the vector space spanned by modular forms of the same weight is finite-dimensional [89]. It is, however, also a natural constraint on the β-functions, as ∞ is considered a fixed point of the RG flow.

Note that the infinitesimal (form) dz transforms as

$$dz \to d\gamma(z) = \frac{d\gamma}{dz}dz = (cz+d)^{-2}dz. \tag{4.5.3}$$

Then, if $f(z)$ is a modular form of weight 0, its derivative

$$\frac{df}{dz} \to (cz+d)^2\frac{df}{dz}, \tag{4.5.4}$$

is a modular form of weight 2.

Now, the constant function mapping all $z \mapsto 0$, is a modular form every weight. There are no non-trivial modular forms of odd weight, as choosing $\gamma = -I$, gives $f(z) = (-1)^k f(z)$.

For non-trivial examples of modular forms, the prototype is given by the *Eisenstein series of weight k*, the 2-dimensional analogue of the Riemann zeta function: $\zeta(k) = \sum_{n=1}^{\infty} 1/n^k$ (see e.g. chapter 1 of [89]):

$$E_k(z) = N_k \sum_{(c,d)\neq(0,0)} \frac{1}{(cz+d)^k}, \quad z \in \mathbb{H}, \tag{4.5.5}$$

where every pair of integers except (0,0) is summed over. The normalization factor, N_k, is chosen so that $\lim_{\mathrm{Im}(z)\to\infty} E_k(z) = 1$, and is given by the Riemann zeta function, evaluated at k: $N_k = 2\zeta(k)$.

However, this construction only works for even $k > 2$. For $k = 2$, the sum is not absolutely convergent [89], and for $k < 0$, it is not obviously not holomorphic at ∞.

Indeed, let $\mathcal{M}_k(SL(2,\mathbb{Z}))$ denote the vector space spanned by modular forms of weight k. Theorem 3.5.2 of [89] tells us that for $k < 4$, $\mathcal{M}_k(SL(2,\mathbb{Z})) = \{0\}$. This means, quite simply, that the functions we are looking for do not exists. In order to keep working within this mathematical framework, the modular forms for the congruence groups are considered.

4.5.2 Modular Forms at Level 2

Similar to the discussion of the last subsection, define a modular form of weight k for the congruence group Γ to be a holomorphic function $f : \mathbb{H} \to \mathbb{C}$, holomorphic at ∞, such that for every $z \in \mathbb{H}$ and all $\gamma \in \Gamma$

$$f(\gamma(z)) = (cz+d)^k f(z).$$

It is again sufficient to check the transformational abilities of f for the generators of Γ [89].

For subgroups of $\Gamma(1)$, modular forms of weight 2 exists. Indeed, $\dim[\mathcal{M}_2(\Gamma(2))] = 2$ [94]; there are exactly two linearly independent modular forms of weight 2 at level 2.

For the congruence groups Γ_T, Γ_R, Γ_S, there are still no modular forms of negative weight, however, there exists exactly *one* modular form of weight $k = 2$ for each of them. Now, the *physical*, contravariant, β-function for the QHE parameter space is expected to be a modular form of weight $k = -2$, but lowering the index by the Poincaré metric:

$$\beta_i = G_{ij}\beta^j,$$

we obtain a new function:

$$\beta^{\text{mod}}(\sigma, \bar{\sigma}) = \frac{1}{\sigma_D^2}\beta^{\text{phys}}(\sigma, \bar{\sigma}), \tag{4.5.6}$$

in coordinates $\sigma = \sigma_H + i\sigma_D$. Recall from section 2.10 that the physical β-function is expected to be anti-holomorphic in σ. Under a modular transformation, it will then be expected to transform as

$$\beta(\sigma, \bar{\sigma}) \to (c\bar{\sigma} + d)^{-2}\beta(\sigma, \bar{\sigma}).$$

The transformational property of β^{mod} is then (recall eq. (4.2.13)):

$$\begin{aligned}\beta^{\text{mod}} = \frac{1}{\sigma_D^2}\beta^{\text{phys}} &\to \frac{|c\sigma + d|^4}{\sigma_D^2}(c\bar{\sigma} + d)^{-2}\beta^{\text{phys}} \\ &= (c\sigma + d)^2\frac{1}{\sigma_D^2}\beta^{\text{phys}} \\ &= (c\sigma + d)^2\beta^{\text{mod}},\end{aligned} \tag{4.5.7}$$

so β^{mod} transforms exactly like a modular form of weight 2. Note that the construction with the Poincaré metric, $1/\sigma_D^2$, necessarily breaks the holomorphic properies of exactly one of β^{phys}, β^{mod}. We will assume that β^{mod} is the holomorphic one, as this implies that it is a weight 2 modular form.[10] Also, the assumption that the β-function is divergence free means that it can be written as the gradient of a potential function. As β^{mod} is the covariant (lower-index) function, it will be directly related to the potential: $\beta_i = \partial_i\Phi$, and as such may be taken to be more fundamental than β^{phys} [82].

The physical β-function is then found from the relation (4.5.6), and the problem is reduced to a possible one: finding the modular forms of weight 2 for the groups Γ_T, Γ_R, Γ_S.

[10] Here we started from the assumption that β^{phys} is holomorphic to conclude that it is not holomorphic. This may seem to be an uncomfortable situation, but the holomorphic properties of the β-functions, even though they are well motivated, are *assumptions*. Here we have merely switched to a slightly different *ansatz*: that β^{mod} has holomorphic properties rather than β^{phys}.

4.5.3 The Theta Functions

This has already been done, in general terms in [82], and in greater detail in [79] and [84]. The construction is based on the Jacobi theta-functions. In their most general form, they may be defined as:

$$\Theta_{\mu,\nu}(\tau|z) = \sum_{n=-\infty}^{\infty} e^{i\pi(n+\frac{1}{2}\mu)^2 z + i\nu\pi\tau}, \quad \tau \in \mathbb{C},\ z \in \mathbb{H},\ \mu,\nu \in \mathbb{R}.$$

The modular nature of $\Theta_{\mu,\nu}(\tau|z)$ is explored in [28] and [91]. Let $q = e^{i\pi z}$ for $z \in \mathbb{H}$ and from Θ, define three new functions:

$$\theta_2(z) = \Theta_{1,0}(0|z) = \sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}, \tag{4.5.8}$$

$$\theta_3(z) = \Theta_{0,0}(0|z) = \sum_{n\in\mathbb{Z}} q^{n^2}, \tag{4.5.9}$$

$$\theta_4(z) = \Theta_{0,1}(0|z) = \sum_{n\in\mathbb{Z}} (-1)^n q^{n^2}. \tag{4.5.10}$$

Now, θ_2, θ_3 and θ_4 are close to being modular forms of weight 1/2 for Γ_T, Γ_S and Γ_R, respectively. This would mean the theta functions to the power of four might be good candidates for building modular forms of weight 2. Performing this construction directly is rather difficult, the simplest way to go about it is first finding weight 0 forms. The derivative of the weight 0 form will then give the weight 2 form (subsection 4.5.1).

From chapter 7 of [91], the theta functions have the following transformation properties under T and S:

$$\begin{aligned} \theta_2(z+1) &= e^{i\pi/4}\theta_2(z), & \theta_2(-1/z) &= (-iz)^{1/2}\theta_4(z), \\ \theta_3(z+1) &= \theta_4(z), & \theta_3(-1/z) &= (-iz)^{1/2}\theta_3(z), \\ \theta_4(z+1) &= \theta_3(z), & \theta_4(-1/z) &= (-iz)^{1/2}\theta_2(z). \end{aligned} \tag{4.5.11}$$

From these we can calculate

$$\theta_4(Rz) = \theta_4(TSTz) = (-i(z+1))^{1/2}\theta_4(z), \tag{4.5.12}$$

and

$$\theta_2(R^2 z) = \theta_2(TST^2STz) = (2z+1)^{1/2}\theta_2(z). \tag{4.5.13}$$

So the theta functions to the power of four have the following transformation properties under the generators of Γ_T, Γ_S and Γ_R, respectively:

$$\begin{aligned} \theta_2^4(Tz) &= -\theta_2^4(z), & \theta_2^4(R^2z) &= (2z+1)^2\theta_2^4(z), \\ \theta_3^4(T^2z) &= \theta_3^4(z), & \theta_3^4(Sz) &= -z^2\theta_3^4(z), \\ \theta_4^4(T^2z) &= \theta_4^4(z), & \theta_4^4(Rz) &= -(z+1)^2\theta_4^4(z). \end{aligned} \tag{4.5.14}$$

These are exactly the transformation properties required up to some annoying minuses. That is, the θ_i^4 are not modular forms, however certain combinations of them are. Define

$$\lambda(z) = \frac{\theta_2^4(z)}{\theta_3^4(z)}. \tag{4.5.15}$$

Using the relation [91]

$$\theta_3^4(z) = \theta_2^4(z) + \theta_4^4(z), \tag{4.5.16}$$

along with the other transformation abilities for the θ_i, we get:

$$\begin{aligned}
\lambda(Tz) &= \frac{\theta_2^4(Tz)}{\theta_3^4(Tz)} = \frac{\lambda(z)}{\lambda(z)-1}, \\
\lambda(Sz) &= \frac{\theta_2^4(Sz)}{\theta_3^4(Sz)} = 1 - \lambda(z), \\
\lambda(Rz) &= \lambda(TSTz) = \frac{1}{\lambda(z)},
\end{aligned}$$

which in turn gives

$$\begin{aligned}
\lambda(T^2 z) &= \lambda(z), \\
\lambda(R^2 z) &= \lambda(z).
\end{aligned}$$

So λ is a modular form of weight 0 for $\Gamma(2)$. Similarly, modular forms of weight 0 can be constructed for Γ_T, Γ_S and Γ_R from λ. From [91] they are

$$\varphi_T = \frac{\lambda - 1}{\lambda^2}, \tag{4.5.17}$$

$$\varphi_S = \lambda(\lambda - 1), \tag{4.5.18}$$

$$\varphi_R = -\frac{\lambda}{(1-\lambda)^2}. \tag{4.5.19}$$

Their invariance under the action of the corresponding generators is readily checked. Since φ_T, φ_S and φ_R are modular forms of weight 0 for Γ_T, Γ_S and Γ_R, respectively, their derivatives will be modular forms of weight 2.

Lastly, note that from the transformational properties (4.5.14), knowing that that there are two linearly independent modular forms of weight 2 for $\Gamma(2)$, we may take $\mathcal{M}_2(\Gamma(2)) = \mathrm{Span}\{\theta_2^4, \theta_3^4\}$. Of course, any two of the three θ_i^4 will work as they are related by eq. (4.5.16).

4.5.4 The Quantum Hall β-Functions

The potential functions φ_X, $X = T, S, R$ are *unique* in the sense that the vector space of modular forms of weight 0 for each Γ_X is 1-dimensional [94]. Similarly,

the vector space of weight 2 modular forms for each Γ_X is also 1-dimensional [91]. However, as φ_X is invariant under Γ_X, any weight 2 modular form multiplied by any power of φ_X will still be weight 2 modular form.

To uniquely determine the β-functions, there is an additional assumption that must be made about the potential. A reasonable condition to put on the potential is that it is real valued. This will necessarily mean that we have to work with the absolute value $|\varphi_X|$. To keep holomorphic properties, we further demand that the potential factorizes into a holomorphic and an antiholomorphic term. A more refined argument for this assumption is found in [84]. Under these assumptions, the simplest candidate for a potential function is

$$\Phi_X(\sigma, \bar{\sigma}) = \ln |\varphi_X|^2 = \ln \varphi_X(\sigma) + \ln \bar{\varphi}_X(\bar{\sigma}), \tag{4.5.20}$$

and the β-functions are given by (up to a normalization constant)

$$\beta_X(\sigma) \propto \frac{\partial \Phi_X}{\partial \sigma} = \frac{\varphi_X'(\sigma)}{\varphi_X(\sigma)}.$$

For $X = T, S, R$, this gives [84]

$$\beta_T \propto \frac{\lambda'}{\lambda}\frac{2-\lambda}{1-\lambda} \propto 1 + 24\sum_{n=0}^{\infty}\frac{nq^n}{1+q^n}, \tag{4.5.21}$$

$$\beta_S \propto \frac{\lambda'}{\lambda}\frac{1-2\lambda}{1-\lambda} \propto 1 - 24\sum_{n=0}^{\infty}\frac{(2n+1)q^{2n+1}}{1+q^{2n+1}}, \tag{4.5.22}$$

$$\beta_R \propto \frac{\lambda'}{\lambda}\frac{1+\lambda}{1-\lambda} \propto 1 + 24\sum_{n=0}^{\infty}\frac{nq^{2n}}{1+q^{2n}}, \tag{4.5.23}$$

where $q = \exp(i\pi\sigma)$.

Also, as $\Gamma_Q = G\Gamma_S G^{-1}$, with $G : z \to 2z$, the β-function corresponding to Γ_Q is given by

$$\beta_Q(\sigma) = \beta_S(G^{-1}\sigma) = \beta_S(\sigma/2). \tag{4.5.24}$$

4.5.5 The Family of Potential Functions

In subsection 4.4.3, the completely undegenerate plateau sequence $\nu = 0, \pm 1, \pm 2, ...$ was linked to Γ_T, and the spin degenerate sequence $\nu = 0, \pm 2, \pm 4, ...$ was linked to Γ_R. As spin-polarization may be tuned by an applied back gate voltage, it may in principle be possible to tune a system from being Γ_T to Γ_R symmetric. Assuming the system keeps $\Gamma(2)$ symmetry during the transition, the simplest parametrization for the RG-potential is given by (see e.g. [84])

$$\Phi_a = \ln[\lambda(\lambda-1)^{a-1}] + \text{c.c.}, \tag{4.5.25}$$

c.c. denoting the complex conjugate expression and $a \in \mathbb{R}$. This gives a family of C-functions, where the β-functions are given as

$$\beta_a(\sigma) \propto \frac{\partial \Phi_a}{\partial \sigma} = \frac{\lambda'}{\lambda}\frac{1-a\lambda}{1-\lambda} = i\pi(\theta_3^4 - a\theta_2^4), \tag{4.5.26}$$

the last equality obtained by using the relation $\lambda'/\lambda = i\pi\theta_4^4$ [91]. Observe that (up to normalization) $a = 1/2$ gives β_T and $a = -1$ gives β_R. The parametrization also includes β_S for $a = 2$. These values for a will be termed *maximally symmetric*. Also note that for certain values of a, the β-function degenerates to one of the theta-functions: $a = 0$ gives θ_3^4, $a = 1$ gives θ_4^4, and the limit $a \to \infty$ gives θ_2^4. Sadly, this potential family cannot include β_Q.

The parameter a is believed to be some (possibly complicated) function of the Zeeman energy and gate voltage, however it may well depend on other parameters $(B, T, g^*$, material constants,...) as well.

4.6 The QHE Phase Diagram

4.6.1 Maximally Symmetric Phase Diagrams

In subsection 4.4.3, the structure of attractive and repulsive fixed points for each of the congruence groups were determined, and in subsection 4.5.4, their respective β-functions were found. The complete phase diagram for the quantum Hall effect may now be constructed.

Fig. 4.6.1 displays the phase diagram in parameter space for QHEs exhibiting symmetry under the congruence groups and Γ_Q. Here, we define *phase boundaries* to distinguish regions of the phase diagram flowing towards distinct attractive fixed points when temperature is lowered. The *separatrices* distinguish regions of the phase diagram originating from distinct repulsive fixed points.

As such, separatrices connect attractive fixed points and phase boundaries connect repulsive fixed points. The shapes of the separatrices and phase boundaries are determined by the β-functions, and they are indeed $\mathbb{H}$-lines. Which $\mathbb{H}$-lines constitutes relevant separatrices and phase boundaries is determined by the β-functions according to the locations of the saddle points: the points in $\mathbb{H}$ where β vanishes. With the parametrization of the β-functions given by eq. (4.5.26), these points are given by [84] (see also Appendix C of [94]).

$$\sigma_\otimes = \frac{iK(1-1/a)}{K(1/a)}, \tag{4.6.1}$$

and every point equivalent to $\sigma_\otimes$ under the relevant symmetry group. The specific saddle point found by eq. (4.6.1) is termed the *seed*. K is the complete elliptic integral of the first kind:

$$K(x) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - x^2 \sin^2 \theta}},$$

and its values may be found numerically. At a saddle point, a phase boundary and a separatrix meet in a 90° angle. Locations and shapes of the phase boundaries and separatrices may of course also be found by studying a plot of the flow.

For the σ-flow, the large separatrix distinguishing flows originating at ∞ from those originating at $\mathbb{Q}$ will be termed the *enveloping separatrix*.

Note that in the σ-plane, Γ_R and Γ_Q are essentially obtained from doubling Γ_T and Γ_S, respectively, and vice versa in the ρ-plane. The ρ-plane is obtained from the σ-plane, simply by inversion: $\rho = -1/\sigma$.

4.6.2 The Family of Phase Diagrams

To go one step further, we can by similar methods determine a phase diagram for each value of a in the one-parameter family β_a. For each a, saddle points may be found by eq. (4.6.1). The enveloping separatrix may be found from the RG-flow through points sufficiently close to the saddle point. This has been done in Fig. 4.6.2, which displays the enveloping separatrix and the phase boundary intersecting it at 90° for $a \in [-1.2, 2.2]$.

The insets displays the specific cases of maximal symmetry:

- $a = -1$: Γ_R, for which $\beta = \theta_2^4 + \theta_3^4$,
- $a = 1/2$: Γ_T, for which $\beta = \theta_3^4 + \theta_4^4$,
- $a = 2$: Γ_S, for which $\beta = \theta_4^4 - \theta_2^4$,

and maximal degeneration:

- $a = 0$: $\beta = \theta_3^4$,
- $a = 1$: $\beta = \theta_4^4$.

The latter cases are termed "degenerate" as the seed saddle point approaches $\mathbb{Q}$ in this case, giving rise to a very complicated flow.

At the time of conception, the idea of representing the family of phase diagrams as in Fig. 4.6.2 was thought to be original. It is, however, worth noting that

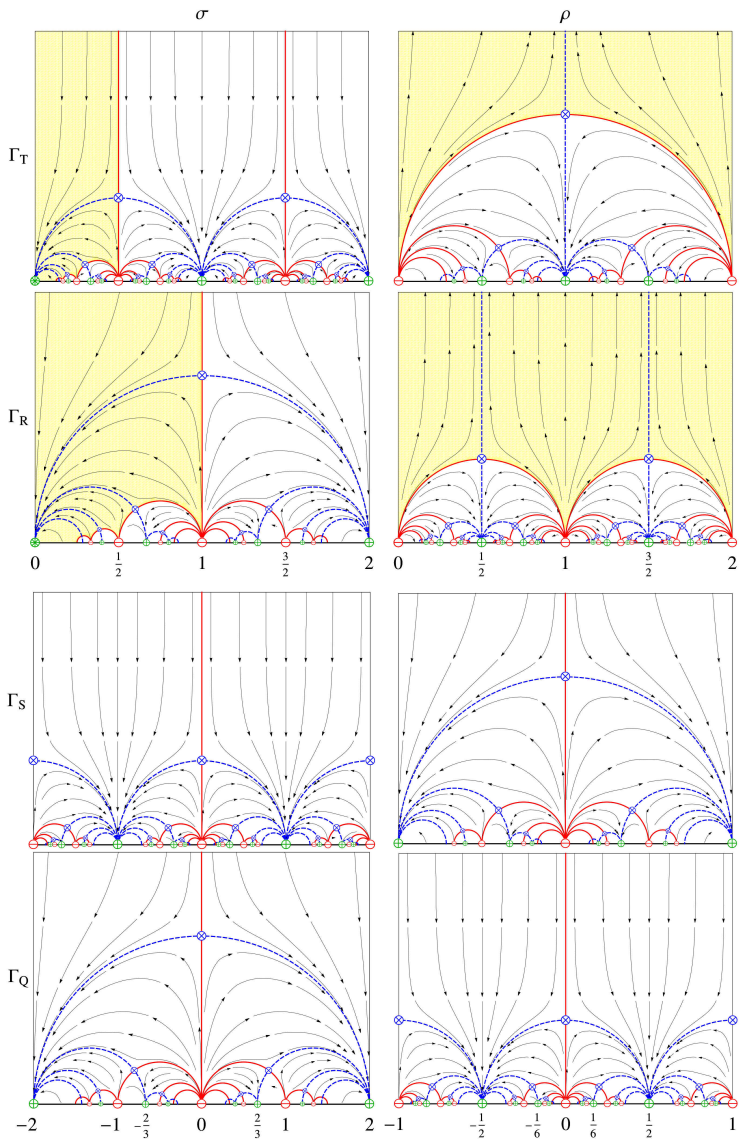

Figure 4.6.1: Phase diagrams for the QHEs exhibiting symmetry under the congruence groups and Γ_Q. The horizontal axis represents σ_H (ρ_H) and the vertical σ_D (ρ_D). $\oplus$, $\ominus$ and $\otimes$ shows attractive, repulsive and saddle fixed points, respectively. $\circledast$ represents the QH insulator where it exists, and the colored background represents the insulator phase. Dashed lines marks separatrices and solid lines phase boundaries. The arrows shows the parameter flow according to the β-function, when temperature is lowered. See the text for further explanation. The figure is generated by C. A. Lütken. See Fig. 2 of the appended paper (section 5.3) for a color image.

similar ideas can be traced back to 1993 [95]. Compare Fig. 4.6.2 with Fig. 4.6.3. There, the interpolation between Γ_R and Γ_T has been parametrized by an

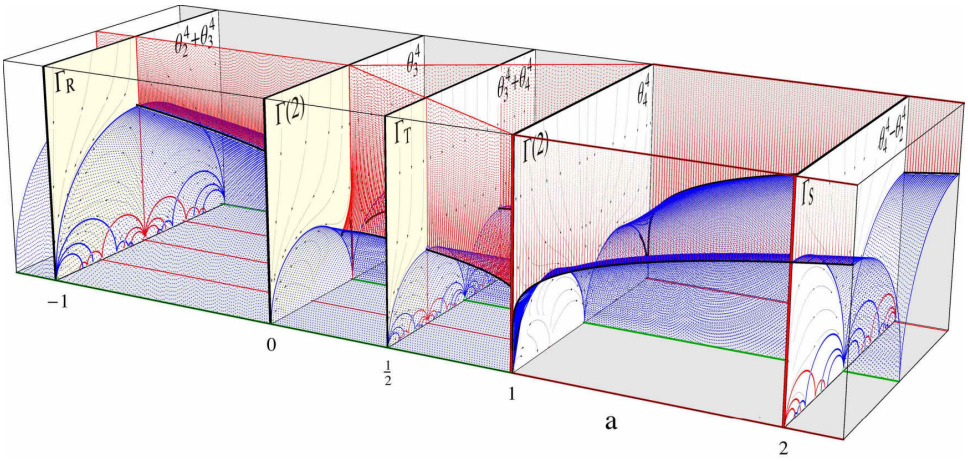

Figure 4.6.2: The family of σ-flow phase diagrams for the QHE parametrized by the real variable a in the region $-1.2 < a < 2.2$. Displayed is the enveloping separatrix and the phase boundary crossing it. The seed saddle point is found at their intersection. The figure is generated by C. A. Lütken, see Fig. 4 of the appended paper (section 5.3) for a color image.

effective g-factor, which can be thought of as $g_{\text{eff}} \sim (g/\cos\theta)T^{-\kappa}$, measuring the "effective spin polarization". κ is the scaling exponent introduced in subsection 3.7.1, and θ is the angle betweeen the sample surface normal and the applied magnetic field. See [95] for the details.

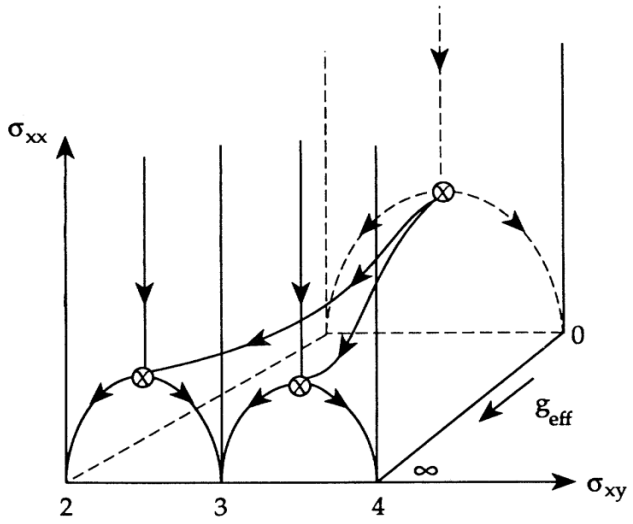

Figure 4.6.3: An early attempt at parametrizing the transition between Γ_R and Γ_T symmetry. In this representation, an effective g-factor given as $g_{\text{eff}} \sim (g/\cos\theta)T^{-\kappa}$ is taken to be the parameter interpolating between the two β-functions. See the text for further explanation. The units are $e^2/h = 1$ on the σ-axes. The figure is from [95].

Chapter 5

Experimental Verification

In this chapter, the hypothesis of modular symmetries in the quantum Hall effect is tested. This is done by sampling data points from published experimental work in the QHE and test how well they fit into the phase diagrams presented in the previous chapter. The method will be more in greater detail shortly, after a quick summary of the assumptions that were made in order to produce the phase diagrams. The results of the search for experimental verification (or invalidation) are presented in the paper written in collaboration with K. S. Olsen and C. A. Lütken.

5.1 Summary of Assumptions

The aim of this thesis has been to create a comprehensive introduction to renormalization group flows and the Quantum Hall effect, and to motivate the relevance of modular symmetry in this context. It is, however, important to keep track of exactly what assumptions have been made along to way. These are summarized here, before the results are presented.

The assumptions all concern the geometrical structure of the β-function, and are naturally divided into two categories: (1) holomorphic structure and (2) symmetries.

Assumption 1.1: *β is curl free on $\mathbb{H}$.*

> This is required by the C-theorem, and by the interpretation of the RG flow in the following sense. A non-zero curl would allow for limit cycles, that is, closed curves as the limit of the RG flow. As RG flows amounts to integrating out degrees of freedom, this behaviour is prohibited.

Assumption 1.2: *β is divergence free on $\mathbb{H}$.*

> This assumption amounts to placing all sources of the RG flow on the boundary of $\mathbb{H}$, i.e. $\mathbb{Q} \cup \{\infty\}$. The sinks are the plateaux, which automatically are placed on $\mathbb{Q}$ by the nature of the QHE. The sources, however, suffers no such restriction. For instance, Laughlin *et. al.* proposed a theory in 1985, placing the sources of QH RG flow in the interior of $\mathbb{H}$ [96].

In totality, assumption 1 amounts to β having a holomorphic structure on $\mathbb{H}$. Note that this is a *global* constraint, valid everywhere on $\mathbb{H}$. This is in sharp contrast to the general discussion of β-functions, which is usually restricted to the vicinity of a fixed point, where perturbation theory is valid. The β-functions found here are exact; valid *everywhere* in the phase diagram of the QHE.

Holomorphic structure is in itself a very restrictive condition, and it is difficult to motivate from theory alone. However, the assumptions are ultimately validated by the fact that they have rigid consequences that, as demonstrated in the following paper, agree with virtually all experimental data so far. The restrictive predictions of the β-functions reaches their full potential when complex analyticity is combined with our symmetry constraints, also motivated by QH data.

Assumption 2.1: *β is a modular function.*

> This means that the β-function transforms covariantly under the action of a congruence subgroup, Γ, of the modular group $SL(2,\mathbb{Z})$, or the group Γ_Q obtained by a modular correspondance.

Assumption 2.2: The symmetry group Γ contains $\Gamma(2)$.

> Combined with the above assumptions, this assumption means that Γ is a congruence group at level 2, and effectively constrains the β function to belong to the one-parameter family described by Φ_a of subsection 4.5.5. The main reason for this assumption is that it severely constrains the number of possible β-functions.

This is the minimal set of assumptions that has to be made about the β-functions. Note that we have also made the implicit assumption that the metric on parameter space (the Zamolodchikov metric) is of the simplest kind: diagonal in a real basis, or anti-diagonal in a holomorphic basis. In geometry, the canonical metric is the Poincaré metric, but more general metrics could in principle be considered. The exact form of the metric affects the value of the critical exponents near the saddle points of the flow, so this hypothesis can be tested exprimentally. This is, however, beyond our current scope.

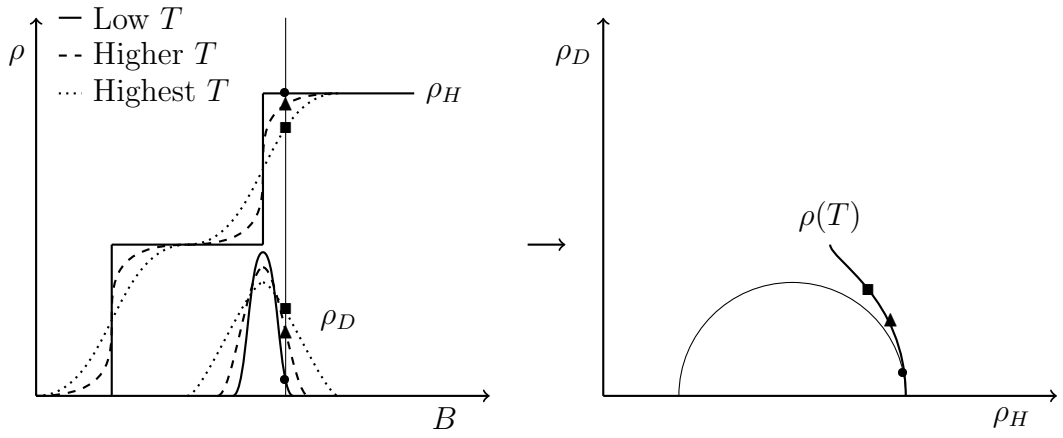

Figure 5.2.1: Illustration of the data sampling procedure and subsequent fitting into a phase diagram. Each plot $\rho_H(B)$ and $\rho_D(B)$ is made at constant temperature. Sampling is done at constant magnetic field (gate voltage) to obtain the temperature dependence. Subsequently, a flow line is fitted to the sampled data. See the text for further explanation.

5.2 Method

The method for experimentally testing the above assumptions is outlined here.

We need to find published experimental data of the QHE for several temperatures, for the same sample, in the same interval of magnetic field (or gate voltage). The data is usually presented in the form of plots, driven by magnetic field (gate voltage). The required kind of experimental data is illustrated in Fig. 5.2.1.

A set of resistivity (or conductivity) data is then sampled at constant magnetic field for several temperatures. Depending on the plateau structure of the data, a suitable β-function and phase diagram is chosen, according to the classification from subsection 4.4.3.

Next, a flow line, $\rho(T)$, is fitted to the data set by numerically solving the coupled set of differential equations

$$\dot{\rho} = \frac{d\rho}{dT} = \beta(\rho) \Leftrightarrow \left\{ \begin{array}{rl} \dot{\rho}_H = & \mathrm{Re}[\beta(\rho_H, \rho_D)] \\ \dot{\rho}_D = & -\mathrm{Im}[\beta(\rho_H, \rho_D)] \end{array} \right\}, \qquad \rho(T_0) = \rho_0. \qquad (5.2.1)$$

The initial condition $\rho(T_0) = \rho_0$ is chosen such that the resulting flow line $\rho(T)$ lies close to all data points in the same sampled set.

Our task has been to trawl through every relevant experimental paper written about the QHE, dating back to at least 2012,[1] digitize the data presented in

[1]It was assumed that previous workers have thoroughly analysed the experimental data

experimental plots, and to compare them with the theoretical flow lines in a suitable phase diagram. This is by no means a small task, not least because the relevant experiments are scattered throughout the vast literature on the QHE. The results of this endeavour are presented in the paper found in the following section.

5.3 Paper

The paper is currently under peer review for publication in *Physical Review B*. It incorporates all relevant data we could find, resulting in a total of 1484 sampled experimental data points compared to 282 modular flow lines, generated from the one-parameter RG potential Φ_a.

Take care not to confuse the references embedded in the paper with the references used elsewhere in this thesis. Also note that the page numbering follows the page numbering within the paper, not within the thesis.

In the paper, the notation $\mathbb{C}^+$ is used for the upper half plane $\mathbb{H}$. All conductivity plots are in the natural unit e^2/h, and all resistivity plots are in the natural unit $h/e^2 \approx 26\text{k}\Omega$.

published earlier than 2012. The paper does also include older experimental data.

[16:08:22 August 11, 2017]

On the universality of modular symmetries in two-dimensional magnetotransport

K.S. Olsen, H.S. Limseth and C.A. Lütken
Department of Physics, University of Oslo

We analyze experimental quantum Hall data from a wide range of different materials, including semiconducting heterojunctions, thin films, surface layers, graphene, mercury telluride, bismuth antimonide, and black phosphorus. The fact that these materials have little in common, except that charge transport is effectively two-dimensional, shows how robust and universal the quantum Hall phenomenon is. The scaling and fixed point data we analyze appear to show that magnetotransport in two dimensions is governed by a small number of universality classes that are classified by *modular* symmetries, which are infinite discrete symmetries not previously seen in Nature. The Hall plateaux are (infrared) stable fixed points of the scaling-flow, and quantum critical points (where the wave-function is delocalized) are unstable fixed points of scaling. Modular symmetries are so rigid that they in some cases fix the global geometry of the scaling flow, and therefore predict the exact location of quantum critical points, as well as the shape of flow lines anywhere in the phase-diagram. We show that most available experimental quantum Hall scaling data is in good agreement with these predictions.

PACS numbers: 23.23.+x, 56.65.Dy
Keywords: Quantum Hall effect; modular symmetry; scaling and phase diagrams; quantum critical points

I. INTRODUCTION

The continuous and discrete symmetries observed in Nature may be exact or approximate. The continuous case includes exact symmetries like Lorentz and gauge invariance, which severly constrains possible dynamical models, while discrete symmetries usually are finite and approximate. We shall here investigate a class of experimental data that appear to respect a new type of symmetry that is called *modular*. Although these are finitely generated approximate (emergent) discrete symmetries, because they are non-abelian and infinite they provide unusually strong constraints on low-energy model building.

Infinite discrete groups, including modular symmetries, play an important role in modern mathematics, but because they are extremely rigid it is not clear if they can exist in the real physical world. Indeed, it is only in bespoke physical systems ("designer universes"), engineered to be effectively (for all practical purposes) two-dimensional, that modular symmetries have been found.[1–26]

The quantum Hall effect (QHE) appears in materials where charge carriers are forced to move in a single atomic plane, for example on the surface of a crystal or in a sheet of graphene. Experiments measuring the electromagnetic properties (magnetotransport) of Hall-systems produce what at first sight appears to be an impenetrable morass of data. But first appearances can be misleading, and if the quantum Hall data is viewed from a particularly advantageous vantage point a hidden pattern of great beauty and utility is revealed.[1,2] This rigid emergent order is encoded in a fractal phase-diagram tightly harnessed by a modular symmetry that allows it to teeter on the brink of chaos, without actually taking the leap.

Our purpose here is to explore the robustness and universality of these new symmetries, by comparing and contrasting data from the most disparate materials available. We do this in the simplest possible way, by superimposing scaling data directly onto mathematical diagrams with modular symmetry. This "phenomenological" approach is unbiased, since no theoretical assumptions are invoked, and we are free to represent (plot) the data in any way we want. We will not here discuss theoretical ideas that are needed in order to connect the well-known microphysics ("electrons in a dirty lattice") to the emergent macrophysics observed in transport experiments.

Since modular mathematics is unfamiliar to most physicists, a brief introduction to modular symmetry in physics is provided in the next section. In order to motivate this, we start by summarizing the main conclusion: *the scaling properties of a quantum Hall system appears to have a remarkable simplicity and universality encoded in a modular symmetry.*

More precisely, the Hall and magnetoconductivities σ_H and σ_D show a strong dependence on the dominant scale parameter t (usually temperature), and the *scaling functions* $\beta_H = d\sigma_H/dt$ and $\beta_D = d\sigma_D/dt$ appear to be harnessed by a modular symmetry, which is so rigid that they only barely survive:

(i) if the symmetry observed in an experiment is one of the maximal subgroups of the modular group, then *the physical β-function is unique*, up to an overall normalization (compare Fig. 2).

(*ii*) if the symmetry observed in an experiment is reduced to the largest subgroup shared by the maximal subgroups, then *there is a unique family of β-functions parametrized by a single real number*, up to an overall normalization (compare Fig. 4).

These all but unique β-functions generate equally unique flow diagrams, which we compare with scaling data. They almost always agree, within experimental error (compare Figs. 6-17).

This transmutation of modular mathematics into quantum Hall physics follows from a fundamental property of scaling functions: *they must respect any geometric structure with which the parameter space is endowed.* In the QHE this is the space of transport coefficients (conductivities or resistivities), which appears to be equipped with both a *complex structure* and an *emergent modular symmetry.* These circumstances conspire to give a very strong constraint on low-energy physics, and any model of this physics, as we now explain in the context of the QHE.

It is convenient to combine the real conductivities σ_H and σ_D into a complex quantity $\sigma = \sigma_H + i\sigma_D$ that takes values in the upper half of the complex plane: $\sigma \in \mathbb{C}^+$ ($\Im\sigma = \sigma_D > 0$). This is useful because it reduces (transport) matrix operations to ordinary (complex) algebra. But it is much more than that, because the flow for which the scaling function $\beta = d\sigma/dt = \beta_H + i\beta_D$ is a tangent vector field is:

(1) *divergence free* (no sources or sinks on $\mathbb{C}^+$),

(2) *curl free* (a curl would render the physical interpretation of β-functions meaningless),

(3) *covariant under a modular symmetry* $\Gamma \subset \mathrm{SL}(2,\mathbb{Z})$,

(4) *finite on $\mathbb{C}^+$ except for isolated simple zeros* (quantum critical points are regular saddle points).

(1) and (2) is equivalent to the Cauchy-Riemann equations, and it follows that β is a holomorphic function of $\overline{\sigma}$, i.e., anti-holomorphic in σ.[27] Such *Laplacian flows* are automatically gradient flows, i.e., completely determined by a scalar potential φ, which in this context is called the *renormalization group (RG-) potential* or *C-function* (compare eq. (2)). Combining this with (3) and (4) we may conclude that *β is a modular form of weight two.*

It is the paucity of weight two forms on large modular groups that gives modular symmetry extremely sharp teeth. The first useful result is that *no* such forms exist if Γ is the full modular group $\Gamma(1)_0 = \mathrm{SL}(2,\mathbb{Z})$, and there are therefore no candidate β-functions with this symmetry. This provides a theoretical reason, independent of the experimental observation that this symmetry is too large, for considering smaller groups. So we turn our attention to subgroups of $\Gamma(1)_0$, where further surprises await us,[28] including (*i*) and (*ii*).

We shall see that this provides a host of rigid predictions that are easy to falsify. The most surprising consequence of a modular symmetry is perhaps that the plateaux *must* be rational. This follows from the fact that in order for a modular symmetry to act "properly" on the real line (in a strict mathematical sense),[29] the upper half plane $\mathbb{C}^+$ is compactified to the *extended upper half plane* $\overline{\mathbb{C}}^+$ by adding *only* rational numbers, and a *single* point at infinity: $\overline{\mathbb{C}}^+ = \mathbb{C}^+ \cup \mathbb{Q} \cup \{i\infty\}$. It is also appealing that the integer (IQHE) and fractional (FQHE) quantum Hall effects are automatically unified by a modular symmetry.

The mathematical primer in Sect. II is followed by some introductory remarks about the novel materials that have yielded most of the new data discussed in the following sections. They give a fairly comprehensive overview of the current experimental status of the modular hypothesis, including all scaling experiments we have found to be of sufficient quality to enable us to extract a partial flow-diagram. Sects. IV-VII provide what is essentially a catalogue of fixed point data and scaling diagrams, organized by the modular symmetry they exhibit. Within each of these universality classes the data are grouped according to the type of material used in the experiment.

Sect. VIII summarizes previous work and some of the successes of the modular paradigm so far, as well as some of the outstanding problems and challenges to be addressed in future work.

II. MODULAR SYMMETRY

The nested hierarchical structure that is emerging in phase portraits of the QHE (compare Figs. 6-17) is the signature of an approximate global discrete symmetry, which, given some familiarity with modular groups, is surprisingly easy to identify by finding some of the fixed points.

The repulsive ultraviolet (UV) fixed points ($\ominus$) and attractive infrared (IR) fixed points ($\oplus$ = plateaux) of scaling lie on the boundary of parameter space. They are sources and sinks of *flow lines* (streamlines) in a *flow diagram* of the vector field β of scaling functions. The quantum critical points (semi-stable saddle points $\otimes$) all lie in the interior of parameter space. This fixed point structure, which can be extracted directly from the geometry of the data without any theoretical bias, is the DNA of the symmetry, from which all else will follow.

The full modular group $\mathrm{SL}(2,\mathbb{Z}) = \langle T, S\rangle$ can be represented by fractional linear (Möbius) transformations, generated by *translations* $T(z) = z + 1$ and *duality transformations* $S(z) = -1/z$, acting on the upper half of the complex plane $\mathbb{C}^+(z)$. It is the fact that T and S do not commute that makes this group infinite, and interesting. Any "word" in T and S is a fractional linear (Möbius) transformation $\gamma(z) = (az + b)/(cz + d)$, with integer coefficients and unit determinant ($ad - bc = 1$). Words can only be simplified using the "grammatical"

rules $S^2 = 1 = (ST)^3$ that define the abstract group.

As far as the full modular group is concerned, all fractions (plateaux values) are equivalent, so if this were a physically viable symmetry we should observe all possible fractional plateaux. However, we never observe the full set of fractions in any given quantum Hall experiment, but only plateaux (fractions) that satisfy certain constraints on the parities of the numerator, or denominator, or both. These *parity rules*, which depend on the two-dimensional material under consideration, are the key to identifying any would-be modular symmetry. They link microphysics to macrophysics, because the observed spectrum of integer fixed points follows directly from the spectrum of charge carriers supported by the system in the non-interacting limit ("Landau level spectroscopy").

The resistivity $\rho = S(\sigma) = -1/\sigma$ is conveniently given by the modular duality transformation S, since this is equivalent to taking the matrix inverse of the conductivity tensor. Note that it is conventional to choose $\sigma_H = \sigma_{12}$ and $\rho_H = \rho_{21}$ in order to eliminate an annoying minus sign.

A. Hierarchy of symmetries

So the full modular symmetry is too strong for the QHE, but the largest subgroups of $\mathrm{SL}(2,\mathbb{Z})$ are not. A map showing the tip of the modular iceberg, including all the groups we need, is presented in Fig. 1.

Subgroups of the modular group are obtained by relaxing the translation symmetry ($T \to T^n$), or the duality symmetry ($S \to R^n$, where $R(z) = TST(z) = z/(1+z)$), or both. Three of these so-called "congruence subgroups at level two" preserve parities, which means that each of them groups the fractions into two equivalence classes. Because p and q in $\sigma_\oplus = p/q$ are relatively prime, there are only three types of fractions with well defined parities.

With "o" representing odd integers and "e" representing even integers, we have $p/q \in o/o$, o/e or e/o, and it is easy to verify that the equivalence classes are[30]

$$\begin{aligned}
\Gamma_{\mathrm{T}} = \Gamma_{\mathrm{T}}(2)_1 = \langle T, R^2\rangle: &\quad \left\{\frac{e}{o}, \frac{o}{o}\right\}_\oplus \cup \left\{\frac{o}{e}\right\}_\ominus \\
\Gamma_{\mathrm{R}} = \Gamma_{\mathrm{R}}(2)_1 = \langle R, T^2\rangle: &\quad \left\{\frac{e}{o}\right\}_\oplus \cup \left\{\frac{o}{o}, \frac{o}{e}\right\}_\ominus \\
\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{S}}(2)_1 = \langle S, T^2\rangle: &\quad \left\{\frac{o}{o}\right\}_\oplus \cup \left\{\frac{o}{e}, \frac{e}{o}\right\}_\ominus
\end{aligned} \tag{1}$$

A class is indexed by $\oplus$ if the fractions are attractive fixed points of scaling in the σ-plane, and by $\ominus$ if they are repulsive fixed points. This assignment follows from the requirement that the direction of the flow is downward at the top of the conductivity plane, which is a result that can be obtained in a perturbative analysis of localisation in the weak coupling limit $\sigma \to i\infty$. The fixed point at

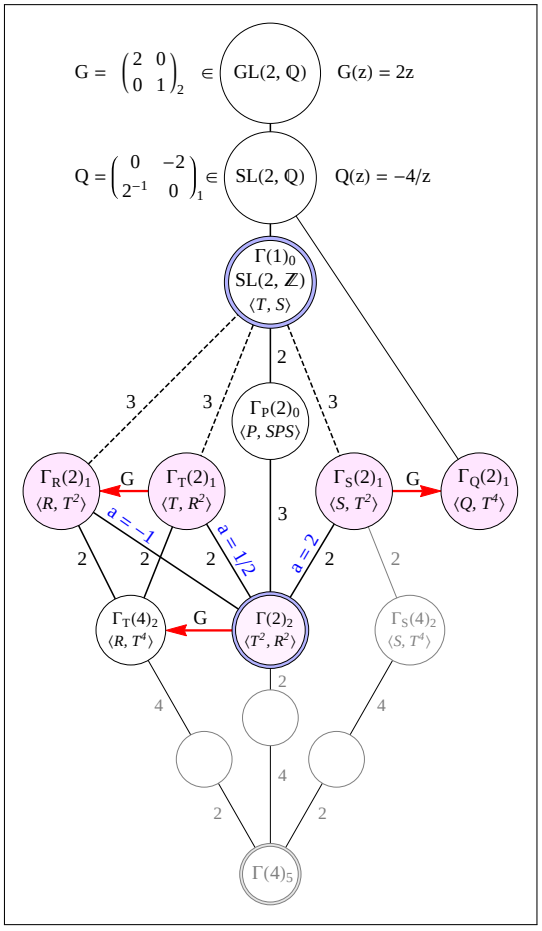

Figure 1. (Color online) Some of the groups between $\mathrm{GL}(2,\mathbb{Q})$ and $\Gamma(4)$ that are relevant for the QHE. Both $\Gamma(1)_0$ and $\Gamma_{\mathrm{P}}(2)_0$, where $P = ST$ ($P^3 = 1$), are too large to support a physical β-function. The pink groups are not too big, and it is their flow-diagrams that we compare to experiments.

vanishing coupling must therefore be repulsive, $i\infty = \ominus$. Since $\infty = 1/0 \in o/e$, and all fixed points in a given class are mapped into each other by the symmetry, all fractions in the class containing o/e must be repulsive. Notice that the denominators of attractors are always odd.

Fig. 1 shows some of the groups between $\mathrm{GL}(2,\mathbb{Q})$ and $\Gamma(4)_5$. The subscript is the number of linearly independent β-functions (weight two forms) that the symmetry allows.[31] A thick solid line means that the subgroup is normal, and the index of the subgroup labels the line. There are another twenty groups between $\Gamma(1)_0 = \mathrm{SL}(2,\mathbb{Z})$ and $\Gamma(4)_5$ that are not shown here.[32,33]

The red arrow is a *modular correspondence* obtained by conjugating with $G \in \mathrm{GL}(2,\mathbb{Q})$, where $G(z) = 2z$. The relation $\Gamma_{\mathrm{T}}(4)_2 = G\,\Gamma(2)_2\,G^{-1}$ is important in the theory of theta-functions (modular forms of weight $w = 1/2$).[34] Conjugating $\Gamma_{\mathrm{T}}(2)_1$ gives the familiar group $\Gamma_{\mathrm{R}}(2)_1 = G\,\Gamma_{\mathrm{T}}(2)_1\,G^{-1}$, but $\Gamma_{\mathrm{Q}}(2)_1 = G\,\Gamma_{\mathrm{S}}(2)_1\,G^{-1}$ is new. The G-conjugate $Q = G\,S\,G^{-1}(z) = -4/z$ of the duality generator S is an unnormalized *Fricke involution*. G changes

or *transmutes* a modular symmetry into an equivalent group, rather than *breaking* it to a smaller subgroup. This corresponds to moving horizontally and vertically in Fig. 1.

The symmetries that have been found to be relevant for the QHE are colored pink. When $a = -1,\ 1/2,\ 2$ the symmetry of the $\Gamma(2)_2$-invariant RG-potential φ_a discussed below (compare Sect. II C and Fig. 4) is enhanced to one of the maximal subgroups of $\Gamma(1)_0$, as indicated on the diagram. $\Gamma_X(2)_1$ (X = Q, R, T) are the symmetries most often observed in experiments. Since only level two appears to be physically relevant (so far), we often simplify notation by dropping the level ($\Gamma_X = \Gamma_X(2)_1$).

B. Modular phase-diagrams

Because the duality transformation S swaps e/o and o/e, leaving o/o unchanged, the direction of the flow in the $\rho = S(\sigma)$-plane is reversed if the symmetry acting on σ is Γ_T or Γ_R, but not if the symmetry is Γ_S or Γ_Q. This dichotomy is a persistent theme.

The fixed point at the origin of the σ-plane (at $i\infty$ in the ρ-plane) has a special significance. If it is attractive this means that the system has an insulating phase, which we call the quantum Hall insulator (QHI) and assign the special symbol ❀. Since $0 = 0/1 \in e/o$, we conclude that a model with Γ_T- or Γ_R-symmetry in the σ-plane does have this insulator phase, but that Γ_S- and Γ_Q-symmetric models do not (compare eq. (1) and Fig. 2).

Notice that all plateaux in the Hall conductivity are accompanied by vanishing magnetoconductivity ($\sigma_D = 0 \implies \sigma_H \in \mathbb{Q}$), also for the insulator phase, as expected from naive localization theory (compare Fig. 2). The same is true for the resistivities, except for the peculiar insulator phase ($\rho_D = 0 \implies \rho_H \in \mathbb{Q} \neq 0$). In this case the "plateau" ❀ = $i\infty$ is associated with a diverging magnetoresistivity, while the Hall resistivity is not quantized ($\rho_D \to \infty \implies \rho_H \in \mathbb{R}$), because there is only one point at infinity in $\overline{\mathbb{C}}^+$.

The experimental signature of this phase is a plateau in the Hall conductivity with $\sigma_H = 0$, accompanied by a large *peak* in the magnetoresistivity, $\rho_D >> 1\,[h/e^2]$. This is, for example, what is observed experimentally when graphene is placed in a very strong magnetic field,[35,36] signalling that the modular symmetry is changing from Γ_Q to Γ_T, as discussed in Sect. VII.

Observe also that Γ_R and Γ_T are conjugate inside the parent group $GL(2,\mathbb{Q})$ under the rescaling $G(z) = 2z$ by a factor of two (compare Fig. 1). This means that flow-diagrams with these two symmetries are identical, up to a doubling of all coordinates. A similar rescaling of Γ_S gives a conjugate group Γ_Q that is not strictly speaking modular (compare Fig. 1), but its flow-diagram is just a doubling of the Γ_S-symmetric flow.

IQHE	$\sigma = \sigma_H + i\sigma_D \in \overline{\mathbb{C}}^+(\sigma)$			$\rho = \rho_H + i\rho_D \in \overline{\mathbb{C}}^+(\rho)$		
Γ_X	⊕	⟵ ⊗ ⟶	⊕′	⊕	⟵ ⊗ ⟶	⊕′
Γ_T	n	$\frac{2n+1+i}{2}$	$n+1$	$\frac{1}{n+1}$	$\frac{2n+1+i}{2n^2+2n+1}$	$\frac{1}{n}$
Γ_R	$2n$	$2n+1+i$	$2n+2$	$\frac{1}{2n+2}$	$\frac{2n+1+i}{4n^2+4n+2}$	$\frac{1}{2n}$
Γ_S	$2n-1$	$2n+i$	$2n+1$	$\frac{1}{2n+1}$	$\frac{2n+i}{4n^2+1}$	$\frac{1}{2n-1}$
Γ_Q	$4n-2$	$4n+2i$	$4n+2$	$\frac{1}{4n+2}$	$\frac{2n+i}{8n^2+2}$	$\frac{1}{4n-2}$

Table I. Left half: Integer plateaux values ⊕ of the Hall conductivity σ_H constrained by a symmetry $\Gamma(2) \subset \Gamma_X \subset SL(2,\mathbb{Q})$, with X = Q, R, S or T, together with the location of semi-stable fixed points for transitions between these plateaux, i.e. the position of "integer" quantum critical points ⊗ in the complexified conductivity-plane. Right half: Corresponding values of the resistivity (see Sect. 2 for details).

In summary, since both $\Gamma(1)_0$ and $\Gamma_P(2)_0$ are too large, there are just two types of physically acceptable conductivity flow-diagrams with maximal modular symmetry: Γ_T (and its G-conjugate Γ_R), and Γ_S (and its G-conjugate Γ_Q).

For convenience an "atlas" of Q-, R-, S- and T-symmetric flows, in both σ and ρ, is provided in Fig. 2. In these cases the shape of the flow lines (but not the flow rate) is completely fixed by the large symmetry. They are most easily derived as a gradient flow of RG-potentials with the requisite symmetry (compare eq. (2)). We defer details to the discussion below of symmetry transmutations.

For future reference we have also listed the integer fixed points for these cases in Tab. I. The complete spectrum of attractors (plateaux) for these symmetries may be found in Fig. 5.

Γ_T and Γ_R are the relevant groups for respectively the ordinary spin-polarized and unpolarized QHE, where quasi-particles have the usual parabolic ("non-relativistic") dispersion, i.e., the QHE that appears in materials without Dirac-modes. We will therefore call these the *non-relativistic polarized and unpolarized* groups.

Graphene is different. Because of the peculiar topology of its Fermi surface, there is a doubling of degrees of freedom due to an additional "pseudospin" or "valley" degeneracy, and there are gapless (massless) excitations at half filling with linear dispersion, i.e., their energy is linear in momentum. These modes therefore behave like relativistic (Dirac) fermions, with the Fermi velocity replacing the speed of light. The linear dispersion and unusual band structure leads to a different non-interacting spectrum, but that is all we need to identify the potential modular symmetry, and the phenomenological analysis of

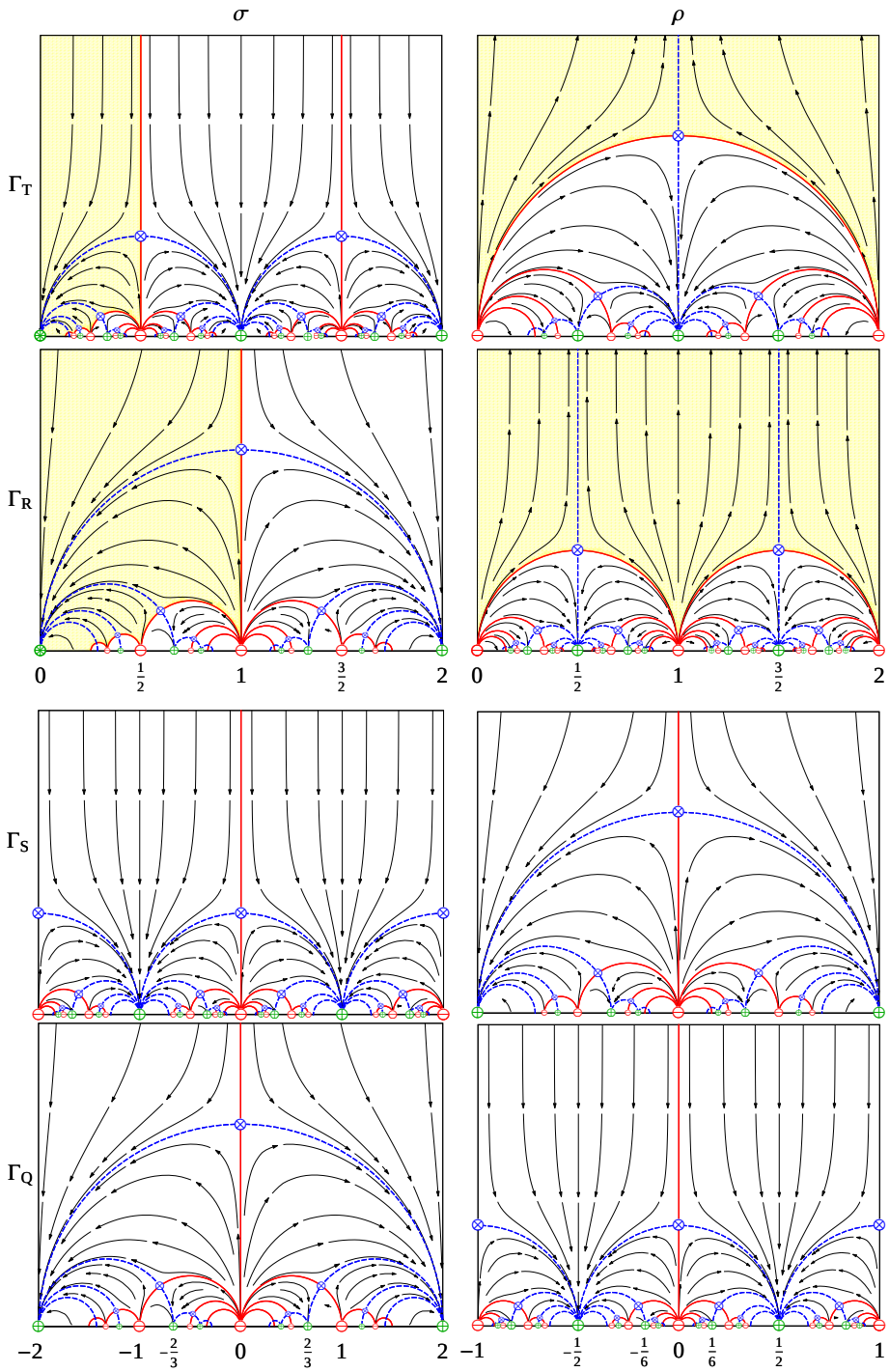

Figure 2. (Color online) Conductivity (σ) and resistivity ($\rho = -1/\sigma$) phase-diagrams with symmetry Γ_X (X = Q, R, S, T). Only $\Gamma_T(\sigma)$ and $\Gamma_S(\sigma)$ are truly different, since $\Gamma_R(\sigma)$ is simply a doubling of $\Gamma_T(\sigma)$, and likewise for $\Gamma_Q(\sigma)$ and $\Gamma_S(\sigma)$. The main physical distinction is that Γ_T (and Γ_R) has an insulator phase (yellow; $\sigma_{\circledast} = 0$, $\rho_{\circledast} = i\infty$), while Γ_S (and Γ_Q) does not.

graphene is analogous to the parabolic case.[12,13]

Note that a topological zero-mode eliminates the insulator phase, so the relevant groups in this "relativistic" case with Dirac modes are Γ_S and Γ_Q, for respectively the spin-polarized and unpolarized QHE. We will therefore call these the *relativistic polarized and unpolarized* groups.

A phase is by definition the set of all points in $\overline{\mathbb{C}}^+$ that flow to a given plateau $\oplus$ (IR fixed point), and it is

uniquely labelled by this rational limit point on the real axis. A phase transition between two plateaux $\oplus$ and $\oplus'$ is permitted by the symmetry iff it has a fixed point $\otimes$ located on the semi-circle in $\overline{\mathbb{C}}^+$ connecting $\oplus$ and $\oplus'$, which we write as $\oplus \leftarrow \otimes \rightarrow \oplus'$ or $\oplus \overset{\otimes}{\longleftrightarrow} \oplus'$. If one of the attractors is $i\infty$ the semi-circle has infinite radius, i.e., it is a vertical line. We also adopt the convention that $\oplus \overset{\otimes}{\longleftrightarrow} \oplus'$ refers to a transition in the conductivity plane, whence an integer plateau-value $\oplus = \sigma_\oplus = \sigma_H = n\,[e^2/h] \in \mathbb{Z}$ refers to the IQHE, for which $\rho_H = 1/n\,[h/e^2]$.

C. Transmutation of symmetries

Degeneracies in the spectrum of delocalized states may be broken either by external fields, or by internal many-body interactions between charge carriers within each band, or both.

In the simplest materials we only have spin-degeneracy to consider. A symmetry transmutation occurs when the spins are neither fully polarized, nor fully degenerate, in which case the modular symmetry is at least partially broken. The maximal groups are no longer relevant, but it is conceivable that some smaller symmetry survives. The simplest situation is if we have *minimal symmetry breaking*, which means that the largest common subgroup survives. From our map in Fig. 1 we see that this group is $\Gamma(2)$, and our task is to find a $\Gamma(2)$-symmetric family of physically sensible β-functions that interpolates between Γ_R, Γ_T and Γ_S. Γ_Q is not in this family because it is not in the modular group $\Gamma(1)$.

$\Gamma(2)_2$ admits a two-dimensional vector space of weight two forms, which is spanned by two Jacobi theta-functions, for example θ_3^4 and θ_4^4. Since θ_2^4 is also a weight two form it must be somewhere in this space, and because $\theta_2^4 = \theta_3^4 - \theta_4^4$ it is. Any $\Gamma(2)$-symmetric β-function must be a linear combination of these:[22]

$$\begin{aligned} \beta_a &\propto (1-a)\,\theta_3^4 + a\,\theta_4^4 = \theta_3^4 - a\,\theta_2^4 \propto \partial\varphi_a \\ \varphi_a &= \ln\lambda + (a-1)\ln(\lambda-1), \quad \lambda = \theta_2^4/\theta_3^4. \end{aligned} \tag{2}$$

This is, as expected, a gradient flow, derived from the $\Gamma(2)$-invariant RG-potential φ_a. It interpolates between the maximally symmetric cases labelled R, S and T, while retaining as much modular symmetry as possible.[37] The phenomenological parameter a has an unknown and presumably complicated dependence on non-universal microscopic details, like many-body interactions and Zeeman-splitting. Provided we choose the normalization of β_a to be imaginary, a must be real for the flow to agree with perturbative localization theory at weak coupling ($1/\sigma_D \to 0$).

The thousands of non-semicircular flow lines, phase boundaries and separatrices shown in this article were all obtained directly from φ_a by numerical integration, for a handful of real values of a. This includes the 282 modular flow lines that are compared to 1 484 experimental data points in Figs. 6 - 17. In all but one case (compare Fig. 17) the agreement is within the estimated experimental uncertainty. We find that 1 434 data points on 274 flow lines appear to be consistent with modular ($\Gamma(2)$ or Γ_Q) symmetry, for only a few (seven) real values of a, and most of these (1 295 data points on 243 flow lines) are consistent with one of the maximal symmetries Γ_X (X = Q, R, S, T). It is this universality we wish to investigate here.

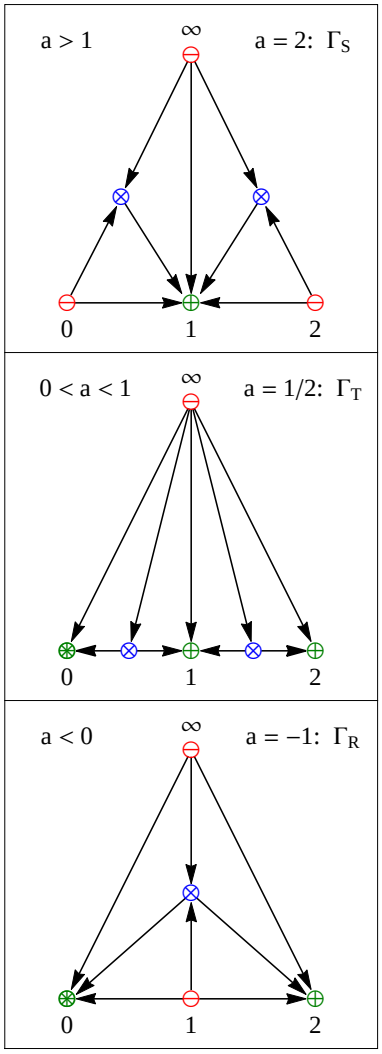

Figure 3. (Color online) Schematic of the three subfamilies of the $\Gamma(2)_2$-family φ_a. Each of these has one (and only one) member with enhanced symmetry: φ_{-1} has Γ_R-symmetry, $\varphi_{1/2}$ has Γ_T-symmetry, and φ_2 has Γ_S-symmetry (compare Fig. 4). We call these subfamilies the R-, T- and S-families.

When $a = 0,\, 1,\, \infty$ the β-function degenerates to one of the theta-functions ($\beta_0 \propto \theta_3^4$, $\beta_1 \propto \theta_4^4$, and $\beta_\infty \propto \theta_2^4$), which are finite. So for these three exceptional values of a (and only these), β_a has no zeros, and the flow has no fixed points (critical points) in $\mathbb{C}^+$. This phenomenon is unavoidable, because the family has three subfamilies

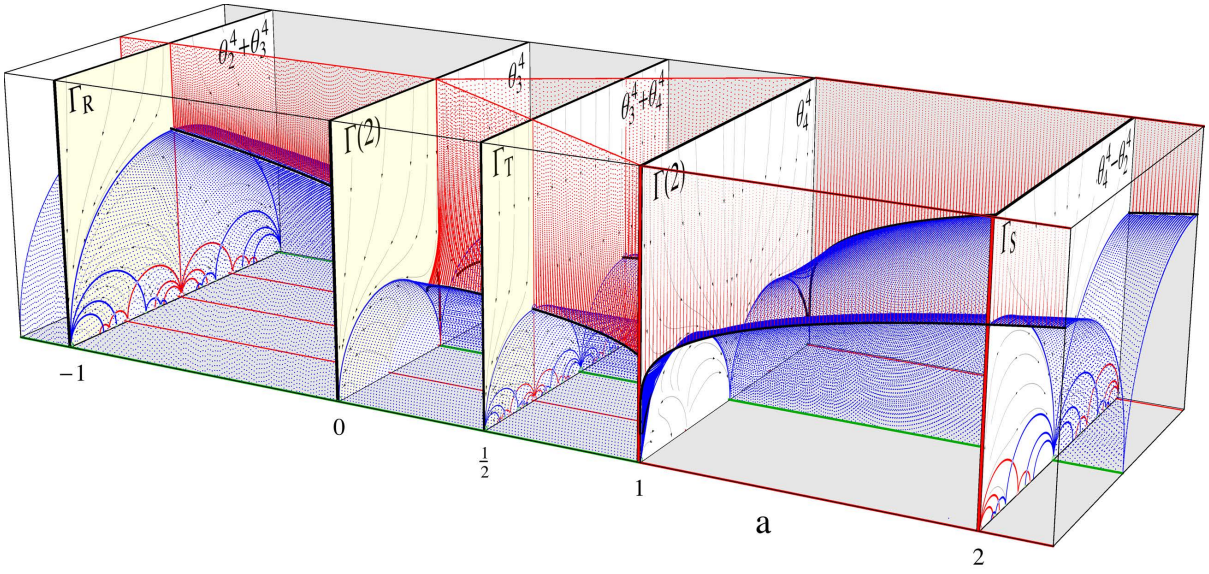

Figure 4. (Color online) The one-parameter family β_a of $\Gamma(2)$-symmetric RG-flows, shown here for values of the symmetry-breaking parameter a in the range $-1.2 < a < 2.2$, is divided into three subfamiles by the "ramification points" $a = 0, 1$.

where the plateaux-structures are quite distinct. Consider for example the triplet of fixed points $\sigma_* \in (0, 1, 2)$. We have $(0, 1, 2) = (\circledast, \ominus, \oplus)$ for $a < 0$, $(0, 1, 2) = (\circledast, \oplus, \oplus)$ for $0 < a < 1$, and $(0, 1, 2) = (\ominus, \oplus, \ominus)$ for $1 < a$ (compare Fig. 3). At $a = 0, 1$ the critical points disappear by merging with the fixed points $\sigma_* \in (0, 1, 2)$, allowing a phase to appear ($\ominus \longrightarrow \oplus$) or disappear ($\oplus \longrightarrow \ominus$), as shown schematically in Fig. 3.

Fig. 4 shows the complete family of $\Gamma(2)$-symmetric flow-diagrams for the range of most physical interest. Each subfamily has one member for which the symmetry is enhanced from $\Gamma(2)$ to one of the maximal subgroups: Γ_R when $a = -1$, Γ_T when $a = 1/2$, and Γ_S when $a = 2$. For example, a very strong magnetic field gives a large Zeeman splitting that leads to Γ_T-symmetry in the non-relativistic case. For weak fields this symmetry is transmuted into an unpolarized spectrum, $\Gamma_T \xrightarrow{G} \Gamma_R$.

Each panel is labelled on the left by the symmetry of the flow, and on the right by a function proportional to the β-function that generates the flow. The yellow region is the insulator phase, which disappears when $a > 1$.

Only at the unphysical singular points $a = 0, 1$ are there no quantum critical points. Slicing this "family-plot" at any value of the symmetry-breaking parameter $a \neq 0, 1$ gives a "warped" but physicially sensible diagram, i.e. a scaling flow that is finite except for simple zeros (compare Fig. 17). These are the *quantum critical points*, located at $\sigma_{\otimes} = iK'(a^{-1})/K(a^{-1})$ and all its $\Gamma(2)$-images, where K and K' are elliptic integrals of the first kind.[22]

This family is sufficiently large to accomodate almost all quantum Hall data that we have examined so far (one possible exception is discussed in Sect. VII).

In some materials the band structure is more subtle, with additional "competing" degeneracies, and the pattern of symmetry breaking may be more complicated. For example, the four-fold spin-pseudospin degeneracy in graphene giving rise to Γ_Q-symmetry can be broken by internal many-body or external magnetic field interactions. Independently of the microscopic mechanism, Fig. 5 shows an idealized pattern of symmetry breaking that transmutes flow diagrams. In this scenario bands split if degenerate spin or pseudospin states (or both) are resolved. When a band splits a new delocalized state appears, giving rise to a new plateau in the Hall conductivity. Each symmetry (top row) has a unique spectrum of integer attractors (IQHE plateaux) (bottom row).

In real materials like graphene many-body interactions, which presumably are responsible for pseudospin splitting, may obfuscate this simple picture. Since electron correlations appear to be strongest for the lowest Landau level, degeneracies may not be equally robust for all bands, leading to a hierarchy of plateaux spectra that only manifests a modular symmetry for limiting cases (strong and weak magnetic field, say). As in atomic physics, it may nevertheless be useful to retain the group theoretic labelling of states for intermediate cases where the symmetry is broken. Some graphene experiments exploring this question are discussed in the penultimate section.

III. NEW MATERIALS

We have argued that the convergence of modular mathematics and quantum Hall physics suggests that it would be unnatural to restrict attention to only one of the de-

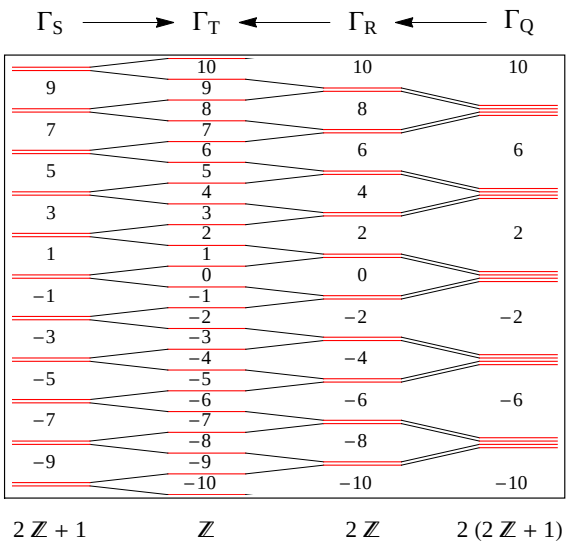

Figure 5. (Color online) Simplified pattern of symmetry transmutations from band splitting (lifting of spin and/or pseudospin degeneracies). Band gaps are labelled by the filling factor. Each symmetry $\Gamma_{\rm X}$ (X = Q, R, S, T) leaves a unique fingerprint on the spectrum $\oplus_{\rm X} \in \mathbb{Q}$ of attractive fixed points, i.e., rational plateaux values of σ_H $[e^2/h]$.

scendants of the modular group. We have also seen that there are very few viable candidates to choose from, and that most of these fit snugly into a simple and unique one-parameter family of $\Gamma(2)_2$-symmetric β-functions (up to normalization). In other words, while these infinite non-abelian symmetries are extremely constraining, they do leave enough flexibility that we can accomodate almost all experiments to date (but only barely so).

The discovery in recent years of new types of materials that support Dirac modes and "robust" topological edge states presents new opportunities for testing the modular paradigm sketched above. We will review a number of recent experiments that have explored large tracts of the modular landscape that were previously inaccessible.

These experiments have provided substantial evidence for those level two symmetries that until now have been beyond our reach. The data we have analyzed suggests that the full complement of level two symmetries may be present in Nature. In preparation for that discussion we give a brief summary of some of the most salient features of these materials.

A. Dirac matter

Dirac matter is a name used to describe materials in which the low-energy excitations are Dirac fermions. In Bloch theory these states appear as a consequence of a finite number of crossing points in the Brillouin zone where the Hamiltonian becomes gapless. If the energy dispersion close to these points is linear, similar to the relativistic dispersion in particle physics, this is called a Dirac cone, and the effective low energy Hamiltonian is Dirac-like, where the Fermi-velocity replaces the speed of light. When a Hall effect takes place in such materials, each zero mode contributes 1/2 to the Hall conductivity.[38] The Nielsen-Ninomiya theorem guarantees that Dirac cones come in pairs, ensuring an integral conductivity.

The most familiar material with linear ("relativistic") dispersion is graphene, where two Dirac cones sit at corners of the Brillouin zone. In the presence of a magnetic field, each Dirac fermion contributes $n + 1/2$ to the Hall conductivity. Taking into account both spin and valley (pseudospin) degeneracy the IQHE in graphene is $\sigma_H = 4(n + 1/2) = 2 \mod 4 \in \oplus_{\rm Q}$ $(n \in \mathbb{Z})$. The most unusual property of this plateaux spectrum is the absence of the attractor $\circledast = 0$, i.e., an insulator phase, which is a consequence of the zero modes shifting the Hall spectrum. In the ordinary (un-)polarized IQHE the plateaux spectrum is $\mathbb{Z}\,(2\mathbb{Z}) = 0 \mod 1\,(2)$.

B. Topological insulators

Topological insulators are special phases of matter characterized by a gapped bulk material with gapless edge or surface modes.[40] These gapless modes are topologically protected in the sense that they are robust to perturbations that preserve the symmetries of the system. The theory of topological insulators relies on Bloch theory as well as recent mathematical tools like Chern numbers and homotopy theory to characterize classes of Hamiltonians that preserve the bulk gap.

A normal insulator is said to be topologically trivial. The QED vacuum presents an insulator in this class. Here two bands are associated with electrons and positrons, while a large gap is associated with the pair production energy. The gapless surface modes of a topological insulator appear as a necessary consequence of a topologically non-trivial material ending on a trivial one (e.g. the vacuum). The only way a topological property can change across the interface is for the gap to close. This relation between bulk topology and edge modes are called the bulk-edge correspondence or duality.

The first topological insulator to be discovered was the IQHE itself. Here the Landau levels serve as energy bands, while a strong magnetic field induces a gap up to the first empty level. The bulk-boundary correspondence is in this case attributed to electrons skipping along the edges of the Hall sample due to the magnetic field. In this case it is not the material that is considered a topological insulator but the IQHE as a whole.

Depending on the material in which the Hall effect takes place, different imprints are seen on the Hall conductivity. Graphene, for example, has a unique Hall spec-

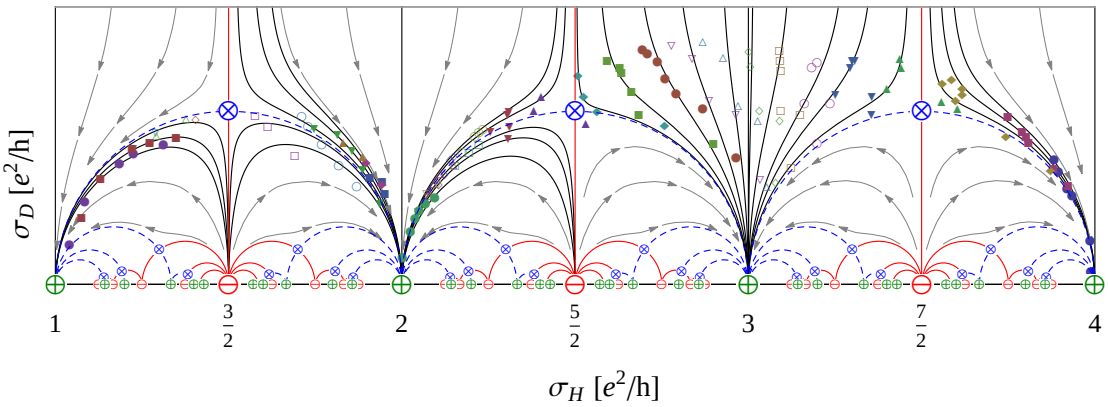

Figure 6. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transitions $1 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 2 \overset{\otimes}{\longleftrightarrow} 3 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 4$ in a semiconducting InGaAs/InP heterojunction.[39]

trum $\oplus_{\mathrm{Q}} = 4n + 2$ due to its two Dirac cones.

Another example is provided by the surface of a 3-dimensional topological insulator, which can serve as an effective two-dimensional arena for the QHE. The bulk-boundary correspondence tells us that this surface has massless excitations. Depending on the bulk topology the surface Brillouin zone either an even or an odd number of Dirac cones are present,[40] and the effective two-dimensional material can be seen as a Dirac material. In the case of an odd number of Dirac cones the Nielsen-Ninomiya theorem appears to be broken. This is solved by the existence of partner Dirac fermions at the opposite surface of the 3-dimensional topological insulator.[40] Under the assumption that the two sides are independent the Hall conductivity will be a sum of both contributions.

IV. UNIVERSALITY CLASS Γ_{T}

A. Plateaux transitions in InGaAs/InP

The result of the first scaling experiment in the context of the QHE, obtained in 1985 using a semiconducting heterojunction cooled below 4.2 K,[39] is reconstructed in Fig. 6 from the published data. Clear indications of a modular symmetry are already evident in this diagram (compare Fig. 2), even with the large uncertainty in the data.

Fig. 6 shows our reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transitions $1 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 2 \overset{\otimes}{\longleftrightarrow} 3 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 4$ in a semiconducting InGaAs/InP heterojunction with 2D electron density $n = 3.4 \times 10^{11}\,\mathrm{cm}^2$, mobility $\mu = 35\,000\,\mathrm{cm}^2/\mathrm{Vs}$ and effective mass $m^* = 0.047\,m_e$ (m_e is the free electron mass), in the temperature range 4.2 K (top) to 0.5 K (bottom).[39]

Comparison with a modular scaling flow (solid lines) with quantum critical points at $\otimes = 1/2, 3/2$ and $5/2$ reveals a Γ_{T}-symmetry in the transport data (compare Fig. 2).

In the three decades following this pioneering experiment technology has improved and error bars have shrunk. In the following we shall see that not only have experiments failed to contradict the symmetry, the agreement with the coldest experiments, where the symmetry is expected to be most accurate, is now in some cases at the per mille level.

B. Plateaux transitions in GaAs/GaAlAs

Figs. 7 and 8 provide further evidence for the existence of a universality class with Γ_{T}-symmetry that unifies the IQHE (Fig. 8) with the FQHE (Fig. 7).

C. Plateau-insulator transition in Cr(BiSb)Te

The QHE can take place on the top of 3-dimensional topological insulators,[43] like bismuth antimonide $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$ which was the first 3-dimensional topological insulator to be discovered.[44] The effective edges of these two-dimensional surface systems are magnetic domain walls along which the charge carriers move.

Fig. 9 and 10 show our reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1$ in a 2D ferromagnetic topological insulator (thin film of $\mathrm{Cr}_x(\mathrm{Bi}_{1-y}\mathrm{Sb}_y)_{2-x}\mathrm{Te}_3$ grown on a semi-insulating InP (111) substrate).[45] After applying an external magnetic field $B = 14\,\mathrm{T}$ to saturate the magnetization, the magnetic field strength was set to zero and experiments were performed at different temperatures with tunable gate voltage. In order to compensate for what is presumably

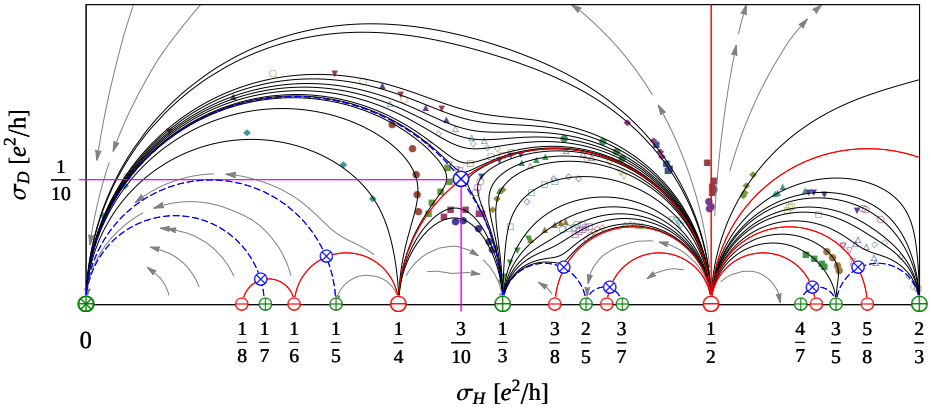

Figure 7. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the fractional transitions $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1/3 \longleftarrow \overset{1/2}{\ominus} \longrightarrow 3/5 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 2/3$ in a GaAs/GaAlAs heterojunction.[41]

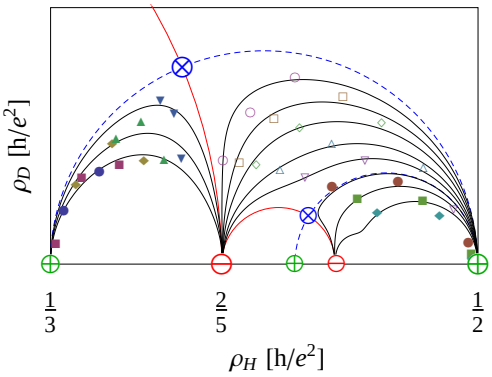

Figure 8. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transition $2 = \oplus \overset{\otimes}{\longleftrightarrow} \oplus = 3$ in a GaAs/GaAlAs heterojunction.[42]

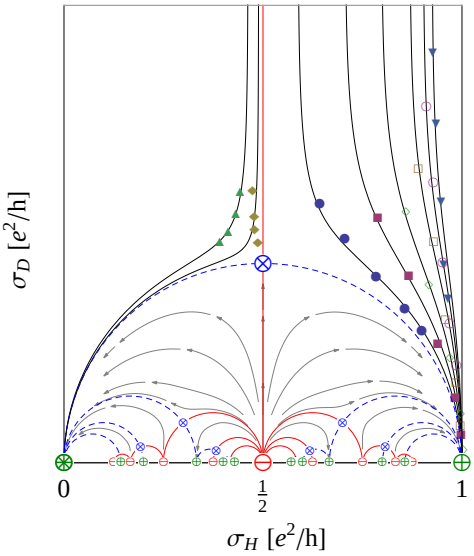

Figure 9. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1$ in a 2D ferromagnetic topological insulator (a thin film of $Cr_x(Bi_{1-y}Sb_y)_{2-x}Te_3$ grown on a semi-insulating InP(111) substrate).[45]

a systematic error of unknown origin, the data in Fig. 9 has been shifted slightly to the left so that the plateaux are integer-valued.

In both cases, comparison with a modular scaling flow (solid lines) with a quantum critical point at $\otimes = (1+i)/2$ (compare Tab. I and Fig. 2) reveals that these transport data are in excellent agreement with Γ_T-symmetry.

D. Plateaux transitions in mercury telluride

Bulk mercury telluride is a semi-conductor of the II-VI type,[46] but when used to create a quantum well (HgCdTe/HgTe/HgCdTe) the electronic properties depend crucially on the thickness d of the sample. This thickness introduces a parameter which can be tuned to find quantum phase transitions. For thin wells with thickness below the critical thickness $d_c \approx 6.3\,\text{nm}$ the material has a normal band structure, whereas for wide wells ($d > d_c$) the band structure is inverted.[46,47]

In addition to having a highly specific energy spectrum with an inverted band structure, the 2DEG in a wide HgTe quantum well is characterized by a low effective mass, $m^* = 0.02\,m_e$.[48] The low effective mass causes a large Landau level separation $\Delta E = \hbar qB/m^*c$, and the QHE survives to relatively high temperatures. In [49 and 50] a strong integer effect was observed up to $T \sim 10 - 15\,\text{K}$.

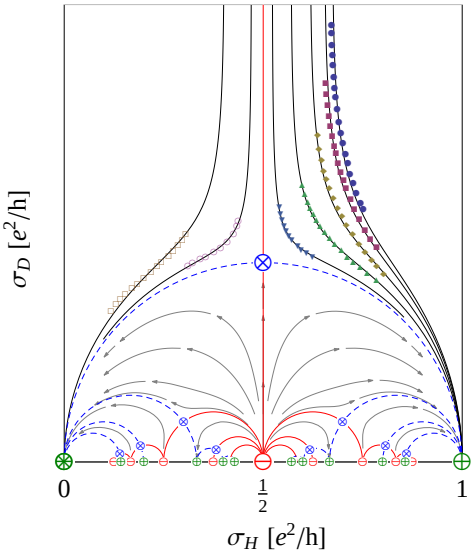

Figure 10. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1$ in a 2D ferromagnetic topological insulator (a thin film of $Cr_x(Bi_{1-y}Sb_y)_{2-x}Te_3$ grown on a semi-insulating InP(111) substrate).[45]

Fig. 11 shows our reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transitions $1 \overset{\otimes}{\longleftrightarrow} 2 \overset{\otimes}{\longleftrightarrow} 3$ in a heterostructure $Hg_xCd_{1-x}Te/HgTe/Hg_xCd_{1-x}Te$ ($x \approx 0.7$) with a 20.3 nm wide HgTe quantum well.[50] Since this thickness is well above d_c there should be no Dirac cones in the bulk Brillouin zone. The sample was grown by molecular beam epitaxy on a GaAs substrate, symmetrically modulation doped with In at both sides of the quantum well, yielding a mobility of $22 \times 10^4\, cm^2/Vs$ and an electron gas density of about $1.5 \times 10^{15}/m^2$.[50,51]

The longitudinal and Hall resistivities were measured with a constant 1 A current in the temperature range 2.9 – 50 K, and a magnetic field strength in the 0 – 9 T range. There is clear evidence for plateaux at $\nu = 1, 2, 3$ and 4, obtained for magnetic fields in the range 1.8 – 8 T. For most magnetic field values the system exhibited scaling behaviour for the five lowest temperatures T = 2.9, 4.1, 6.1, 8.1 and 10 K, and in one instance also for 15 and 20 K. In some cases, close to the fix points only the three lowest temperatures were usable.

Comparison with a modular scaling flow (solid lines) with quantum critical points at $\otimes = (2n + 1 + i)/(2n^2 + 2n + 1) = 1 + i,\ (3 + i)/5,\ (5 + 1)/13,\ (7 + i)/25, \ldots$ reveals a Γ_T-symmetry in the transport data (compare Tab. I and Fig. 2).

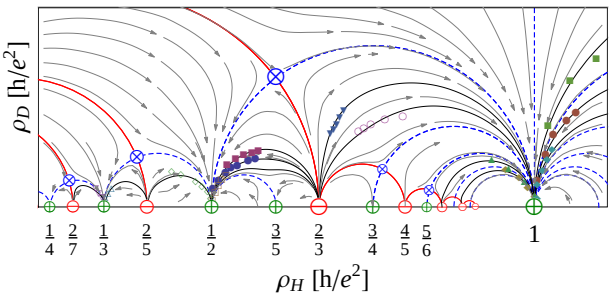

Figure 11. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transitions $1 \overset{\otimes}{\longleftrightarrow} 2 \overset{\otimes}{\longleftrightarrow} 3$ in a HgTe/HgCdTe heterostructure with a wide HgTe quantum well.[50]

E. Plateau-insulator transitions in bismuth antimonide

In [52] the QHE was studied by measuring surface conductivities on the top and bottom of the 3-dimensional topological insulator bismuth antimonide. Two 8 nm thick TI films of $(Bi_{1-x}Sb_x)_2Te_3$ ($x = 0.84, 0.88$) were grown on insulating InP (111) substrates using molecular beam epitaxy. Quantum Hall signatures were found at magnetic field strengths above 14 T, for temperatures ranging from 700 mK down to 40 mK, at various gate voltages V_G.

Fig. 12 shows our reconstruction of their temperature-driven scaling data (discrete icons) exploring the two plateau-insulator transitions $-1 = \oplus \overset{\otimes}{\longleftrightarrow} \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1$. Inaccessible data points and clear statistical outliers were not considered when sampling the data.

Comparison with a modular scaling flow (solid curves) with quantum critical points at $\otimes = (\pm 1 + i)/2,\ (\pm 3 + i)/2, \ldots$ reveals a Γ_T-symmetry in the transport data (compare Tab. I and Fig. 2).

F. Plateaux transitions in black phosphorus

In addition to graphene, black phosphorus is the only other 2D atomic crystal where a QHE has been observed.

Fig. 13 shows a scaling flow derived from our reconstruction of data obtained in an experiment on a few layers of black phosphorus, which were sandwiched between two layers of insulating hexagonal boron nitride (hBN) and placed on a graphite back-gate to create a van der Waals heterostructure.[53]

The 25 nm bottom layer of hBN allows the electrons in the graphite to screen the impurity potential at the black phosphorus-hBN interface, which gives a record high Hall mobility of $6\,000\, cm^2/Vs$ for this material. It is this large mobility that gives an observable QHE.[53]

The data is extracted from Fig. 7 of the supplementary

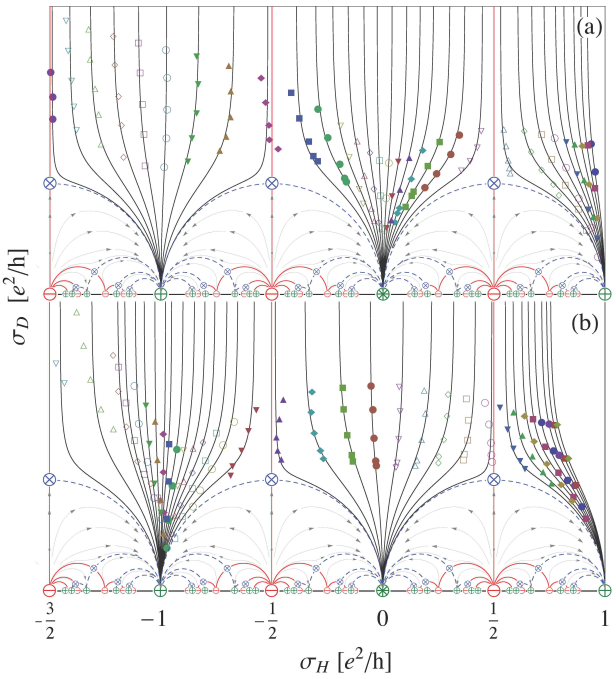

Figure 12. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux-insulator transitions $-1 = \oplus \overset{\otimes}{\longleftrightarrow} \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1$ in a bismuth antimonide topological insulator $(Bi_{1-x}Sb_x)_2Te_3$, with (a) $x = 0.88$, and (b) $x = 0.84$.[52]

material of [53]. The Hall resistances were measured at fixed magnetic fields of 27, 29, 31 and 33 T, and temperatures 1.7, 4.1, 4.6, 6, 8 and 10 K, by varying the back gate voltage from −2 to −0.7 V. Plateaux were discovered for filling factors $\nu = 1$, 2 and 3. Due to overlap of the Hall resistance curves, an area of ∼ ±013 V at the inflection point of the $1 \overset{\otimes}{\longleftrightarrow} 2$ transition had to be excluded. The curves for 8 and 10 K were also excluded because the magnetoresistance did not vanish on the plateaux.

Deriving the magnetoresistivity $\rho_D = \lambda R_D$ from the measured magnetoresistance R_D requires knowing the aspect ratio $\lambda = L_y/L_x$ of the Hall bar (of length L_x and width L_y). Since this information was not provided in [53] we also fitted λ. The best fit of the data gave $\lambda \approx 3$, which is consistent with the optical image of the device (black phosphorus/hBN/graphite heterostructure) shown in Fig. 1 (a) of [53].

Comparison with a modular scaling flow (solid lines) with quantum critical points at $\otimes = (2n+1+i)/(2n^2+2n+1) = 1+i$, $(3+i)/5$, $(5+1)/13$, $(7+i)/25$, ... reveals a Γ_T-symmetry in the transport data (compare Fig. 2).

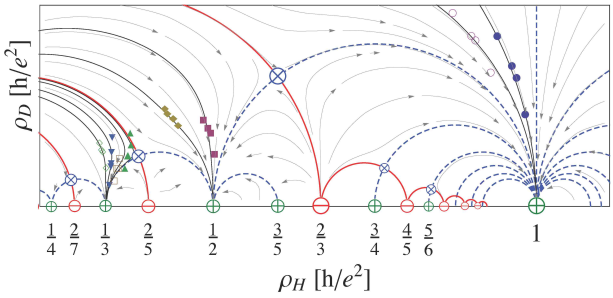

Figure 13. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateaux transitions $1 \overset{\otimes}{\longleftrightarrow} 2 \overset{\otimes}{\longleftrightarrow} 3$ in black phosphorus.[53]

V. UNIVERSALITY CLASS Γ_R

A. Plateau-insulator transition in GaAs/GaAlAs

Fig. 14 shows our reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in a GaAs/GaAlAs heterojunction.[54] Comparison with a modular scaling flow (solid lines) with a quantum critical point at $\otimes = 1+i$ reveals a Γ_R-symmetry in the transport data (compare Fig. 2).

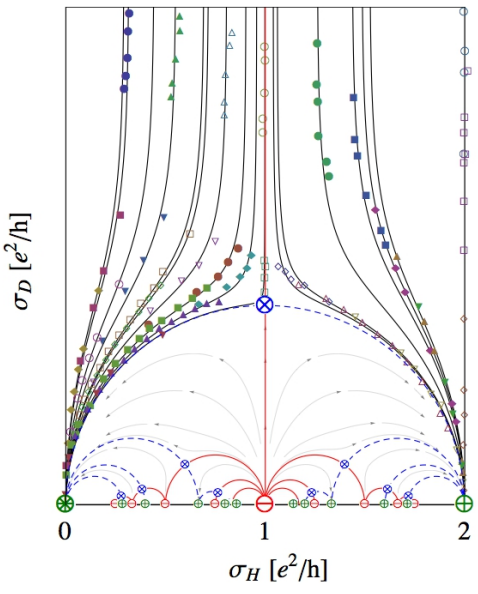

Figure 14. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in a GaAs/GaAlAs heterojunction.[54]

B. Plateau-insulator transition in graphene

Fig. 15 shows our reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in graphene.[55] In order to compensate for what is presumably a systematic error of unknown origin, the dataset close to the dashed blue semi-circle has been shifted up slightly, so that the flow does not violate the semi-circle law (i.e., so that the flow does not cross the separatrix connecting the plateau $\oplus$ to the insulator $\circledast$ via the critical point $\otimes$). Comparison with a modular scaling flow (solid lines) with a quantum critical point at $\otimes = 1 + i$ reveals a Γ_R-symmetry in the transport data (compare Fig. 2).

In this experiment large-area ($0.6 \times 0.1\,\mathrm{mm}^2$) monolayer graphene devices were made by epitaxial growth on SiC-substrate. In the devices, a buffer layer of graphene made partial covalent bonds with the exposed Si atoms and only the top graphene layer was conducting. Experiments were made in the temperaure range $2.6 - 25\,\mathrm{K}$ with magnetic fields in the range $0.1 - 9\,\mathrm{T}$.

According to [55] the graphene-substrate coupling, which includes Si-C covalent bonds and defects, such as interfacial dangling bonds, can be strong enough to break the sublattice symmetry of the conducting graphene sheet. In order to enhance this effect the Coulomb screening of potential fluctuations was reduced by engineering the carrier density to be as low as $n \approx 10^{15}/m^2$.

This may be the reason for the appearance of an insulator phase, which signals that the Γ_Q-symmetry observed in ordinary monolayer graphene has been transmuted to a Γ_R-symmetry (compare Fig. 5).

The data that best fit the flow lines are taken from one of the least disordered samples which also had the highest surface roughness (called EG2 in [55]). The data from the other sample (EG3) appears to fit slightly better if shifted up by about $0.03e^2/h$. This may be the result of a small systematic error, but it is so small that it may be within the random error of this experiment.

VI. UNIVERSALITY CLASS Γ_Q

We have already mentioned the spectrum of plateaux observed in some experiments on graphene. The competition between several scales is not easy to disentangle, especially in crossover regions where the lowest Landau level may be more susceptible to symmetry breaking than higher levels. However, so far it seems that the symmetries we have discussed (compare Fig. 1) suffice to account for the plateaux data.

A much more stringent test is, as we have seen in the non-relativistic case, to compare the unstable fixed points with experimental quantum critical points. Scaling experiments on graphene are still in their infancy, and the paucity of data means that this analysis is far from conclusive. Unfortunately, so far a meaningful comparison is only possible for the doubly degenerate IQHE, which should be compared with the phase- and flow-diagram in Fig. 2.

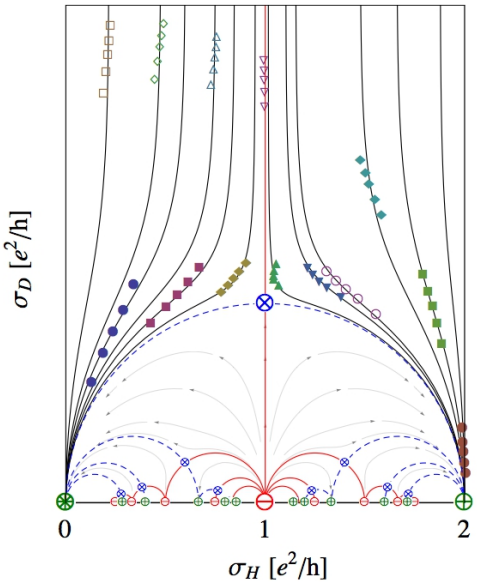

Figure 15. (Color online) Reconstruction of temperature-driven scaling data (discrete icons) exploring the plateau-insulator transition $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in graphene.[55]

Because of the zero-mode there is no QHI ($\sigma_\oplus = \circledast = 0$) in this case, so Γ_T and Γ_R are immediately eliminated as potential symmetries. A glance at the defining characteristics of the groups in eq.(1) shows that, up to a factor of two, Γ_S is the only viable candidate. Because of the double degeneracy in graphene the conductivity should be doubled,[56] $\sigma \to 2\sigma = G(\sigma)$, giving the Γ_Q-symmetric phase- and flow-diagram shown in the bottom panel of Fig. 2.

An immediate consequence is that fractional plateaux in the doubly degenerate QHE should appear only at $\sigma_H = 2(2n+1)/(2m+1) \neq \pm 1/3$. In fact, $\sigma_H = 1/3$ has also been observed, but only when the magnetic field is so strong that one expects the spin-valley degeneracy to be lifted, which transmutes Γ_Q to Γ_R or Γ_T.

A. IQHE in graphene

Fig. 16 is a reconstruction of some experimental quantum Hall data for graphene,[57–59] compared with modular critical points (blue $\otimes$). As explained in Section 2, ideally we would like to have a family of scaling data deep inside the scaling domain, in which case we could obtain the experimental critical point from the temperature independent crossing point of the curves. Unfortunately such data are still not available for graphene. The family

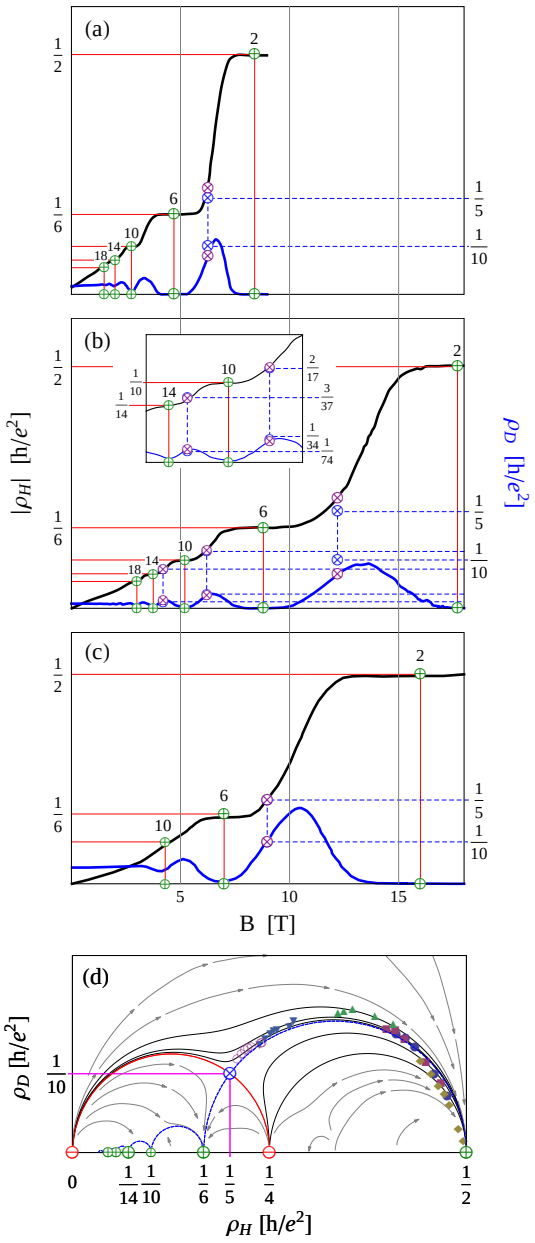

Figure 16. (Color online) (a-c) Experimental quantum Hall data for graphene reconstructed from [57–59], compared with modular critical points (blue ⊗). (d) Scaling flow derived from reconstructed graphene data published in [59], superimposed on the phase diagram with Γ_Q-symmetry (compare Fig. 2).

of data published recently are consistent with our estimate, but not good enough to resolve any discrepancy in detail.[59] This is why only the data obtained at the lowest temperature (4.1 K) has been used in Fig. 16(c).

In lieu of such "family portraits" we use the translational symmetry in the conductivity of the IQHE to estimate the location of experimental critical points. If we are close enough to the scaling domain critical points should lie at the apex of the arcs in the conductivity graph connecting neighbouring plateaux, i.e., where the experimental conductivity graph crosses the vertical lines $\sigma_H = 4n$ (compare Fig. 2). Mapping these points back onto the resistivity data gives the experimental critical points ⊗ (purple icons) shown in Fig. 16, which mostly eclipse the modular predictions ⊗ (blue icons).

Fig. 16 (a) is our reconstruction of the first data on the $2 \overset{\otimes}{\longleftrightarrow} 6$ transition, discovered in 2005.[57,60] Fig. 16 (b) shows the $2 \overset{\otimes}{\longleftrightarrow} 6 \overset{\otimes}{\longleftrightarrow} 10 \overset{\otimes}{\longleftrightarrow} 14$ transitions explored in 2009.[58] The latter two transitions are magnified in the inset, but the distinction between experimental and Γ_Q critical points is still not resolved in this plot. Fig. 16 (c) shows more recent data on the $2 \overset{\otimes}{\longleftrightarrow} 6 \overset{\otimes}{\longleftrightarrow} 10$ transitions.[59] In this case the fixed point of Γ_Q is totally eclipsed by the experimental critical point. In all cases the overlap of experiment and theory is reasonable, and possibly within experimental error, although no error analysis of these experiments has been published.

Fig. 16 (d) shows a scaling flow derived from reconstructed graphene data published in [59], superimposed on the phase diagram with Γ_Q-symmetry (compare Fig. 2).

We see that it is possible that the earliest graphene experiments (Figs. 16 (a-b)) had not yet reached the scaling limit, which is where an approximate low-energy symmetry would appear. The good agreement with the most recent data in Fig. 16 (c) notwithstanding, since these experiments have only probed the IQHE in graphene, it is premature to claim that these experiments unambiguously demonstrate the emergence of a modular symmetry in this material. This question can only be settled by more accurate scaling experiments involving transitions to fractional plateaux.

B. FQHE in graphene

Since they were discovered in 2009 many fractional plateaux have been found in graphene.[62,63] A recent study found some intriguing new fractional plateaux in graphene:[64]

$$\sigma_H = \begin{cases} \frac{1}{3}, \frac{2}{3}, \frac{2}{5}, \frac{3}{5}, \frac{3}{7}, \frac{4}{7}, \frac{4}{9} & \text{for } 0 < \nu < 1 \\ \\ \frac{4}{3}, \frac{8}{5}, \frac{10}{7}, \frac{14}{9} & \text{for } 1 < \nu \end{cases}$$

The first sequence is consistent with Γ_T, in which case both the spin and pseudospin has been resolved. Barring coincidences, the second sequence appears to be constrained to have only even numerators. Since $4/3, 8/5 \notin \oplus_Q$, the only possibility appears to be Γ_R, which has plateaux

$$\oplus_R = \frac{2n}{2m+1} \ni \frac{4}{3}, \frac{8}{5}, \frac{10}{7}, \frac{14}{9} \ldots$$

A possible interpretation is that either the spin or the pseudospin degeneracy has been fully resolved, while

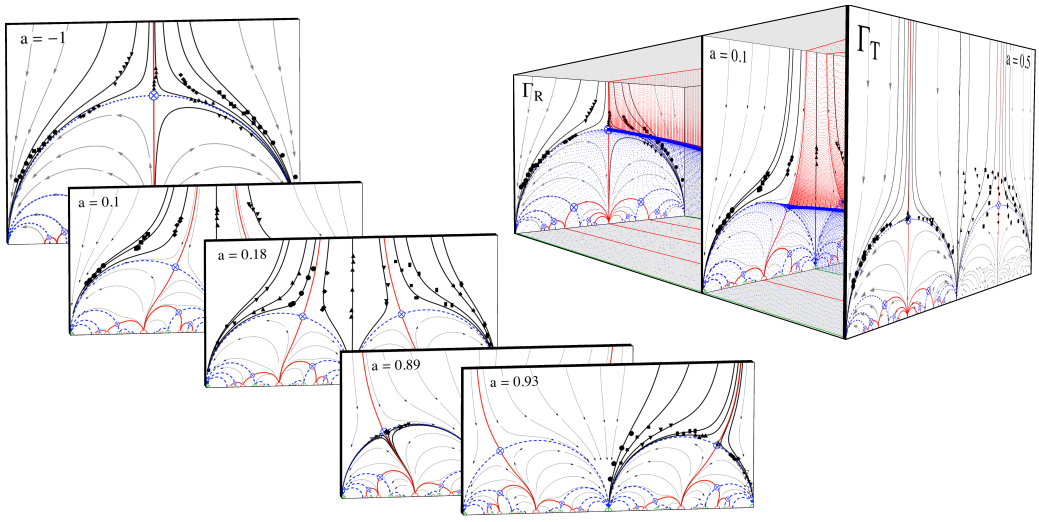

Figure 17. (Color online) Left: Reconstruction of temperature-driven scaling data (discrete icons) exploring various parts of the landscape of $\Gamma(2)$-symmetric scaling flows in $\sigma(t)$, derived from a wide range of different 2D materials (compare Fig. 4).[61] Right: Reconstruction of temperature-driven scaling data (discrete icons) exploring the transitions $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1 \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in GaAs with self-assembled InAs dots, for various values of the spin splitting (parametrized by a), which was tuned using a backgate voltage.[61]

the other remains at least partially degenerate (compare Fig. 5). This is consistent with the expectation that the lowest level will be most susceptible to symmetry-breaking (compare Fig. 20).

VII. TRANSMUTATIONS

When one or more control parameter of an experiment changes, an external electric field, say, a modular symmetry can be transmogrified into another modular symmetry. This "morphing" may be a consequence of some discrete microscopic symmetry being broken, but unless the new modular group is a subgroup of the original group, we will call this a *crossover* or *transmutation*, rather than *symmetry-breaking*. We discuss some examples of this that have been explored experimentally.

A. $\Gamma_R \longrightarrow \Gamma_T$

We turn now to some experiments that have explored the transition from non-relativistic degenerate (unpolarized) to non-degenerate (fully polarized/spin split) bands, by tuning the spin splitting using a backgate voltage. By the arguments discussed in the first section, we expect these data to interpolate between the two maximal submodular symmetries Γ_R (unpolarized) and Γ_T (polarized). When the Zeeman splitting is between these extremes the modular symmetry must be at least partially broken, but possibly only to their maximal common subgroup $\Gamma(2)$ (compare Fig. 1).

The panels inside the box on the right hand side of Fig. 17 shows a reconstruction of temperature-driven scaling data (discrete icons) exploring the transitions $0 = \circledast \overset{\otimes}{\longleftrightarrow} \oplus = 1 \overset{\otimes}{\longleftrightarrow} \oplus = 2$ in GaAs with self-assembled InAs dots.[61] The transition from degenerate (unpolarized) to non-degenerate (fully polarized/spin split) bands is explored by tuning the spin-orbit interaction using a backgate voltage, and compared to the family of physically viable $\Gamma(2)_2$-invariant RG-potentials (compare Sect. II) $\varphi_a \propto \ln\lambda + (a-1)\ln(\lambda-1)$,[22] with values of the real parameter a ranging from $a_R = -1$ to $a_T = 1/2$ in this experiment. All solid lines are flow trajectories derived by numerical integration from the gradient flow generated by this potential. For clarity we display only those parts of the modular phase boundaries (red curtains) that are above all separatrices (blue canopies).

By comparing the data for the $0 \overset{\otimes}{\longleftrightarrow} 1$ transition with the flow derived from $\varphi_{1/2}$ (left front panel), we see that the scaling flow in this case appears to respect Γ_T-symmetry.

This is not so for the $1 \overset{\otimes}{\longleftrightarrow} 2$ transition (right front panel), since several of the experimental flow lines are crossing the separatrix (dashed blue semi-circle). This is a rare example where the $\Gamma(2)_2$-symmetry appears to be broken, presumably due to the intervention of new physics that is not relevant for the other experiments. It is conceivable that some (maximal?) subgroup of $\Gamma(2)_2$ has survived, but we have insufficient data to investigate this. It is also conceivable that a systematic error of unknown of origin is responsible, but we have no way of investigating this either. New physics would be more

interesting, and it could aid in the construction of a phenomenological function $a = a(B, T, \dots)$ (the dots include material properties that are capable of breaking modular symmetry), which could be used to predict which type of modular symmetry (if any) that is to be expected in the transport coefficients of new materials.

It is instructive to examine these flows in more detail, compare panels on the right hand side of Fig. 17, which also includes some other experiments that were discussed in [22]. This reconstruction of temperature-driven scaling data (discrete icons) is derived from a wide range of different 2D materials, and explores various parts of the landscape of $\Gamma(2)$-symmetric scaling flows. As in all our diagrams, solid lines are flow trajectories derived by numerical integration from the gradient flow generated by the RG-potential φ_a.

It is interesting to note the severe deformation of the fractional phases when the symmetry is broken to $\Gamma(2)$. This is a consequence of the dramatic transmogrification that must take place in passing from one sub-family to another (compare Figs. 3 and 4). When fractional quantum Hall data become available for these and similar materials, these predictions will provide a very stringent test of modular symmetry.

B. $\Gamma_{\rm Q} \to \Gamma_{\rm R}$

Fig. 18 illustrates a conjectured modular explanation of a peculiar phenomenon observed in a graphene device.[65] This experiment appears to show that an insulator phase can inject itself into the standard graphene sequence $\cdots -6 \overset{\otimes}{\longleftrightarrow} -2 \overset{\otimes}{\longleftrightarrow} 2 \overset{\otimes}{\longleftrightarrow} 6 \dots$, giving $\cdots -6 \overset{\otimes}{\longleftrightarrow} -2 \overset{\otimes}{\longleftrightarrow} \circledast \overset{\otimes}{\longleftrightarrow} 2 \overset{\otimes}{\longleftrightarrow} 6 \dots$, without being accompanied by other new plateaux. This would mean that the modular symmetry is completely broken.

If a modular symmetry is still at large, then the appearance of an insulator phase at $\oplus = \circledast = 0$ could mean that the original symmetry $\Gamma_{\rm Q}$ has been transmuted into $\Gamma_{\rm R}$ or $\Gamma_{\rm T}$ (compare Fig. 2). Since five of the six peaks in σ_D have roughly the same height $\max(\sigma_D) \approx 1\ [e^2/h]$ (compare inset in Fig. 18), the leading candidate is $\Gamma_{\rm R}$:

$$\otimes_Q = 4n + 2i \xrightarrow{\text{“split”}} \otimes_R = 1 + 2n + i \quad (n \in \mathbb{Z}).$$

If so, there should be structure emerging in the $-2 \overset{\otimes}{\longleftrightarrow} 2$ and $\pm 2 \overset{\otimes}{\longleftrightarrow} \pm 6 \overset{\otimes}{\longleftrightarrow} \pm 10 \dots$ transitions, signalling that new plateaux are germinating at $\oplus = 0 \mod 4$. In this experiment we see two new peaks in σ_D developing near the new plateau at $\oplus = \circledast = 0$, as expected, but no other new plateaux in σ_H are resolved, and the original peaks in σ_D are suppressed rather than split.

A possible explanation is that even if both the new plateaux and the new zeros in σ_D are insufficiently developed to be resolved by this experiment, the new zeros in σ_D that eventually develop at $\oplus = \pm 4, \pm 8, \dots$ may be forcing an observable suppression of the peaks of σ_D.

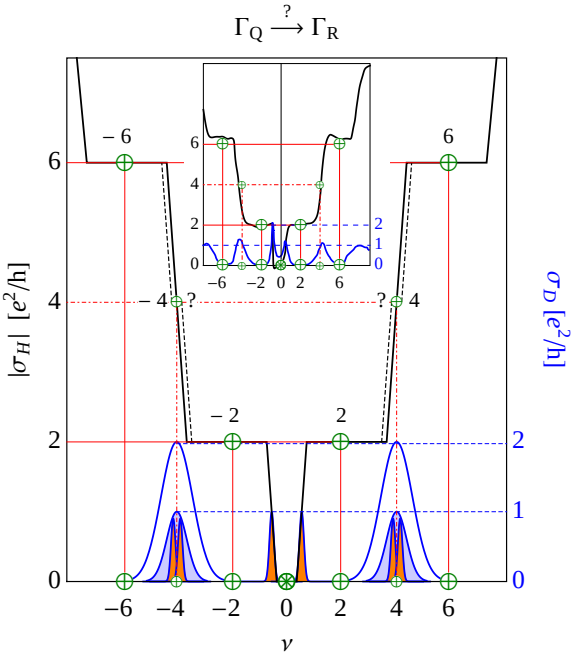

Figure 18. (Color online) Schematic diagram of conductivity data (reconstructed in the inset) as a function of the filling factor $\nu = n_s h/eB$ (B = 18 T, T = 250 mK), from an experiment that perhaps may be interpreted as probing the crossover $\Gamma_{\rm Q} \to (\Gamma_{\rm R}) \to \Gamma_{\rm T}$ in a graphene sample (compare Fig. 5).[65] The diagram illustrates how a suppression of a peak in σ_D may be the first sign that a new phase is germinating, before the new plateau in σ_H or the associated splitting of σ_D can be experimentally resolved. If the magnetic field strength is increased further we expect the band structure to eventually be fully resolved (spin-valley splitting), and a second transmutation $\Gamma_{\rm R} \to \Gamma_{\rm T}$ would suppress the peaks further, $\sigma_D \to 1/2$.

In other words, when a critical point “splits” in order to make room for a new phase, the presence of this new pair of critical points could at first appear as a suppression of the original peak, as is seen in this experiment. When the plateau is fully developed there should be two peaks instead of one, both smaller than the original peak (compare Fig. 2), as seen for the insulator transitions in Fig. 18.

In this experiment, when the insulator phase is present the height of all but one of the peaks in σ_D is consistent with $\Gamma_{\rm R}$-symmetry (compare inset in Fig. 18). This does not explain why the remaining peak is twice as large, but the asymmetry in the data, as well as the crude Hall quantization that only roughly approximates the expected integers ($\oplus = 2 \mod 4$), suggests that there are large systematic errors of unknown origin that may be responsible for skewing the data.

C. $\Gamma_Q \to \Gamma_T$

The valley degeneracy in graphene is protected by an inversion symmetry of the carbon lattice. It seems to be very difficult to lift some or all of this degeneracy without also destroying the Dirac-cones, and the transmutation $\Gamma_Q \longrightarrow \Gamma_S$ has not been seen (compare Fig. 5). We consider first the simplest case where both spin- and valley-symmetry is broken by a very strong magnetic field, which gives the transmutation $\Gamma_Q \longrightarrow \Gamma_T$.

Fig. 19 shows a reconstruction of resistivity data from a pair of graphene experiments that appear to be showing the crossover $\Gamma_Q \to \Gamma_T$ (compare Fig. 5).[35,57] In both experiments the density of states was controlled by a back-gate potential $|V_g| < 80\,V$, in a fixed background magnetic field B and constant temperature T.

With $B = 9\,T$ and $T = 1.6\,K$ at least fourteen plateaux at $\sigma_{\oplus} = 4n + 2$ ($n = -7, -6, \ldots, 5, 6$) were observed, compare Fig. 19 (a), but no insulator phase.[57] This is consistent with Γ_Q-symmetry. The dotted blue lines connected to the stack of unstable fixed points ($\ominus \longrightarrow \otimes \longleftarrow \ominus$) is a visual mnemonic to remind us of the modular quantum critical point expected to appear at $\rho_{\otimes} = i/2\,[h/e^2]$, which was not accessible in this experiment.

In a much stronger magnetic field $B = 45\,T$, and similar temperature $T = 1.4\,K$, the double spin-valley degeneracy appears to be completely broken, compare Fig. 19 (b). In addition to the weak field plateaux (Fig. 19 (a)), new plateaux were observed at $\sigma_{\oplus} = 0, \pm 1, \pm 4\,[e^2/h]$, and ρ_D is showing preliminary signs (splitting and suppression) of additional plateaux germinating at $\sigma_{\oplus} = \pm 3\,[e^2/h]$, and perhaps also at $\sigma_{\oplus} = \pm 5\,[e^2/h]$.[57] This is consistent with the Γ_T-symmetry expected when both the spin and pseudospin degeneracies have been resolved. The dotted lines connecting some plateaux to the insulator fixed point $\rho_{\circledast} = i\infty$ is a reminder that modular symmetry does not predict a quantized value of the Hall potential in this phase ($\rho_H \in \mathbb{R}$). Equivalently, both the Hall and magnetoconductivities vanish at the IR fixed point $\sigma_{\circledast} = 0$ on the boundary of this phase (yellow region in Fig. 2).

In the absence of sufficient information about the geometry of these Hall devices, we have in both diagrams chosen to normalize the magnetoresistance ρ_D so that the principal left peak takes the maximum value expected from modular symmetry (i.e., the height of the relevant semi-circular separatrix, compare top and bottom right panels in Fig. 2). In both cases the sub-leading left peak (and in (b) also the next peak to the left) are in reasonable agreement with the expected (maximal) modular values of ρ_D, indicated here by dashed blue lines. There is, however, a substantial asymmetry between the data on the right and left hand side, which is not expected if the emergent symmetry is fully developed.

Fig. 20 shows a schematic illustration of a conjectured band structure as a function of an external magnetic field strength B, inferred from graphene data obtained at $T = 1.4\,K$ with $B < 45\,T$.[66] Each time a band splits

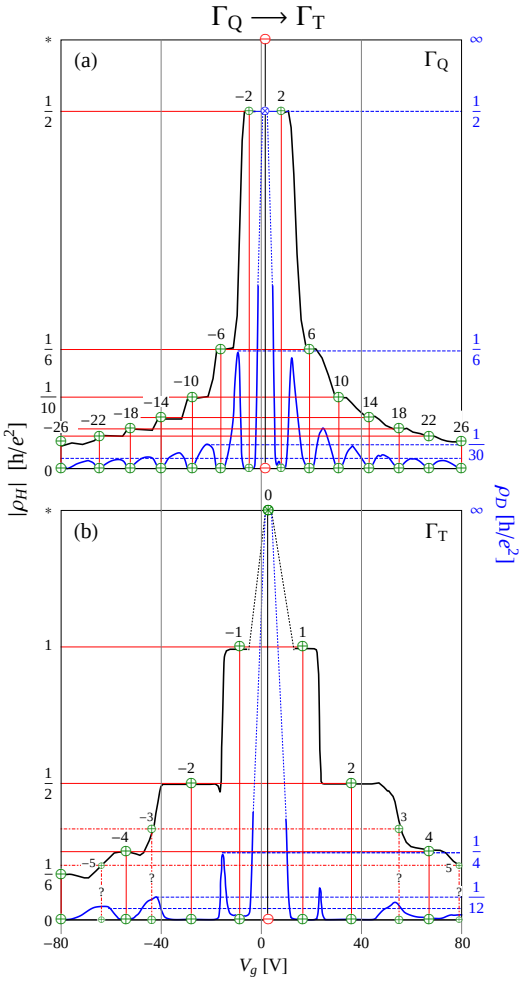

Figure 19. (Color online) Reconstruction of resistivity data (solid curves) from two graphene experiments in a constant external magnetic field. Both the plateaux spectrum and relative heights of the peaks in the magnetoresistivity ρ_D, (a) in a "weak" field $B = 9\,T$,[57] and (b) in a strong field $B = 45\,T$,[35] is consistent with the transmutation $\Gamma_Q \to \Gamma_T$ of modular symmetry (compare Fig. 5).

a new delocalized state appears, giving rise to an additional plateau in the Hall conductivity. Splitting spin states $|n;\uparrow,*\rangle$ and $|n;\downarrow,*\rangle$ gives a blue gap, splitting valley states $|n;*,\Uparrow\rangle$ and $|n;*,\Downarrow\rangle$ gives a green gap. Gaps are uniquely labelled by the filling factor ν.

The weak (left) and strong (right) field limits are consistent with the transmutation $\Gamma_Q \to \Gamma_T$ of modular symmetry (compare Fig. 5). However, if the degeneracy in the lowest Landau level is less robust than for higher levels, so that the splitting happens at different values of the magnetic field B, then at intermediate values of B neither symmetry will be manifest in the spectrum of plateaux.

Tilted field data obtained in this experiments[66] seems to show that spin splitting is easier to achieve than valley splitting in graphene, at least for the lowest Landau levels, so that there is a hierarchy in the spin-valley splitting

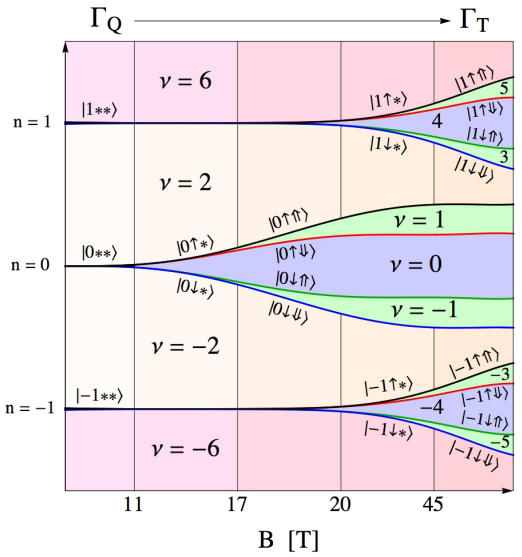

Figure 20. (Color online) Schematic illustration of conjectured band structure of a grapene sample investigated at $T = 1.4\,K$ as a function of an external magnetic field strength $B < 45\,T$.[66] Colored regions labelled by filling factors ν are gaps between the bands, which are represented here by curves labelled with spin-up state vectors $|n;\uparrow\Uparrow\rangle$ (black) and $|n;\uparrow\Downarrow\rangle$ (red), and spin-down states $|n;\downarrow\Uparrow\rangle$ (green) and $|n;\downarrow\Downarrow\rangle$ (blue), for the three Landau levels $n = 0, \pm 1$. The high and low field limits are consistent with respectively Γ_T and Γ_Q.

sequence for increasing values of the external magnetic field (compare Fig. 20) ($[B] = T$):

$$
\begin{aligned}
&\oplus_Q \overset{B<11}{=\!=} \pm 2 \overset{\otimes}{\longleftrightarrow} \pm 6 \overset{\otimes}{\longleftrightarrow} \pm 10 \cdots \xrightarrow{B>11} \\
&0 \overset{\otimes}{\longleftrightarrow} \pm 2 \overset{\otimes}{\longleftrightarrow} \pm 6 \overset{\otimes}{\longleftrightarrow} \pm 10 \cdots \xrightarrow{B>17} \\
&0 \overset{\otimes}{\longleftrightarrow} \pm 1 \overset{\otimes}{\longleftrightarrow} \pm 2 \overset{\otimes}{\longleftrightarrow} \pm 6 \overset{\otimes}{\longleftrightarrow} \pm 10 \cdots \xrightarrow{B>20} \\
&0 \overset{\otimes}{\longleftrightarrow} \pm 1 \overset{\otimes}{\longleftrightarrow} \pm 2 \overset{\otimes}{\longleftrightarrow} \pm 4 \overset{\otimes}{\longleftrightarrow} \pm 6 \overset{\otimes}{\longleftrightarrow} \pm 10 \cdots \xrightarrow{B>45} \\
&\cdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdots \xrightarrow{B\to\infty} \\
&0 \overset{\otimes}{\longleftrightarrow} \pm 1 \overset{\otimes}{\longleftrightarrow} \pm 2 \overset{\otimes}{\longleftrightarrow} \pm 3 \overset{\otimes}{\longleftrightarrow} \pm 4 \overset{\otimes}{\longleftrightarrow} \pm 5 \cdots = \oplus_T,
\end{aligned}
$$

where we have allowed for the possibility (not shown in Fig. 20) that there are additional levels in the hierarchy if higher Landau levels are not equally robust.

Since the valley degeneracy is protected by inversion symmetry of the graphene lattice, which is not broken by the external field, the valley splitting presumably depends on the subtle energetics of many-body interactions in this material. A detailed understanding of this hierarchy is therefore a difficult dynamical problem that is only tractable in limiting cases (e.g. $0 \leftarrow B \to \infty$), where the identification of emergent modular symmetries may be useful.

VIII. DISCUSSION

We have reviewed experimental quantum Hall scaling data from a wide selection of materials, and compared these with modular flow diagrams derived from the *holomorphic Hall potential* $\varphi_a = \ln\lambda + (a-1)\ln(\lambda-1)$ that is parametrized by a single real number a.[22] This potential is invariant under the congruence subgroup $\Gamma(2)$ of the full modular group $\mathrm{SL}(2,\mathbb{Z})$, because it is built only from the classical *elliptic modular lambda function* λ.[67]

In Sects. I and II we have explained and emphasized that the physical properties that must be required of any $\Gamma(2)$-invariant scaling function $\beta_a = \partial\varphi_a$ renders the functional form of the RG-potential φ_a essentially unique (up to normalization). The key is to recognize the holomorphic modular structure of the parameter space, which must be respected by the scaling functions. It is the pincer movement of a complex structure and modular symmetry that pins down the RG-potential.[68]

For $a = -1, 1/2, 2$ the symmetry is enhanced to the maximal subgroups Γ_R, Γ_T and Γ_S, respectively. So, if the emergent symmetry is observed to be Γ_X (X = Q, R, S, T), then the β-function, and therefore the phase- and flow-diagram, *is unique* (up to normalization). Since these are the symmetries that are most often encountered in the QHE, it is easy to falisfy the modular hypothesis (i.e., the relevance of φ_a for the QHE), but this is not what has happened. As technology has improved over the past three decades, so has the agreement between experimental scaling data and modularity. Compare, for example, our reconstruction of data from 1985 shown in Fig. 6 with Fig. 10 (data from 2014) or Fig. 12 (data from 2015).

In addition to the fact that some of the modular predictions from 1992 have been verified at the *per mille* level,[1,2,20,24] it is perhaps the overall agreement of the unique modular family of level two flow-diagrams with a wide range of different materials and experimental circumstances that is the most convincing evidence for "modular universality" in the QHE.

There have been various attempts over the years to analyze the phase structure of the QHE, starting with a proposal based on the translation symmetry of RG-flows in the IQHE.[69] This was motivated by a sigma-model of localization,[70–77] but the target space geometry does not appear to be rich enough to include the FQHE.

In [1 and 2] it was proposed that a modular symmetry would be capable of describing both the integer and the fractional Hall effects, by including dualities in addition to translations, as described in the introduction. The three maximal subgroups Γ_R, Γ_S and Γ_T were immediately identified as the largest symmetries of physical interest. Γ_T was shown to give the correct phenomenology for the spin-polarized QHE, Γ_S was proposed as the relevant symmetry for analogous transport problems with bosonic quasi-particles, and the symmetry $\Gamma(2)$ and the idea of a potential interpolating between the enhanced

symmetries was introduced.[4]

At roughly the same time superficially similar dualities acting on the filling fraction ν were considered,[78–83] and the resulting transformations are known collectively as "the law of corresponding states". Since ν is essentially the plateau value $\sigma_\oplus = \sigma_H \in \mathbb{R}$, this approach is oblivious to the complex structure that gives modular symmetry most of its predictive power. These dualities appear to disagree with experiment, unlike the complexified duality identified in [1 and 2], which is in excellent agreement with available data.[17,19,20,24]

There have been two other attempts to construct families of interpolating β-functions for the QHE.[5,6,9–11,14,21] They have both retained the original idea that the β-function should be modular form of weight two,[1,2,7,8,15] and their work looks superficially similar to ours. This is because a flow line derived from *any* function with maximal modular symmetry always has the same shape, so plots of the vector fields will appear to be identical. As was explained in [1–4] this is a mathematical trick, which we also employed in order to obtain the original phase diagrams, but by itself this is not sufficient to build a physical model. Subsequent alternative proposals did not heed this advice. They did not pay sufficient attention to the physical properties that critical points must have, nor did they consider the elementary but demanding experimental constraints on scaling that we discussed in the introduction. Consequently, the conjectured "β-functions" are not well motivated, nor do they appear to have any reasonable physical interpretation, as we now explain.

An interpolating β-function, which in our notation is $\beta_{\tilde{a}} \propto (\tilde{a}\,\lambda - 1)/\theta_3^4$ was proposed in [5, 6, 9, and 14]. They conclude that the location of the zero of this function is not predicted by $\Gamma(2)$ symmetry, which is correct since it depends on the free parameter $\tilde{a}$, similar to our a. But they seem ambivalent about the order of the transition point, they do not consider the family of functions under deformations of $\tilde{a}$, and they do not discuss the points of enhanced symmetry.

A family of meromorphic functions that was conjectured to be an interpolating β-function was postulated in [11 and 21]. However, these "scaling functions" have no physical foundation or interpretation. In particular, they have poles where there should be critical points, i.e., where a physical β-function must vanish. Critical exponents are therefore ill defined, and there is no physical scaling.

These attempts to implement the modular ideas introduced in [1–4] have all failed for the same reason: they ignored the complex structure and physical requirements that the scaling function must respect, i.e., *the geometric structure of the effective (emergent) theory.*

This is in sharp contrast to the work discussed here. Since our β-function is derived from a physically motivated and physically sensible potential, it is by construction well behaved everywhere on the interior of parameter space. This includes the critical points where it has simple zeros, and therefore well defined scaling and critical exponents. The reason that the "mock β-functions" used in [1–4] turned out to be essentially the same as $\partial\varphi_a$, is that they are subjugated by the same holomorphic modular structure. Given this geometric structure the fate of the model is sealed, and the modular predictions made in 1992 are either right or wrong, at least for the maximally symmetric cases that are most frequently encountered in experiments. Fortunately, as we have seen here, there is now a large and growing body of evidence in good agreement with many modular scaling predictions.

We have also discussed some experiments where the modular symmetry is transmuted or broken. We have described in detail the most benign scenario, where the symmetry is "morphing" between maximally symmetric cases, while maintaining a maximal amount of unbroken modular symmetry. This is captured by a unique interpolating family of scaling diagrams with minimal symmetry breaking. The simplest physical example is when cranking up the magnetic field changes an unpolarized (spin degenerate) QHE to a polarized QHE. This is consistent with some experiments, but in a few cases modular symmetry is more severely broken or hidden. It is an important outstanding problem to find out when and how this happens, not only to delineate the domain where emergent modular symmetries are relevant, but also in order to understand why they are so accurate when they do appear.

In summary, all scaling properties of almost all quantum Hall systems seems to be encoded in the deceptivly simple looking RG-potential φ_a.[22]

Perhaps the biggest outstanding problem in the QHE is to determine the value(s) of the critical (delocalization) exponent(s), which would completely nail down the quantum Hall universality class(es). This exponent is determined by the curvature of the RG-potential at a critical point, and therefore depends on the normalization of the β-function. This number does not follow from symmetry alone, and information about the dynamics of the collective (emergent) modes relevant at low energy is required. The Ising model provides a helpful analogy. Kramers and Wannier managed to calculate the exact value the critical temperature (location of the critical point) by exploiting a $\mathbb{Z}_2$-duality that is similar to S-duality acting on σ_D, but the exact value of the critical exponent remained beyond reach until Onsager solved the model completely.

Similarly, modular symmetry is by itself not sufficient to find the low-energy effective field theory, but (unlike $\mathbb{Z}_2$) because it is an infinite non-abelian group it does severly limit the supply of candidate models, and may therefore provide valuable assistance in the search for this theory.

[1] C. A. Lütken and G. G. Ross, "Duality in the quantum Hall system," Phys. Rev. B **45**, 11837 (1992).

[2] C. A. Lütken and G. G. Ross, "Delocalization, duality, and scaling in the quantum Hall system," Phys. Rev. B **48**, 2500 (1993).

[3] C. A. Lütken, "Global phase diagrams for charge transport in two dimensions," J. Phys. A: Math. Gen. **26**, L811 (1993).

[4] C. A. Lütken, "Geometry of renormalization group flows constrained by discrete global symmetries," Nucl. Phys. B **396**, 670 (1993).

[5] Y. Georgelin and J.C. Wallet, "Group $\Gamma(2)$ and the fractional quantum Hall effect," Phys. Lett. A **224**, 303 (1997).

[6] Y. Georgelin, , T. Masson, and J.C. Wallet, "Modular groups, visibility diagram and quantum Hall effect," J. Phys. A: Math. Gen. **30**, 5065 (1997).

[7] C. P. Burgess and C. A. Lütken, "One-dimensional flows in the quantum Hall system," Nucl. Phys. B **367**, 500 (1997).

[8] C. P. Burgess and C. A. Lütken, "On the implications of discrete symmetries for the β-function of quantum Hall systems," Phys. Lett. B **451**, 365 (1999).

[9] Y. Georgelin, , T. Masson, and J.C. Wallet, "$\Gamma(2)$ modular symmetry, renormalization group flow and the quantum Hall effect," J. Phys. A: Math. Gen. **33**, 39 (2000).

[10] Y. Georgelin, , T. Masson, and J.C. Wallet, "Visibility diagrams and experimental stripe structures in the quantum Hall effect," J. Phys. A: Math. Gen. **33**, 8649 (2000).

[11] B. P. Dolan, "Duality in the quantum Hall effect: The role of electron spin," Phys. Rev. B **62**, 10278 (2000).

[12] C. P. Burgess and B. P. Dolan, "Particle-Vortex Duality and the Modular Group: Applications to the Quantum Hall Effect and Other 2-D Systems," Phys. Rev. B **63**, 155309 (2001).

[13] C.P. Burgess and B.P. Dolan, "Duality and nonlinear response for quantum Hall systems," Phys. Rev. B **65**, 155323 (2002).

[14] F. Chandelier, Y. Georgelin, , T. Masson, and J.C. Wallet, "Global quantum Hall phase diagram from visibility diagrams," Phys. Lett. A **301**, 451 (2002).

[15] C. A. Lütken, "Holomorphic anomaly in the quantum Hall system," Nucl. Phys. B **759** [**FS**], 343 (2006).

[16] C. A. Lütken and G. G. Ross, "Anti-holomorphic scaling in the quantum Hall system," Phys. Lett. A **356**, 382 (2006).

[17] C. A. Lütken and G. G. Ross, "Geometric scaling in the quantum Hall system," Phys. Lett. B **653**, 363 (2007).

[18] C. A. Lütken and G. G. Ross, "Implications of experimental probes of the RG-flow in quantum Hall systems," (2009), arXiv:0906.5551v1 [cond-mat.other].

[19] C. A. Lütken and G. G. Ross, "On the origin of duality in the quantum Hall system," Phys. Lett. A **374**, 4700 (2010).

[20] C. A. Lütken and G. G. Ross, "Experimental probes of emergent symmetries in the quantum Hall system," Nucl. Phys. B **850** [**FS**], 321 (2011).

[21] B. P. Dolan, "Holomorphic and anti-holomorphic conductivity flows in the quantum Hall effect," J. Phys. A **44**, 175001 (2011).

[22] J. Nissinen and C. A. Lütken, "Renormalization-group potential for quantum Hall effects," Phys. Rev. B **85**, 155123 (2012).

[23] J. Nissinen and C. A. Lütken, "The quantum Hall curve," (2012), arXiv:1207.4693v1 [cond-mat.str-el].

[24] C. A. Lütken and G. G. Ross, "Quantum criticial Hall exponents," Phys. Lett. A **378**, 262 (2014).

[25] C. A. Lütken, "Infinite symmetry in the quantum Hall effect," EPJ Web of Conferences (2014), http://dx.doi.org/10.1051/epjconf/20147100079. 2nd International conference on new frontiers in physics; 2013-08-28 - 2013-09-05.

[26] C. A. Lütken, "Introduction to the role of modular symmetries in graphene and other two-dimensional materials," Contemporary Physics (2014), http://dx.doi.org/10.1080/00107514.2014.949445.

[27] This has implications for the critical exponents (flow rates near saddle points), which are in agreement with experimental observations [16 and 24].

[28] Not least of which is that the QHE implies Fermat's Last Theorem. A counter-example would allow the construction of an elliptic curve whose properties imply that a 2-form on $\Gamma_0(2) = \Gamma_{\rm T}$ must be a cusp form.[29] This contradicts the physics of the spin-polarized QHE, which has this symmetry, because the β-function is not a cusp form.

[29] F. Diamond and J. Schurman, *A First Course in Modular Forms*, Graduate Texts in Mathematics; 228 (Springer Verlag, New York, NY, USA, 2005).

[30] Other names for these groups are $\Gamma_0(2) = \Gamma_{\rm T}$, $\Gamma^0(2) = \Gamma_{\rm R}$, and $\Gamma_\theta(2) = \Gamma_{\rm S}$. Our simplified notation is similar to [33].

[31] The number of weight two forms on the principal congruence subgroup $\Gamma(N)$ is 2 for $\Gamma(2)$, and $(N+6)N^2/24 \times \prod_{p|N}(1-p^{-2})$, for primes p that divide $N > 2$.[29] The subscript will usually be suppressed, except when it is important to recall the dimension of this vector space.

[32] H. Petersson, "Über die kongruenzgruppen der Stufe 4," JRAM **212**, 63 (1963).

[33] R. Rankin, *Modular Forms and Functions* (Cambridge University Press, Cambridge, 1977).

[34] D. Zagier, *The 1-2-3 of Modular Forms*, Annals of Mathematics Studies No. 97 (Springer Verlag, 2008) Lectures at a Summer School in Nordfjordeid, Norway.

[35] Y. Zhang, Z. Jiang, J.P. Small, M.S. Purewal, Y.-W. Tan, M. Fazlollahi, J.D. Cudlow, J.A. Jaszczak, H.L. Stormer, and P. Kim, "Landau-Level Splitting in Graphene in High Magnetic Fields," Phys. Rev. Lett. **96**, 136806 (2006).

[36] J.G. Checkelsky, L. Li, and N.P. Ong, "Zero-Energy State in Graphene in a High Magnetic Field," Phys. Rev. Lett. **100**, 1206801 (2008).

[37] The idea of finding a C-function that interpolates between maximal symmetries first appeared in [4], but only phase diagrams for $a = -1, 1/2, 1$ and 2 were discussed in detail.

[38] E. Fradkin, *Field Theories of Condenced Matter Physics* (Cambridge university press, 2013).

[39] H.P. Wei, D.C. Tsui, and A.M.M. Pruisken, "Localization and scaling in the quantum Hall regime," Phys. Rev. B **33**, 1488 (1985).

[40] M. Z. Hasan and C. L. Kane, "Topological insulators," Rev.Mod.Phys. **82**, 3045 (2010).

[41] S.S. Murzin, S.I. Dorozhkin, D.K. Maude, and A.G.M. Jansen, "Scaling flow diagram in the fractional quantum Hall regime of $GaAs/Al_xGa_{1-x}As$ heterostructures," Phys. Rev. B **72**, 195317 (2005).

[42] W. Li, C. L. Vicente, J. S. Xia, W. Pan, D. C. Tsui, L. N. Pfeiffer, and K. W. West, "Scaling in Plateau-to-Plateau Transition: A Direct Connection of Quantum Hall Systems

with the Anderson Localization Model," Phys. Rev. Lett. **102**, 216811 (2009).
[43] J. Wang, B. Lian, and S.-C. Zhang, "Quantum anomalous Hall effect in magnetic topological insulators," Phys. Scr. T **164**, 014003 (2015).
[44] Y. Ando, "Topological insulator materials," J. Phys. Soc. Jpn. **82**, 102001 (2013).
[45] J.G. Checkelsky, R. Yoshimi, A. Tsukazaki, K.S. Takahashi, Y. Kozuka, J. Falson, M. Kawasaki, and Y. Tokura, "Trajectory of the anomalous Hall effect towards the quantized state in a ferromagnetic topological insulator," Nature Physics **10**, 731 (2014).
[46] M. Konig, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, "Quantum Spin Hall Insulator State in HgTe Quantum Wells," Science **318**, 766–770 (2007).
[47] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, "Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells," Science **314**, 1757–1761 (2006).
[48] E. B. Olahanetsky, S. Sassine, Z. D. Kvon, N. N. Mikhailov, S. A. Dvoretsky, J. C. Portal, and A. L. Aseev, "Quantum Hall liquid-insulator and plateau-to-plateau transitions in a high mobility 2DEG in a HgTe quantum well," Pis'ma v Zh. ksper. Teoret. Fiz. **84**, 661–665 (2006).
[49] Y. G. Arapov, S. V. Gudina, V. N. Neverov, S. M. Podgornykh, M. R. Popov, G. I. Harus, N. G. Shelushinina, M. V. Yakunin, N. N. Mikhailov, and S. A. Dvoretsky, "Temperature Scaling in the Quantum-Hall-Effect Regime in a HgTe Quantum Well with an Inverted Energy Spectrum," Semiconductors **49**, 1545–1549 (2015).
[50] S. V. Gudina, Y. G. Arapov, V. N. Neverov, S. M. Podgornykh, M. R. Popov, N. G. Shelushinina, M. V. Yakunin, S. A. Dvoretsky, and N. N. Mikhailov, "2D-localization and delocalization effects in quantum Hall regime in HgTe wide quantum wells," Phys. Status Solidi C **13**, 473–476 (2016).
[51] Y.G. Arapov, S.V. Gudina, V.N. Neverov, S.M. Podgornykh, M.R. Popov, G.I. Harus, N.G. Shelushinina, M.V. Yakunin, N.N. Mikhailov, and S.A. Dvoretsky, "Temperature scaling in the quantum-Hall-effect regime in a HgTe wide quantum well with an inverted energy spectrum," Semiconductors **49**, 1545 (2015).
[52] R. Yoshimi, A. Tsukazaki, Y. Kozuka, J. Falson, K.S. Takahashi, J.G. Checkelsky, N. Nagaosa, M. Kawasaki, and Y. Tokura, "Quantum Hall effect on top and bottom surface states of topological insulator $(Bi_{1-x}Sb_x)_2Te_3$ films," Nat. Commun. 6:6627 doi: 10.1038/ncomms7627 (2015).
[53] L. Li, F. Yang, G. J. Ye, Z. Zhang, Z. Zhu, W. Lou, X. Zhou, L. Li, K. Watanabe, T. Taniguchi, K. Chang, Y. Wang, X. H. Chen, and Y. Zhang, "Quantum Hall effect in black phosphorus two-dimensional electron system," Nature Nanotechnology **11**, 593–597 (2016).
[54] S.S. Murzin, M. Weiss, A.G.M. Jansen, and K. Eberl, "Universal flow diagram for the magnetoconductance in disordered GaAs layers," Phys. Rev. B **66**, 233314 (2002).
[55] L. I. Huang, Y. Yang, R. E. Elmquist, S. T. Lo, F. H. Liu, and C. T. Liang, "Insulator-quantum Hall transition in monolayer epitaxial graphene," RSC Adv. **6**, 71977 (2016).
[56] V.P. Gusynin and S.G. Sharapov, "Unconventional Integer Quantum Hall Effect in Graphene," Phys. Rev. Lett. **95**, 146801 (2004).
[57] Y. Zhang, H.L. Stormer, and P. Kim, "Experimental observation of the quantum Hall effect and Berry's phase in graphene," Nature **438**, 201 (2005).
[58] X. Wu, M. Ruan Y. Hu, N.K. Madiomanana, J. Hankinson, M. Sprinkle, C. Berger, and W.A. de Heer, "Half integer quantum Hall effect in high mobility single layer epitaxial graphene," Appl. Phys. Lett. **95**, 223108 (2009).
[59] M. Amado, E. Diez, F. Rossella, V. Bellani, D. Lopez-Romero, and D.M. Maude, "Magneto-transport of graphene and quantum phase transitions in the quantum Hall regime," J. Phys.: Condens. Matter **24**, 305302 (2012).
[60] K.S. Novoselov, A. Geim, S. Morozov, D. Jiang, M.I. Katsnelson, V. Grigorieva, and S.V. Dubonos, "Two-dimensional gas of massless Dirac fermions in graphene," Nature **438**, 197 (2005).
[61] Y.-T. Wang, G-H. Kim, C. F. Huang, S.-T. Lo, W.-J. Chen, J. T. Nicholls, L.-H. Lin, D. A. Ritchie, Y. H. Chang, C.-T. Liang, and B. P. Dolan, "Probing temperature-driven flow lines in a gated two-dimensional electron gas with tunable spin-splitting," J. Phys.: Condens. Matter **24**, 405801 (2012).
[62] X. Du, I. Skachko, F. Duerr, A. Lucian, and E.Y. Andrei, "Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene," Nature **462**, 192 (2009).
[63] K.I. Bolotin, F. Ghahari, M.D. Schulman, H.L.Stormer, and P. Kim, "Observation of the fractional quantum Hall effect in graphene," Nature **462**, 196 (2009).
[64] B.E. Feldman, B. Krauss, J.H. Smet, and A.Yacoby, "Unconventional Sequence of Fractional Quantum Hall States in Suspended Graphene," Science **337**, 1196 (2012).
[65] L. Zhang, Y. Zhang, M. Khodas, T. Valla, and I. A. Zaliznyak, "Metal to Insulator Transition on the N = 0 Landau Level in Graphene," Phys. Rev. Lett. **105**, 046804 (2010).
[66] Z. Jiang, Y. Zhang, H.L. Stormer, and P. Kim, "Quantum Hall States near the Charge-Neutral Dirac Point in Graphene," Phys. Rev. Lett. **99**, 106802 (2007).
[67] The $\Gamma(2)$-invariant holomorphic function $\lambda = \theta_2^4/\theta_3^4$ is also known as the *hauptmodul* for the *modular curve* $X(2) = \mathbb{C}^+/\Gamma(2)$.
[68] It is highly unusual to have a universal form of the RG potential beyond leading order in an expansion near a critical point. The scheme independence encountered here is a consequence of the draconian constraints enforced by the holomorphic modular structure.
[69] D.E. Khmel'nitzkii, "Quantization of Hall conductivity," JETP Lett. **38**, 552 (1983).
[70] H. Levine, S. B. Libby, and A. M. M. Pruisken, "Electron Delocalization by a Magnetic Field in Two Dimensions," Phys. Rev. Lett. **51**, 1915 (1983).
[71] A. M. M. Pruisken, "On localization in the theory of the quantized Hall effect: A two-dimensional realization of the θ-vacuum," Nucl. Phys. B **235**, 277 (1984).
[72] H. Levine, S. B. Libby, and A. M. M. Pruisken, "Theory of the quantized Hall effect (I)," Nucl. Phys. B **240**, 30 (1984).
[73] H. Levine, S. B. Libby, and A. M. M. Pruisken, "Theory of the quantized Hall effect (II)," Nucl. Phys. B **240**, 49 (1984).
[74] H. Levine, S. B. Libby, and A. M. M. Pruisken, "Theory of the quantized Hall effect (III)," Nucl. Phys. B **240**, 71 (1984).
[75] R. B. Laughlin, M. L. Cohen, J. M. Kosterlitz, H. Levine, S. B. Libby, and A. M. M. Pruisken, "Scaling of conduc-

tivities in the fractional quantum Hall effect," Phys. Rev. B **32**, 1311 (1985).
[76] A. M. M. Pruisken, "Dilute instanton gas as the precursor to the integral quantum Hall effect," Phys. Rev. B **32**, 2636 (1985).
[77] A. M. M. Pruisken, "Universal Singularities in the Integral Quantum Hall Effect," Phys. Rev. Lett. **61**, 1297 (1988).
[78] S.-C. Zhang, T. H. Hansson, and S. Kivelson, "Effective-Field-Theory Model for the Fractional Quantum Hall Effect," Phys. Rev. Lett. **62**, 82 (1989).
[79] D.-H. Lee, S. Kivelson, and S.-C. Zhang, "Quasi-particle charge and activated conductance of a quantum Hall liquid," Phys. Rev. Lett. **68**, 2386 (1992).
[80] S. Kivelson, D.-H. Lee, and S.-C. Zhang, "Global phase diagram in the quantum Hall effect," Phys. Rev. B **46**, 2223 (1992).
[81] E. Shimshoni, S. L. Sondhi, and D. Shahar, "Duality near quantum Hall transitions," Phys. Rev. B **55**, 13730 (1997).
[82] X.G. Wen and A. Zee, "Classification of Abelian quantum Hall states and matrix formulation of topological fluids," Phys. Rev. B **46**, 2290 (1992).
[83] J.K. Jain, "Theory of the fractional quantum Hall effect," Phys. Rev. B **41**, 7653 (1990).

Chapter 6

Conclusions and Outlook

As displayed in the paper, experimental measurements (Figs. 6-17) generally fit the hypothesis of modular β-functions well. It is also worth noting that over the years, as measuring equipment has become more accurate, the the data seems to fit *better*, compare e.g. Fig. 6 (1985) to Fig. 8 (2009) and Fig. 11 (2016). This observation is a good argument in favor of our hypothesis.

The shape of the phase diagram generated from β_a is very specific and rigid: there is no parameter in the theory, other than a, that may be tuned in order to fit experimental data points.

In only one experiment was it the case that no value of a could be found, fitting the data points to flow lines in a satisfying way: see the right front panel of the right hand side of Fig. 17 in the paper. Note also that the flow in that case significantly crosses the separatrix. However, closer inspection of the paper presenting those data [97], reveals that they are based on the experimental plots portrayed in Fig. 6.0.1. Judging from how ρ_{xx} fails to reach zero, we may conclude that this experiment was far from the scaling regime. In that case, the data is not expected to fit the modular flow.

In all the experimental data analysed, the assumption that all sources are located on the real line fits well. That is, there is no indication that sources should be placed in the interior of $\mathbb{H}$. Then, regardless of the validity of modular symmetry, a phase diagram structure placing sources on the real line should be preferred, contrary to the model presented by Laughlin et. al. [96].

An issue with the modular symmetry arises from the fixed point classification of sources and sinks performed in subsection 4.4.3. There it was determined that for every congruence group, the fractions on the form (odd number)/(even number) are repulsive. However, Willett *et. al.* found a plateau at $\nu = 5/2$ in an GaAs/AlGaAs structure of electron mobility $\mu \sim 1.3 \times 10^6 \text{cm}^2/\text{Vs}$ in 1987 [98].

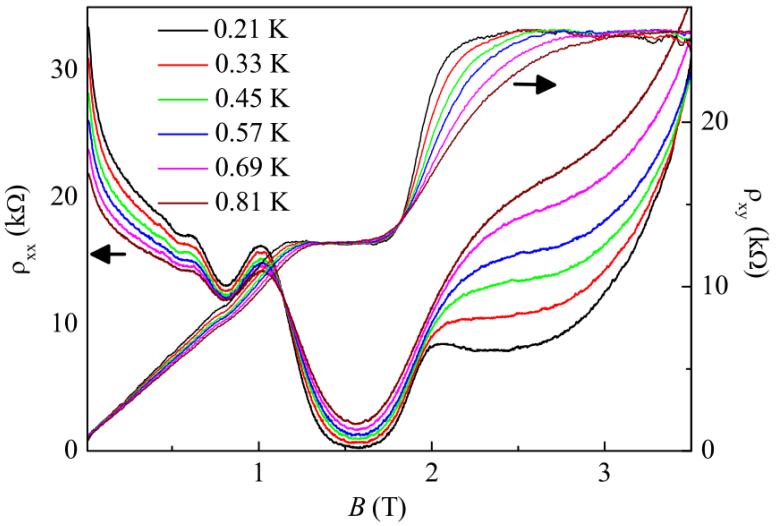

Figure 6.0.1: The experimental data from [97]. Note how the longitudinal resistivity fails to reach zero.

In that experiment, the plateau was not highly resolved, however, experiments in samples with higher electron mobility ($\mu = 1.7 \times 10^7 \text{cm}^2/\text{Vs}$) has yielded an unmistakeable plateau and vanishing longitudinal resistivity at $\nu = 5/2$ [99].

There have also been reports of a $\nu = 1/2$ plateau [100] [101], a $\nu = 3/2$ plateau [102], and a $\nu = 7/2$ plateau [103]. Most striking is perhaps the discovery of a plateau at $\nu = 2 + 3/8$, which nobody seems to be able to explain [104].

These even-denominator plateaux pose a problem for the modular hypothesis, as the modular RG flow strictly prohibits them. They do, however, only appear to be resolved for special experimental conditions, such as extremely high mobilities and sometimes tilted magnetic fields. As such, the modular framework still works for the "simplest" QHEs, although it can not account for the most general cases of the FQHE.

All in all, the data from the integer sequences $\nu = 0, \pm 1, \pm 2, ...$ and $\nu = 0, \pm 2, \pm 4...$ fits well into the modular RG flow, however, more data is always welcome. Especially temperature scaling data for the FQHE is sparse, more experiments done in this regime would help illustrate the unification of IQHE and FQHE that modular symmetries predict. It would be of particular interest to obtain scaling data for the even denominator plateaux, so one can study exactly how this flow breaks with the modular hypothesis.

Also, not many of the (much anticipated) scaling experiments with graphene have been performed. These are needed to properly test the hypothesis of Γ_Q-symmetric β-functions for the $\nu = \pm 2, \pm 6, ...$ sequence. The future also holds promise for graphene's 2-dimensional cousins: silicene, germanene and stanene.

Indeed, Dirac cones have been observed in silicene [105], germanene [106] and stanene [107]. Quantum Hall experiments in these materials are still in their infancy, and there are issues with substrate induced elimination of the Dirac cones (see [107] and the references therein). An interesting task for the future is to classify the scaling flow for this family of relativistic novel 2-dimensional materials.

More scaling experiments with tunable spin-splitting would be interesting. With enough data, an empirical formula for the parameter a could possibly be found. As seen in subsection 4.6.2, an interpolation between the maximal symmetries was proposed in [95], using $g_{\text{eff}} \sim (g/\cos\theta)T^{-\kappa}$ as the parameter, θ denoting the angle between the sample surface normal and the magnetic field. However, in the experiments producing data that required a value of $a \notin \{-1, 1/2, 2\}$, used in our paper (Fig. 17), the magnetic field was parallell to the sample surface normal ([97], [108]). This indicates that the parametrization proposed in [95] can not be the whole truth.

Indeed, the parametrization of [95] ends at Γ_T symmetry with $g_{\text{eff}} = \infty$ (compare Fig. 4.6.3), corresponding to $a = 1/2$. In the two front panels of the left hand side of Fig. 17 in our paper, however, experimental data fitting $a = 0.89$ and $a = 0.93$ is presented.

The parameter g_{eff} of [95] gives us a starting point for which quantities the parameter a might depend on, but more scaling experiments under well documented conditions are needed for a complete analysis.

Finally, in this thesis there has been no attempt of presenting a mechanism for how modular symmetries can emerge from the microphysics of the QHE. This is an unresolved topic in the field. A hope for the future is that someone will accept this daunting challenge, as the determination of such a mechanism would firmly establish a theoretical footing for this theory.

Appendices

Appendix A

Elements of Topology and Group Theory

Topological field theories, by their name, require some knowledge about the mathematical field of topology. This appendix is intended for readers unfamiliar with topology, to get a notion of certain phrases used in the topological arguments of the quantum Hall effect. For further information on topology, see e.g. [109] and [110], which we follow here.

Group theory is by itself important in physics where groups are used to classify symmetries and the action of symmetries on on coordinates or parameters are realized by the group action. Some aspects of group theory is also needed in order to understand the terminology of section REF and chapter REF, and this will also be revised here. A good introduction to discrete groups is found in [92]. Sufficient information about the continuous Lie Groups most useful for special relativity and gauge theory can be found in section 15.4 of [4]. We start off with some definitions. A notion of the concept of a *set* is required.

Relational operators: $\forall$ denotes "for all" or "for each". $\exists$ denotes "there exists".

The **union** of two sets, A and B, is defined as the set of all elements that are in either A or B:

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

The **intersection** of two sets, A and B, is the set of all elements that are in both A and B:

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

The **empty set** is the set containing no elements: $\emptyset = \{\}$.

A map $f : A \to B$ is said to be **surjective** (onto) if the image $f(A) = B$. It is

said to be **injective** (one-to-one) if there for each $b \in B$ exists exactly one $a \in A$ such that $f(a) = b$. It is said to be **bijective** if it is both injective and surjective. If f is bijective, the inverse map f^{-1} is automatically bijective.

Group Theory

A **binary operation** on a set A is a function $* : A \times A \to A$, taking two elements of A and returning an element of A.

A **group** is a set G, with a binary operation $*$ such that

- $\forall\ g_1, g_2, g_3 \in G$, we have $(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$ ($*$ is associative).
- $\exists\ e \in G$ such that $\forall g \in G$, we have $e * g = g * e = g$ (existence of an identity element).
- $\forall\ g \in G\ \exists\ g^{-1} \in G$ such that $g * g^{-1} = g^{-1} * g = e$ (existence of inverses).

The group is denoted $(G, *)$. When the binary operation is implied or unimportant, G by itself referres to the group.

A **subgroup** H of a group G is any subset of G, closed under the binary operation, so that $* : H \times H \to H$, and such that H with the operation $*$ is itself a group.

For a subgroup $H \subset G$, and an element $g \in G$, define the **left coset** of H containing g as $gH = \{gh \mid h \in H\} \subset G$. In general, many elements of G will belong to the same coset of H, and we define the **index** of H in G as the number of left cosets of H in G. The index is then a measure of how large the group H is with respect to G: lower index means that fewer cosets of H are needed to cover G.

A **homomorphism** of groups $(G, *)$ and $(G', \circ)$ is a map $f : G \to G'$ such that $f(g_1 * g_2) = f(g_1) \circ f(g_2)$, i.e. a structure preserving map. If there exists a homomorphism between two groups they are said to be homomorphic. An **isomorphism** is a bijective homomorphism. Likewise, if there exists an isomorphism between two groups they are said to be isomorphic. Isomorphic groups are for all practical purposes the *same* group.

If $\forall\ g_1, g_2 \in G$ we have $g_1 * g_2 = g_2 * g_1$, the group is called **Abelian**. An example is the group of integers $\mathbb{Z}$, with addition as the binary operation. If the condition is not satisfied by the group, it is called **non-Abelian**. Typical examples of non-Abelian groups are groups of matrices with matrix multiplication and groups of functions with function composition.

Some groups are defined by a finite set of **generators**: $\{t_1, ..., t_n\}$, such that any element of the group, $g \in G$, can be expressed as a product of generators: $g = \prod_{i \in I} t_i$ for a set of indices I. The product operator, $\prod$, then means successive use of the binary operation $*$. Such groups are called finitely generated.

Continuously generated groups are groups that contain elements arbitrarily close to the identity e, such that any element of the group may be reached by repeated action these infinitesimal elements. We may then expand an infinitesimal group element as

$$g(\alpha) = e + i\alpha_a T_a,$$

with α_a the infinitesimal group parameters and T_a the generators. The generators span the **algebra** of the group as a vector field over some set of numbers (e.g. $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}, ...$). Important examples of continuous groups are

- The Poincaré group; the symmetry group of special relativity, generated by rotations, translations and boosts of the space-time coordinates.
- The Abelian unitary group $U(1)$; the gauge symmetry group of classical and quantum electrodynamics, isomorphic to the set of complex numbers of unit length: $\{e^{ia} \mid a \in [0, 2\pi)\}$.
- The special unitary group $SU(N)$, consisting of $N \times N$ unitary[1] matrices with determinant 1. $S(2)$ is the gauge group of the electroweak theory, with the Pauli matrices as generators. $SU(3)$ is the gauge group of quantum chromodynamics, with the Gell-Mann matrices as generators.

Given a group G and a set X, the **group action** of G on X is a function $\varphi : G \times X \to X$, satisfying

- $\varphi(e, x) = x \ \forall \ x \in X$ (identity).
- $\varphi(g_1 g_2, x) = \varphi(g_1, \varphi(g_2, x)) \ \forall \ g_1, g_2 \in G, \ \forall \ x \in X$ (compatability).

From these axioms, it follows that for each $g \in G$, the function $\varphi_g = \varphi(g, \cdot) : X \to X$ is a bijection with inverse $\varphi_{g^{-1}}$.

The symmetry groups of physics are realized through group actions: The Poincaré group acts on space-time coordinates e.g. by Lorentz transformations which may be represented as matrices: $x \to \Lambda x$. The gauge groups act on the particle and gauge fields of quantum field theory, e.g. $U(1)$ acts on Dirac spinors by a simple phase multiplication $\psi \to e^{i\alpha}\psi$, and a simultaneous transformation of the electromagnetic vector potential $A_\mu \to A_\mu + \frac{i}{e}\partial_\mu \alpha$.

[1] Recall that a matrix is unitary if its complex conjugated transpose is its inverse.

Topology

We start with a definition of the mathematical field of topology itself [111]: *The area of mathematics concerned with the general properties of shapes and space, and in particular with the study of properties that are not changed by continuous distortions, such as stretching.*

Right off the bat, this sounds like mathematics fit for the QHE, as it has been stressed how robust it is to distortions, i.e. dislocations and impurities in materials. We now move on to the definitions important to this thesis.

A **topology** on a set X is a collection $\mathcal{T}$ of subsets of X with the following properties:

- $\emptyset, X \in \mathcal{T}$.
- If $U_\alpha \in \mathcal{T}$ for $\alpha = 1, 2, ...$, then $\bigcup_\alpha U_\alpha \in \mathcal{T}$.
- If $U_1, ..., U_n \in \mathcal{T}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$.

We say that the topology is closed under arbitrary unions and finite intersections. A set X with a defined topology $\mathcal{T}$ is called a **topological space**, denoted $(X, \mathcal{T})$. If the specific topology in question is implied or unimportant, X by itself is usually referred to as the topological space. The sets constituting the topology $\mathcal{T}$ are termed **open sets**.

Given two topological spaces X and Y, a **product space**, $X \times Y$, can be defined as the set of all ordered doubles: $X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}$, with the **product topology**: $\mathcal{T}_{X\times Y} = \{A \times B \mid A \in \mathcal{T}_X,\ B \in \mathcal{T}_Y\}$. For product spaces, the **projection map** of the first argument $\pi_1 : X \times Y \to X$ is defined as the surjective map $\pi_1(x, y) = x$.

A function $f : X \to Y$ between topological spaces is called **continuous** if for each open set $V \subset Y$, the set $f^{-1}(V) = \{x \in X \mid f(x) \in V\} \subset X$ is open. If f is bijective and both f and its inverse $f^{-1} : Y \to X$ are continuous, f is called a **homeomorphism**. It follows that if f is a homeomorphism, then the set $U \subset X$ is open if and only if $f(U) \subset Y$ is open. There is hence a bijective correspondence between the topologies of X and Y. Then any **topological property** of X, i.e. any property of X that may be expressed entirely in terms of the topology also holds in Y.

Specifically, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called C^p if its p'th derivative exists and is continuous. A function which is infinitely continuously differentiable is said to be C^∞. A C^∞ homeomorphism is called a **diffeomorphism**.

A **topological group** $(G, \mathcal{T}, *)$ is a topological space with a binary operation obeying the group axioms, and such that

- the binary operation $* : G \times G \to G$ is a continuous function of two variables,
- the inversion $i : G \to G$, taking $i(g) = g^{-1}\ \forall\ g \in G$ is continuous.

We are now ready to define the topological entity most important to the QHE.

Assume we have three topological spaces E, F and B, and a continuous, surjective map $p : E \to B$ which locally looks like a product. More specifically, assume that there is a collection of open sets $\{U_i\}$, that covers B: $B \subset \bigcup_i U_i$, and for each U_i a homeomorphism $\phi_i : p^{-1}(U_i) \to U_i \times F$ with the property that $p = \pi_1 \circ \phi_i$. $\circ$ denotes function composition: $(f \circ g)(x) = f(g(x))$.

The whole collection (E, B, F, p) is called a **fiber bundle**, and we say that E fibers over B with fiber F. The spaces E is called the total space and B is called the base space. For each point $x \in B$, the set $p^{-1}(x)$ is called the **fiber over** x.

A **section** of a bundle is a continuous map $s : B \to E$, which is a left inverse of the bundle projection p: $(p \circ s)(x) = x\ \forall x \in B$.

Assume next there is a topological group G acting on the fiber F. The structure is then termed a **G-bundle** if the homeomorphisms ϕ_i can be chosen such that for each pair of intersecting sets U_i, U_j, the composition

$$\phi_i \circ \phi_j^{-1} : (U_i \cap U_j) \times F \to (U_i \cap U_j) \times F$$

has the form

$$(\phi_i \circ \phi_j^{-1})(u, x) = (u, \gamma_{ij}(u)x),$$

where the **transition function** $\gamma_{ij} : U_i \cap U_j \to G$ is continuous. That is, in a sense, the homeomorphisms ϕ_i and ϕ_j differ by an element of G. Readers familiar with differential geometry will recognize the structural similarity this patchwork shares with an atlas.

Finally, in the case of the structure group being the fiber itself, $G = F$, and the group action is multiplication from the left, the bundle is called a **principal G-bundle**.

Appendix B

Elements of Differential Geometry

This appendix aims to serve as a recap of the differential geometric formalism used in this thesis. Good primary sources are [87] and [110] for very mathematical introductions to the subject, and [112] for an introduction more focused on physics. The latter, physics-centered approach is adopted here whenever possible. See also [113] for an introduction to line bundles and their relation to physics. The reader is advised to look through Appendix A before reading this appendix.

Manifolds

Firstly, a **metric** d is any distance-measuring function defined on a topological space X, i.e. $d : X \times X \to \mathbb{R}$, such that $d(x, y)$ gives a notion of the distance between points $x, y \in X$. A topological space with a metric is called a **metric space**.

A **manifold** M is a metric space with the property that for every $x \in M$, there exists an open set $U \subset M$ containing x and some integer $n \in \mathbb{N}$ such that U is homeomorphic to $\mathbb{R}^n$. That is, a manifold is a space that looks Euclidean on small scales. The sets U_α with their corresponding homeomorphims ϕ_α defines local **coordinate systems** or **charts** of M. A C^∞ **atlas** of M is a collection of charts $\{(U_\alpha, \phi_\alpha)\}$ that satisfies:

- The $\{U_\alpha\}$ covers M: $M \subset \bigcup_\alpha U_\alpha$.
- Transition between charts are smooth: For any two intersecting charts U_α and U_β the map $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$ is a diffeomorphism.

A manifold that is everywhere locally homeomorphic to $\mathbb{R}^n$ for the same n will be termed an n-manifold.

A **differentiable manifold** (also called a C^∞ - or smooth manifold) is a manifold with a **maximal atlas**, an atlas consisting of every possible compatible chart.

Standard examples of C^∞ 2-manifolds are the sphere S^2 and the torus T^2. They are usually visualized as imbedded in three dimensional space. An important fact about n-dimensional manifolds, is that they all have imbeddings in a higher-dimensional Euclidean space.

Vectors and One-Forms

Now, as a manifold may well be *curved*, as torus or sphere is, the notion of **vectors** must be properly defined. If we consider the sets of tangent vectors for two non-antipodal points on a sphere, the sets will not be equal. This is different from the situation in $\mathbb{R}^n$, where the set of possible vectors is the same at each point (and equal to $\mathbb{R}^n$ itself). Also observe that the tangent space for each point on S^2 is a copy of $\mathbb{R}^2$, with different orientations in 3-space.

For the case of general manifolds M, it is then more appropriate to assign a vector space to each point $x \in M$, consisting of all the vectors tangent to the manifold (in a higher-dimensional imbedding) at x. This space is called the **tangent space** of M at x, denoted T_xM.[1]

The tangent space is, however, independent of higher-dimensional imbeddings and should be defined that way. Following [112], if we consider a curve $\gamma : \mathbb{R} \to M$, parametrized by some real parameter λ, a tangent vector to the curve at a point x is given by the directional derivative $(d\gamma/d\lambda)_x$. The tangent space at x may then be defined as the set of tangent vectors to all the curves passing through x. Equivalently, the tangent space is the set of all directional derivatives at the point x.

Consider then, any chart (U, ϕ) with local coordinates $y^i \in \mathbb{R}^n$, and an arbitrary function $f : M \to \mathbb{R}$. Taking the directional derivative of f along the curve γ yields

$$\frac{d(f \circ \gamma)}{d\lambda} = \frac{d}{d\lambda}\left[(f \circ \phi^{-1}) \circ (\phi \circ \gamma)\right] = \frac{d(\phi \circ \gamma)^i}{d\lambda}\frac{\partial(f \circ \phi^{-1})}{\partial y^i} = \frac{dy^i}{d\lambda}\partial_i f.$$

[1] It is possible, and mathematically correct, to define the tangent space as the fibers of the tangent bundle: a special case of fiber bundle. This construction is, however, excessively complicated for the intentions of this appendix.

As the function f is arbitrary, we have

$$\frac{d}{d\lambda} = \frac{dy^i}{d\lambda}\partial_i,$$

and we may take the partial derivatives to be the basis vectors of the tangent space.

The action of coordinate transformations $y \to y'$ is then given by the chain rule:

$$\partial'_j = \frac{\partial}{\partial y'^j} = \frac{\partial y^i}{\partial y'^j}\partial_i.$$

Demanding that a vector with components A^i is independent of basis: $A^i\partial_i = A'^i\partial'_i$, yields the inverse transformation for vector components:

$$A'^i = \frac{\partial y'^i}{\partial y^j}A^j.$$

One is usually most interested in the components, rather than the entire vector, and refers to the components A^i as the vector.

Define the **cotangent space** at x, T^*_xM, as the set of all linear maps $\omega : T_xM \to \mathbb{R}$. The elements of this space are called dual vectors or **one-forms**. From their definition, one-forms act on vectors to produce scalars, so a basis, θ^i, on T^*_xM can be constructed by demanding

$$\theta^i(\partial_j) = \delta^i_j,$$

δ^i_j being the Kroenecker delta. This condition is satisfied by

$$\theta^i(\partial_j) = \frac{\partial y^i}{\partial y^j},$$

i.e. with $\theta^j = \mathrm{d}y^j$, the **gradients** (or **differentials**). A member of the cotangent space being a gradient fits together nicely with its action on a vector:

$$\mathrm{d}f\left(\frac{d}{d\lambda}\right) = \frac{df}{d\lambda},$$

the directional derivative (compare $f'(\mathbf{x};\mathbf{r}) = \nabla f \cdot \mathbf{r}$ in Euclidean space).

The transformation under coordinate change follows in the same way as for vectors:

$$\mathrm{d}y'^i = \frac{\partial y'^i}{\partial y^j}\mathrm{d}y^j,$$

and for the components:

$$\omega'_i = \frac{\partial y^j}{\partial y'^i}\omega_j.$$

Vectors are sometimes referred to as **contravariant vectors**, while one-forms are reffered to as **covariant vectors**.

Tensors

A (k,l)-**tensor** is a map $T : (T_x^*M)^k \times (T_xM)^l \to \mathbb{R}$, linear in each argument, taking k one-forms and l vectors, and returning a real number. As with vectors, one is usually most interested in the components, found in the expansion

$$T = T^{i_1 \ldots i_k}{}_{j_1 \ldots j_l} \partial_{i_1} \otimes \ldots \otimes \partial_{i_k} \otimes \mathrm{d}y^{j_1} \otimes \ldots \otimes \mathrm{d}y^{j_l},$$

$\otimes$ denoting a direct product. The transformation properties of the tensor components are found by the applying the corresponding vector and one-form transformations for each index:

$$T'^{i_1 \ldots i_k}{}_{j_1 \ldots j_l} = \frac{\partial y'^{i_1}}{\partial y^{n_1}} \ldots \frac{\partial y'^{i_k}}{\partial y^{n_k}} \frac{\partial y^{m_1}}{\partial y'^{j_1}} \ldots \frac{\partial y^{m_l}}{\partial y'^{j_l}} T^{n_1 \ldots n_k}{}_{m_1 \ldots m_l}.$$

A special tensor is the **metric tensor**, G_{ij}, which is a symmetric $(0,2)$-tensor, used to map vectors to one-forms:

$$G_{ij} A^i = A_j.$$

The inverse metric tensor G^{ij} is defined by $G^{ij} g_{ik} = \delta^i_k$, and is used to map one-forms to vectors:

$$G^{ij} \omega_i = \omega^j.$$

In full generality, the metric tensor is used to lower or raise an index of a tensor of any order and any number of indices may be lowered or raised by successive application of the metric tensor. From this, we can define the **inner product** of two vectors A and B as

$$A \cdot B = G_{ij} A^i B^j = A_j B^j = A^j B_j,$$

and vice versa for one-forms.

Another important property of the metric tensor is that it gives the **line element**:

$$ds^2 = g_{ij} \mathrm{d}y^i \mathrm{d}y^j.$$

Another special tensor is the energy-momentum tensor T_{ij}, which is found from the action S. In general, the action depends on the metric as

$$S = \int d^d x \sqrt{|G|} \mathcal{L},$$

where G is the determinant of the metric tensor. This is also true in flat Euclidean or Minkowski space, but there $G = 1$. The energy-momentum tensor is then

defined as the response of the action to a variation in the metric (see section 4.3 of [112]):

$$T_{ij} = -2\frac{1}{\sqrt{|G|}}\frac{\delta S}{\delta G^{ij}},$$

the functional derivative of the action with respect to the metric components. This definition of the energy-momentum tensor is most intuitive in the context of general relativity, where the geometry of spacetime is indeed determined by the energy density of the universe.

A result that will be useful is a rewriting of the above equation in two dimensional flat space:

$$\delta S \sim -\int d^2x T_{ij}\delta G^{ij},$$

which reproduces the first equation up to a prefactor.

Differential Forms

A special class of tensors are the differential forms (sometimes referred to as just forms). A **differential p-form** is a $(0, p)$ tensor that is completely antisymmetric. One special case is the one-form already introduced. Given two one-forms, ω and η, we may define a two-form by the **wedge product**:

$$(\omega \wedge \eta)_{ij} = \omega_i\eta_j - \omega_j\eta_i.$$

The **exterior derivative** is, in a sense, a generalization of the gradient and the curl, in that it differentiates a p-form to obtain an antisymmetrized $(p+1)$-form. For a one-form:

$$(\mathrm{d}\omega)_{ij} = \partial_i\omega_j - \partial_j\omega_i.$$

Covariant Derivatives and Connections

In general, the partial derivatives of tensors of higher order than (0,0) are not tensors themselves, which can be seen from their transformation abilities:

$$\partial'_i V'^j = \frac{\partial y^n}{\partial y'^i}\partial_n\left(\frac{\partial y'^j}{\partial y^m}V^m\right) = \frac{\partial y^n}{\partial y'^i}\frac{\partial y'^j}{\partial y^m}\partial_n V^m + \frac{\partial y^n}{\partial y'^i}\frac{\partial^2 y'^j}{\partial y^n \partial y^m}V^m,$$

where an extra term appears. This extra term means that $\partial'_i V'^j$ is no longer in the tangent plane T_xM. In order to define derivatives independently of the choice of coordinates, the **covariant derivative** may be defined as:

$$D_i V^j = \partial_i V^j + \Gamma^j_{in} V^n,$$

where the **connection** Γ^j_{in} is constructed such that its transformational properties exactly cancels the extra term from the partial derivative.

The existence of a connection means that the space is *curved*, and the **curvature tensor**, $R^k{}_{lij}$, may be defined from the equation:

$$[D_i, D_j]V^k = R^k{}_{lij}V^l - 2\Gamma^l_{ij}D_lV^k.$$

Connections may be found from the metric tensor as

$$\Gamma^i_{jk} = \frac{1}{2}G^{il}(\partial_j G_{kl} + \partial_k G_{jl} - \partial_l G_{jk}).$$

In a curved space, the notion of straight lines will differ from that ordinary Euclidean space. In a general geometry, a "straight line" is called a **geodesic**. A curve, γ, parametrized by some parameter t, is a geodesic if the **geodesic equation**

$$\frac{d^2\gamma^k}{dt^2} + \Gamma^k_{ij}\frac{d\gamma^i}{dt}\frac{d\gamma^j}{dt} = 0,$$

is satisfied. Note that in the case of $\Gamma^k_{ij} = 0 \ \forall \ i, j, k$, the geodesic equation reproduces the straight lines for Euclidean space. See [112] for a derivation of the geodesic equation.

Gauge Theory

This section follows [113] and [114]. In abstract terms, a gauge theory with gauge group G can be thought of as a principal G-bundle with space-time as the base space, and the gauge fields are the local sections of the bundle. The introduction of particle fields is formally made by an associated bundle with structure group G. The details of that operation will not be revised here, but see section 3.3.2 of [115] for the formalities in the case of Yang-Mills theory.

In this case, the connections are the one-form gauge fields themselves, and the covariant derivative takes the well known component form[2]

$$D_\mu = \partial_\mu + A^j_\mu t^j,$$

with t^j the generators of the gauge group, and A^j the corresponding gauge field. This easiest way to see this is to forget about the bundle formalism, start out

[2]In gauge theories with coupling constant g, the covariant derivative is usually defined $D_\mu = \partial_\mu - igA^j_\mu t^j$. Here, we absorb these prefactors into the A^j's, as the distinction will not be important to us.

with the matter fields, and demand their derivatives to be invariant under the action of G, see e.g. sections 15.1 and 15.2 of [4].

Now, a **gauge transformation** amounts to relating the section of a chart (U_i, ϕ_i) to that of an intersecting chart (U_j, ϕ_j). It can be shown [114] that the effect on the connection is:

$$A' = \gamma_{ij}^{-1} A \gamma_{ij} + \gamma_{ij}^{-1} \mathrm{d}\gamma_{ij},$$

with $\gamma_{ij} : U_i \cap U_j \to G$ the transition functions defined in Appendix A.

The connections specify a curvature on principal bundle, which may be expressed as (compare with $R^k{}_{ijl}$ of the last section)

$$F = \mathrm{d}A + A \wedge A.$$

For the gauge group $U(1)$, the fields A commute[3], and the expression for F simplifies to

$$\begin{aligned} F_{\mu\nu} &= (\mathrm{d}A)_{\mu\nu} \\ &= \partial_\mu A_\nu - \partial_\nu A_\mu, \end{aligned}$$

the electromagnetic field tensor. Also, the gauge transformation simplifies; with $\gamma_{ij} = e^{i\alpha(x)}$ for some well-behaved function α:

$$A'_\mu = A_\mu + i\partial_\mu \alpha,$$

and it is easy to see that the curvature is invariant with respect to gauge transformations. Also, see that the curvature may be expressed as

$$F = \partial_\mu A_\nu \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu,$$

in this case, as $\omega \wedge \eta = -\eta \wedge \omega$ for any two one forms.

[3]Rather, the generators of the gauge group commutes.

Appendix C

The Onsager Relations

This appendix derives the generalized Onsager relations for transport coefficients, used to determine the symmetry properties of the conductivity matrix and, by extension, the resisitivity matrix. First derived by Lars Onsager in 1931, they are still relevant today, e.g. in the context of the quantum Hall effect. The symmetry relations they put on the off-diagonal elements of the transport matrix are, in fact, crucial to the analysis in this thesis. The work here is based on the general discussion in section 15.7 of [2], put in the context of the Hall system.

In studying the Hall system, the natural definition for the equilibrium state will be when the steady Hall voltage $V_H = v_d B/L_y$ is reached so current stops flowing in the y-direction and flows normally, as if not subject to a magnetic field (but with a configured voltage), in the x-direction.

Assume the system is close to this equilibrium state, experiencing a small fluctuation. The currents and electric fields are related by $\mathbf{J} = \sigma\mathbf{E}$, or

$$J_i = \sigma_{ij} E_j, \tag{C.0.1}$$

with $i, j = x, y$. We view the electric fields as driving forces trying to bring the system to equilibrium. This would amount to the entropy reaching its maximum value, and as such, the entropy should be considered a function of the charge accumulated on the Hall bar, building up the voltage. More accurately, it should be considered a function of the charge gradient: $S = S(\nabla q)$, attaining its maximum value at some $\nabla q = \nabla \tilde{q}$. In the following, for notational simplicity, we will name $u_x = \partial q/\partial x$ and $u_y = \partial q/\partial y$. This way, the relation $J_i = \dot{u}_i$ holds, with the dot denoting a time derivative. Furthermore, we will work in natural units with the Boltzmann constant $k = 1$.

Given that the system is close to equilibrium, expand the entropy about $u = \tilde{u}$:

$$S(u) \approx S(\tilde{u}) + \frac{1}{2}\left(\frac{\partial^2 S}{\partial u_i \partial u_j}\right)_{u=\tilde{u}} (u_i - \tilde{u}_i)(u_j - \tilde{u}_j),$$

as the first derivative of S is zero at $u = \tilde{u}$. With our current view of the electric fields, they will be related to the entropy by (possibly after a rescaling)

$$E_i = \frac{\partial S}{\partial u_i} = -\gamma_{ij}(u_j - \tilde{u}_j), \tag{C.0.2}$$

where $\gamma_{ij} = -(\partial^2 S/\partial u_i \partial u_i)_{u=\tilde{u}} = \gamma_{ji}$.

Now, the probability that a given fluctuation away from equilibrium occurs is given by $p \propto \exp(\Delta S)$, where $\Delta S = S(u) - S(\tilde{u}) = -\frac{1}{2}\gamma_{ij}(u_i - \tilde{u}_i)(u_j - \tilde{u}_j)$ measures the deviation from equilibrium. Taking the variables u_i to be continuous, the normalization constant ("partition function") is then given as

$$Z = \int d^2u \exp\left(-\frac{1}{2}\gamma_{ij}(u_i - \tilde{u}_i)(u_j - \tilde{u}_j)\right).$$

The ensemble average of u_i is given by (compare eq. (2.2.4))

$$\tilde{u}_i = \langle u_i \rangle = \frac{1}{Z}\int d^2u u_i \exp\left(-\frac{1}{2}\gamma_{ij}(u_i - \tilde{u}_i)(u_j - \tilde{u}_j)\right),$$

and differentiating this quantity with respect to $\tilde{u}_j$, one obtains

$$\begin{aligned}
\delta_{ij} &= \frac{1}{Z}\int d^2u[\gamma_{nj}(u_n - \tilde{u}_n)]u_i \exp\left(-\frac{1}{2}\gamma_{ij}(u_i - \tilde{u}_i)(u_j - \tilde{u}_j)\right) \\
&= \frac{1}{Z}\int d^2u(-E_j u_i) \exp\left(-\frac{1}{2}\gamma_{ij}(u_i - \tilde{u}_i)(u_j - \tilde{u}_j)\right) \\
&= -\langle E_j u_i \rangle,
\end{aligned} \tag{C.0.3}$$

the ensemble average of the product $E_j u_i$. The relation (C.0.2) was used from line 1 to line 2.

Then comes the tricky part: time reversal invariance. Many physical quantites would remain unchanged if time were to go backwards instead of forwards. Mathematically, one studies the transformation $t \mapsto -t$, and time reversal invariance for the quantity A would mean $A(t) = A(-t)$. If, however, a quantity is proportional to a macroscopic velocity: $B(t) \propto \frac{dA}{dt}$, with A time reversal invariant, we get

$$B(t) \mapsto B(-t) \propto \frac{dA}{d(-t)} = -\frac{dA}{dt} \propto -B(t),$$

that is, $B(t) = -B(-t)$.

This is exactly what happens with a magnetic field under time reversal because it is generated be a current density according to Ampére's law: $\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$, and $\mathbf{J}$ obeys the continuity equation $\nabla \cdot \mathbf{J} = \partial\rho/\partial t$. As a charge density ($\rho$) does not change sign under time reversal, $\mathbf{J}$, and by extension $\mathbf{B}$ does.

As the variables u_i represent the same kind of quantity, their sign change under time reversal will be equal, and as such, the temporal correlation

$$\langle u_i(0)u_j(t)\rangle \overset{B\to -B}{=} \langle u_i(0)u_j(-t)\rangle,$$

is invariant, so long as we remember to flip the sign of B. As physics is not concerned with when we set the zero value of time, we also have $\langle u_i(0)u_j(-t)\rangle = \langle u_i(t)u_j(0)\rangle$, and combining these:

$$\langle u_i(0)u_j(t)\rangle \overset{B\to -B}{=} \langle u_i(t)u_j(0)\rangle.$$

Differentiating this equation and putting $t = 0$ then yields

$$\langle u_i(0)J_j(0)\rangle \overset{B\to -B}{=} \langle J_i(0)u_j(0)\rangle.$$

Invoking the relation (C.0.1):

$$\langle u_i(0)\sigma_{jn}(B)u_n(0)\rangle = \langle \sigma_{in}(-B)u_n(0)u_j(0)\rangle,$$

and finally using eq. (C.0.3) brings us to

$$\sigma_{ij}(B) = \sigma_{ji}(-B), \tag{C.0.4}$$

the Onsager symmetry relations for the conductivities. In terms of the more directly measureable resistivity matrix:

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \sigma = \rho^{-1} = \frac{1}{\det[\rho]} \begin{pmatrix} \rho_{yy} & -\rho_{yx} \\ -\rho_{xy} & \rho_{xx} \end{pmatrix},$$

it follows that

$$\rho_{ij}(B) = \rho_{ji}(-B). \tag{C.0.5}$$

In the steady state, the relation $\mathbf{E} = \rho\mathbf{J}$ leads to $E_x = \rho_{xx}J_x$, $E_y = \rho_{xy}J_x$. The operation $B \to -B$ would change the sign of E_y but not J_x and hence

$$\begin{aligned} \rho_{xy}(B) &= -\rho_{xy}(-B), \\ \rho_{xy} &= -\rho_{yx}, \end{aligned} \tag{C.0.6}$$

and similarly

$$\sigma_{xy} = -\sigma_{yx}. \tag{C.0.7}$$

Note that the diagonal components suffer no restrictions from the Onsager relations in this case.

Finally, resistivities and conductivities are independent of geometry so if the material of the Hall bar is isotropic in the absence of a magnetic field, considering a rotation of the entire system by 90° about the axis of the magnetic field leads to

$$\sigma_{xx} = \sigma_{yy}, \tag{C.0.8}$$

$$\rho_{xx} = \rho_{yy}, \tag{C.0.9}$$

giving us the final form of the transport matrices.

Appendix D

Hyperbolic Geometry

This appendix emphasizes the relation between different geometries, specifically what distinguishes Euclidean geometry from hyperbolic geometry. The discussion is done completely in geometrical terms, starting with Euclid's axioms, pointing out their fallacies, and continuing on to Hilbert's axioms, with all their possibilities. The point is to show that in studying new, abstract spaces, such as the QHE parameter space, there is no *a priori* reason assume a Euclidean geometry, rather than an hyperbolic.

Euclid's Axioms

Geometry is an axiomatic theory, meaning it is based on a fundamental set of postulates from which every result can be derived. The axiomatic method in geometry was introduced by Euclid in "Elements" around 300 BCE and proved to be extremely successful, as this work served as the main textbook for students of geometry for more than two millenia [85]. However, Euclid's definitions and axioms do not satisfy the modern mathematical criteria for rigor and, as will be shown, give rise to an incomplete theory of geometry.

Euclid's postulates [111] can be formulated as follows:

1. A straight line segment can be drawn joining any two points.
2. Any straight line segment can be extended indefinitely into a straight line.
3. Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.
4. All right angles are congruent.

5. If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines, if extended indefinitely, must intersect on that side.

The last postulate is called the *Axiom of parallels* and is most peculiar. Whereas the first four axioms all are minimalistic, the fifth seems unnecessarily complicated. Furthermore, its statement is too intuitive - with the common notion of straight lines it is impossible to imagine that the axiom is not satisfied. It is therefore natural that the fifth axiom should follow from the other four. This has led to many attempts from mathematicians to prove the fifth postulate from the first four, but none were succesful (see chapter 0.5 of [86] for a brief historical overview). The reason is that there exist geometries, different from the Euclidean, in which axioms 1-4 are satisfied but the Axiom of parallels is invalid. The next subsection provides a method of defining a logically complete theory of geometric structures with minimal assumptions about what the world should look like.

Hilbert's Axioms

This section follows closely chapter 1 of [85]. One of the most famous attempts of defining plane geometry rigorously was made by David Hilbert in his 1898 book "Grundlagen der Geometrie", where he presents a new set of axioms. Here, the two-dimensional case will be presented. We work in a plane S, which is a set of points P. Lines are defined as the subset $l \subset S$ consisting of the points satisfying the *incidence relation*, $P \in l$. With these definitions the first set of axioms, called the *incidence axioms*, can be formulated:

I1: *For every pair of distinct points A and B there is a unique line l containing A and B.*

I2: *Every line contains at least two points.*

I3: *There are at least three points that do not lie on the same line.*

Let $\overline{AB}$ denote the unique line containing A and B. Note that the third incidence axiom implies that the geometry is at least two dimensional.

Another property a geometry should include is a sense of *betweenness*, i.e. an interpretation of the expression "point B lies between points A and C". Hilbert's *axioms of order* is concerned with this task. Use the notation $A * B * C$ for the expression "B lies between A and C". The axioms of order are then:

O1: *If $A * B * C$, then A, B and C are distinct points on a line, and $C * B * A$ also holds.*

O2: *Given two distinct points A and B, there exists a point C such that $A*B*C$.*

O3: *If A, B and C are distinct points on a line, then one and only one of the relations $A*B*C$, $C*A*B$ and $B*C*A$ is satisfied.*

O4: *(Pasch's axiom) Let A, B and C be points not on the same line and let l be a line which contains none of them. If $D \in l$ and $A*D*B$, there exists an $E \in l$ such that $B*E*C$, or an $F \in l$ such that $A*F*C$, but not both.*

A more intuitive formulation of the fourth order axiom is as follows: "Any line that intersects a side of a triangle but none of its vertices also intersects one and only one of its other sides." This formulation is often referred to as Pasch's axiom. Note that Pasch's axiom restricts the geometry to a maximum of two dimensions. Also note that axioms 2 and 3 together guarantees that lines may be continued indefinitely.

The third piece of structure a geometry needs is a concept of *congruence*: a method of comparing different configurations of lines and points. Intuitively, two geometric objects are congruent if they are of the same shape and size. Hilbert's *axioms of congruence* formalize this notion.

First, order allows us to define the *line segment* from point A to point B as $AB = \{A, B\} \cup \{C | A*C*B\}$, and the *ray* from A through B as $\overrightarrow{AB} = AB \cup \{C | A*B*C\}$. Then, if A, B and C are points not on the same line, the *angle* $\angle BAC$ may be defined as the pair of rays from A through B and from A through C: $\angle BAC = \{\overrightarrow{AB}, \overrightarrow{AC}\}$.

Furthermore, two points A and B may now be defined to be on the *same side* of a line, l, if the segment AB does not intersect l, i.e. if $AB \cap l = \emptyset$. We also define the *triangle* with vertices A, B, C as as the set $ABC = AB \cup BC \cup AC$.

Let $x \cong y$ be interpreted as the statement "geometric object x is congruent to geometric object y". The axioms of congruence may then be formulated as follows:

C1: *Given a segment AB, a point C and any ray r from C, there is a uniquely determined point $D \in r$ such that $CD \cong AB$.*

C2: *$\cong$ is an equivalence relation on the set of segments.*

C3: *Given points A,B,C such that $A*B*C$ and points A', B', C' such that $A'*B'*C'$, if both $AB \cong A'B'$ and $BC \cong B'C'$, then $AC \cong A'C'$.*

C4: *Given a ray $\overrightarrow{AB}$ and an angle α, there are points E and F on opposite sides of $\overrightarrow{AB}$ such that $\angle BAE \cong \angle BAF \cong \alpha$.*

C5: *$\cong$ is an equivalence relation on the set of angles.*

C6: *Given triangles ABC and $A'B'C'$, if $AB \cong A'B'$, $AC \cong A'C'$ and $\angle BAC \cong \angle B'A'C'$, then $ABC \cong A'B'C'$, i.e. $BC \cong B'C'$, $\angle ABC \cong \angle A'B'C'$ and $\angle ACB \cong \angle A'C'B'$.*

With congruence, a circle, Γ, centered in O with radius congruent to the segment OA may be defined as $\Gamma = \{C \mid OC \cong OA\}$. Define a point B to be inside Γ if $OB < OA$ and outside Γ if $OB > OA$. There are three additional axioms required to identify the geometry with the Euclidean. The first one guarantees that circles intersect properly:

E: *Given two circles Γ and Δ such that Δ contains points both inside and outside Γ, then Γ and Δ have common points.*

The second is the *Archimedean axiom*, which says that given any two segments, one may be covered by a finite amount of copies of the other:

A: *Given a segment AB and a point $C_1 \in AB$, there exist points $C_2, C_3, \ldots, C_n$ on $\overrightarrow{AB}$ such that $AC_1 \cong C_iC_{i+1}$ for every $i < n$ and $B \in AC_n$.*

Last, but not least is the Axiom of parallels, this time in a slightly different formulation called Playfair's axiom:

P: *(Playfair's axiom) Given a line l and a point $P \notin l$ there exists at most one line m through P which does not intersect l.*

It can be shown that if all the other axioms are valid, then Playfair's axiom is equivalent to Euclid's formulation of the Axiom of parallels.

A geometry satisfying axioms I1-3, O1-4, C1-6, E, A and P can be identified with a subset of the standard Euclidean plane.

As the Axiom of parallels once again has entered and historically was the problematic axiom, in order to check its validity one may study its negation and see if one reaches a contradiction. The negations of Playfair's axiom is as follows:

N1: *Given a line l and a point $P \notin l$ there exists no lines through P that does not intersect l.*

N2: *Given a line l and a point $P \notin l$ there exists at least two lines through P that do not intersect l.*

The existence of the required line m in Playfair's axiom can be shown to follow from the other axioms, thus, N2 is the interesting case. It turns out no contradictions can be found, so postulating N2 instead of P is equally valid. An interesting consequence of N2 is that for any line l and point $P \notin l$, there will, in fact, be *infinitely* many lines through P parallell to l. A geometry satisfying Hilbert's axioms with Playfair's axiom replaced by N2 is called a *hyperbolic geometry*.

Bibliography

[1] P. W. Anderson. More Is Different. *Science*, 177(4047):393–396, 1972.

[2] R. K. Pathria and Paul D. Beale. *Statistical Mechanics*. Elsevier, 3rd edition, 2011.

[3] J. M. Yeomans. *Statistical Mechanics of Phase Transitions*. Oxford University Press, 1992.

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[5] R. Rennie, B. S. Beckett, R. A. Hands, M. Lewis, J. Clark, H. M. Clarke, D. Cooper, J. Cullerne, D. E. Edwards, and D. E. Ward. *Dictionary of Physics*. Oxford University Press, 7th edition, 2015.

[6] F. Léonard and B. Delamotte. Critical Exponent Can Be Different on the Two Sides of a Transition: A Generic Mechanism. https://arxiv.org/abs/1508.07852, Accessed 5th October 2015.

[7] A. Pelisetto and E. Vicari. Critical phenomena and renormalization-group theory. *Physics Reports*, 368(6):549 – 727, 2002.

[8] Y. Kats, L. Klein, J. W. Reiner, T. H. Geballe, M. R. Beasley, and A. Kapitulnik. Magnetic resistivity in $SrRuO_3$ and the ferromagnetic phase transition. *Phys. Rev. B*, 63:054435, 2001.

[9] D. P. Belanger and H. Yoshizawa. Neutron scattering and the critical behavior of the three-dimensional Ising antiferromagnet FeF_2. *Phys. Rev. B*, 35:4823 – 4830, 1987.

[10] M. Marinelli, F. Mercuri, and D. P. Belanger. Specific heat, thermal diffusivity, and thermal conductivity of FeF_2 at the Nel temperature. *Phys. Rev. B*, 51:8897 – 8903, 1995.

[11] D. M. Sullivan, G. W. Neilson, H. E. Fischer, and A. R. Rennie. Small angle neutron scattering from D2O in the critical region. *Journal of Physics: Condensed Matter*, 12(15):3531, 2000.

[12] P. Damay, F. Leclercq, F. Formisano, and P. Lindner. Universal critical-scattering function: An experimental approach. *Phys. Rev. B*, 58:12038 – 12043, 1998.

[13] I. M. Abdulagatov, N. G. Polikhronidi, and R. G. Batyrova. Measurements of the isochoric heat capacities Cv of carbon dioxide in the critical region. *The Journal of Chemical Thermodynamics*, 26(10):1031 – 1045, 1994.

[14] S. Kuwabara, H. Aoyama, H. Sato, and K. Watanabe. Vapor-Liquid Coexistence Curves in the Critical Region and the Critical Temperatures and Densities of Difluoromethane and Pentafluoroethane. *J. Chem. Eng. Data*, 40:112 – 116, 1995.

[15] O. Müller and J. Winkelmann. Comparison of critical properties in binary and ternary liquid mixtures using light scattering techniques. *Phys. Rev. E*, 59:2026–2038, 1999.

[16] S. J. Rzoska, K. Orzechowski, and A. Drozd-Rzoska. Critical behavior in broad-band dielectric relaxation on approaching the critical consolute point in ethanol-dodecane mixture. *Phys. Rev. E*, 65:042501, 2002.

[17] R. Kita, T. Dobashi, T. Yamamoto, M. Nakata, and K. Kamide. Coexistence curve of a polydisperse polymer solution near the critical point. *Phys. Rev. E*, 55:3159–3163, 1997.

[18] M. Kleemeier, S. Wiegand, W. Schroer, and H. Weingartner. The liquid-liquid phase transition in ionic solutions: Coexistence curves of tetra-n-butylammonium pricrate in alkyl alcohols. *The Journal of Chemical Physics*, 110(6):3085–3099, 1999.

[19] M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari. 25th-order high-temperature expansion results for three-dimensional Ising-like systems on the simple-cubic lattice. *Phys. Rev. E*, 65:066127, 2002.

[20] Christiaan Sybesma. *Biophysics*. Kluwer Academic Publishers, 1989.

[21] L. P. Kadanoff. Scaling Laws for Ising Models Near T_C. *Physics*, 2(5):263 – 272, 1996.

[22] K. G. Wilson. Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Phys. Rev. B*, 4(9):3174 – 3184, 1971.

[23] K. G. Wilson. The Renormalization Group: Critical Phenomena and the Kendo Problem. *Rev. Mod. Phys.*, 47(4):773 – 841, 1975.

[24] Jaakko Nissinen. *Renormalization, Phase Transitions, and Symmetries in the Quantum Hall Effects*. PhD thesis, University of Oslo, 2013.

[25] Jay L. Devore and Kenneth N. Berk. *Modern Mathematical Statistics with Applications*. Springer, 2nd edition, 2012.

[26] E. Brézin and J. Zinn-Justin, editors. *Champs, Cordes et Phénomenès Critiques - Fields, Strings and Critical Phenomena*. North-Holland, 1990.

[27] Nils Strodthoff. Conformal Field Theory - Foundations, Extensions and Boundaries. Master's thesis, Imperial College London, 2009.

[28] Elias M. Stein and Rami Shakarchi. *Complex Analysis*. Princeton University Press, 2003.

[29] Mary L. Boas. *Mathematical Methods in the Physical Sciences*. John Wiley and Sons Inc., 3rd edition.

[30] A. A. Belavin, A. M. Polyankov, and A. B. Zamolodchikov. Infinite Conformal Symmetry In Two-Dimensional Quantum Field Theory. *Nuclear Physics B*, 241, 1984.

[31] A. B. Zamolodchikov. 'Irreversibility' of the Flux of the Renormalization Group in a 2D Field Theory. *Pis'ma Zh. Eksp. Teor. Fiz.*, 24(12):565–567, 1986.

[32] J. L. Cardy. Is There a c-Theorem in Four Dimensions? *Phys. Rev. B*, 215(4):749–752, 1988.

[33] Y. Nakayama. Scale Invariance Vs Conformal Invariance. *Physics Reports*, 569(Supplement C):1–93, 2015.

[34] Peter E. Haagensen. Duality and the Renormalization Group. arXiv: hep-th/9708110, Accessed 28th Aug. 2017.

[35] K. von Klitzing, G. Dorda, and M. Pepper. New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance. *Phys. Rev. Lett.*, 45(6):494–497, 1980.

[36] R. B. Laughlin. Quantized hall conductivity in two dimensions. *Phys. Rev. B*, 23:5632–5633, 1981.

[37] D. C. Tsui, H. L. Störmer, and A. C. Gossard. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.*, 48:1559–1562, 1982.

[38] R. B. Laughlin. Anomalous quantum hall effect: An incompressible quantum fluid with fractionally charged excitations. *Phys. Rev. Lett.*, 50:1395–1398, 1983.

[39] The Nobel Prize in Physics 2010 - Advanced Information. http://www.nobelprize.org/nobel_prizes/physics/laureates/2010/advanced.html, Accessed 29th Aug. 2017.

[40] The Nobel Prize in Physics 2016 - Advanced Information. http://www.nobelprize.org/nobel_prizes/physics/laureates/2016/advanced.html, Accessed 29th Aug. 2017.

[41] NIST. The New Standard for Resistance Standards. https://www.nist.gov/news-events/news/2017/05/new-standard-resistance-standards, Accessed 29th Aug. 2017.

[42] Justervesenet. DC-Resistans. https://www.justervesenet.no/maleteknikk/kalibrering/elektro/dc-resistans/, Accessed 29th Aug. 2017.

[43] David Tong. The Quantum Hall Effect. http://www.damtp.cam.ac.uk/user/tong/qhe/qhe.pdf, Accessed 4th Oct. 2016.

[44] E. H. Hall. On a new action of the magnet on electric currents. *American Journal of Mathematics*, 2(3):287–292, 1879.

[45] David J. Griffiths. *Introduction to Quantum Mechanics*. Springer, 2nd edition, 2005.

[46] S.-T. Lo, Y.-T Wang, S.-D. Lin, G. Strasser, J. P. Bird, Y-.F. Chen, and C.-T. Liang. Tunable Insulator-Quantum Hall Transition in a Weakly Interacting Two-Dimensional Electron System. *Nanoscale Research Letters*, 8(1):307, 2013.

[47] P. C. Hemmer. *Kvantemekanikk*. Tapir Akademisk Forlag, 5th edition, 2005.

[48] M. Kohmoto. Topological Invariant and the Quantization of the Hall Conductance. *Annals of Physics*, 160(2):343–354, 1985.

[49] E. Brown. Bloch Electrons in a Uniform Magnetic Field. *Physical Review*, 133(4A), 1964.

[50] M. Konig, S. Wiedmann, C. Brne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang. Quantum Spin Hall Insulator State in HgTe Quantum Wells. *Science*, 318(5851):766–770, 2007.

[51] B. I. Halperin. Quantized Hall Conductance, Current-Carrying Edge States, and the Existence of Extended States in a Two-Dimensional Disordered Potential. *Phys. Rev. B*, 25(4):2185–2190, 1982.

[52] G. Ebert, K. von Klitzing, C. Probst, and K. Ploog. Magneto-Quantumtransport on GaAs-$Al_xGa_{1-x}As$ Heterostructures at Very Low Temperatures. *Solid State Communications*, 44(2):95–98, 1982.

[53] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim. Experimental Observation of the Quantum Hall Effect and Berry's Phase in Graphene. *Nature*, 483:201–204, 2005.

[54] K. S. Novoselov, Z. Jiang, Y. Zhang, S. V. Morozov, H. L. Stormer, U. Zeitler, J. C. Maan, G. S. Beobinger, P. Kim, and A. K. Geim. Room-Temperature Quantum Hall Effect in Graphene. *Science*, 315:1379, 2007.

[55] Hideo Aoki and Mildred S. Dresselhaus. *Physics of Graphene*. Springer, 2014.

[56] V .P. Gusynin and S. G. Sharapov. Unconventional Integer Quantum Hall Effect in Graphene. *Phys. Rev. Lett.*, 95:146801, 2005.

[57] H. L. Stormer. Two-Dimensional Electron Correlation in High Magnetic Fields. *Physica B*, 177:401–408, 1992.

[58] X. Du, I. Skachko, F. Duerr, A. Luican, and E. Y. Andrei. Fractional Quantum Hall Effect and Insulating Phase of Dirac Electrons in Graphene. *Nature*, 462:192–195, 2009.

[59] K. I. Bolotin, F. Ghahari, M. D. Shulman, H. L. Stormer, and P. Kim. Observation of the Fractional Quantum Hall Effect in Graphene. *Nature*, 462:196–199, 2009.

[60] B. E. Feldman, B. Krauss, J. H. Smet, and A. Yacoby. Unconventional Sequence of Fractional Quantum Hall States in Suspended Graphene. *Science*, 337:1196–1199, 2012.

[61] David J. Thouless. *Topological Quantum Numbers in Nonrelativistic Physics*. World Scientific, 1998.

[62] John Baez and Javier Muniain. *Gauge Fields, Knots and Gravity*. World Scientific Publishing, 2015.

[63] D. R. Hofstadter. Energy Levels and Wave Functions of Bloch Electrons in Rational and Irrational Magnetic Fields. *Phys. Rev. B*, 14(6):2239–2249, 1976.

[64] Edward Witten. Three Lectures On Topological Phases Of Matter. https://arxiv.org/abs/1510.07698, Accessed 3rd Sept. 2017.

[65] C. A. Lütken and C. G. Ross. Duality in the Quantum Hall System. *Phys. Rev. B*, 45(20):11837–11845, 1992.

[66] M. Amado, E. Diez, F. Rossella, V. Bellani, D. López-Romero, and D. K. Maude. Magneto-Transport of Graphene and Quantum Phase Transitions in the Quantum Hall Regime. *Journal of Physics: Condensed Matter*, 24(30):305302, 2012.

[67] H. P. Wei, D. C. Tsui, M. A. Paalanen, and A. M. M. Pruisken. Experiments on Delocalization and Universality in the Integral Quantum Hall Effect. *Phys. Rev. Lett.*, 61(11):1294–1296, 1988.

[68] X. Wang, H. Liu, J. Zhu, P. Shan, P. Wang, H. Fu, L. Du, L. N. Pfeiffer, K. W. West, X. C. Xie, R.-R. Du, and X. Lin. Scaling Properties of the Plateau Transitions in the Two-Dimensional Hole Gas System. *Phys. Rev. B*, 93:075307, 2016.

[69] Wanli Li, G. A. Csathy, D. C. Tsui, L. N. Pfeiffer, and K. W. West. Scaling and Universality of Integer Quantum Hall Plateau-to-Plateau Transitions. *Phys. Rev. Lett.*, 94:206807, 2005.

[70] Wanli Li, C. L. Vicente, J. S. Xia, W. Pan, D. C. Tsui, L. N. Pfeiffer, and K. W. West. Scaling in Plateau-to-Plateau Transition: A Direct Connection of Quantum Hall Systems with the Anderson Localization Model. *Phys. Rev. Lett.*, 102:216801, 2009.

[71] T. Khouri, M. Bendias, P. Leubner, C. Brune, H. Buhmann, L. W. Molenkamp, U. Zeitler, N. E. Hussey, and S. Wiedmann. High-Temperature Quantum Hall Effect in Finite Gapped HgTe Quantum Wells. *Phys. Rev. B*, 93:125308, 2016.

[72] A. J. M. Giesbers, U. Zeitler, L. A. Ponomarenko, R. Yang, K. S. Novoselov, A. K. Geim, and J. C. Maan. Scaling of the Quantum Hall Plateau-Plateau Transition in Graphene. *Phys. Rev. B*, 80:241411, 2009.

[73] L. Engel, H. P. Wei, D. C. Tsui, and M. Shayegan. Critical Exponent in the Fractional Quantum Hall Effect. *Surface Science*, 229(1):13 – 15, 1990.

[74] T. Machida, S. Ishizuka, S. Komiyama, K. Muraki, and Y. Hirayama. Scaling in Fractional Quantum Hall Transitions. *Physica B*, 298(1):182 – 186, 2001.

[75] N. A. Dodoo-Amoo, K. Saeed, D. Mistry, S. P. Khanna, L. Li, E. H. Linfield, A. G. Davies, and J. E. Cunningham. Non-Universality of Scaling Exponents in Quantum Hall Transitions. *Journal of Physics: Condensed Matter*, 26(47):475801, 2014.

[76] Y. G. Arapov, S. V. Gudina, A. S. Klepikova, V. N. Neverov, S. G. Novokshonov, G. I. Kharus, N. G. Shelushinina, and M. V. Yakunin. Scaling in the Quantum Hall Effect Regime in n-InGaAs/GaAs Nanostructures. 117(1):144–152, 2013.

[77] J. L. Cardy. Duality and the θ Parameter in Abelian Lattice Models. *Nucl. Phys. B*, 205(1):17–26, 1982.

[78] A. Shapere and F. Wilczek. Self-Dual Models with Theta Terms. *Nucl. Phys. B*, 320(3):669–695, 1989.

[79] C. A. Lütken. Geometry of Renormalization Group Flows Constrained by Discrete Global Symmetries. *Nucl. Phys. B*, 396:670–692, 1993.

[80] C. A. Lütken. Global Phase Diagrams for Charge Transport in Two Dimensions. *J. Phys. A*, 26:L811–L817, 1993.

[81] C. A. Lütken and C. G. Ross. Delocalization, Duality and Scaling in the Quantum Hall System. *Phys. Rev. B*, 48:2500–2514, 1993.

[82] C. P. Burgess and C. A. Lutken. One-Dimensional Flows in the Quantum Hall System. *Nucl. Phys. B*, 500:367–378, 1997.

[83] B. P. Dolan. Modular Invariance, Universality and Crossover in the Quantum Hall Effect. *Nucl. Phys B*, 554:487–513, 1999.

[84] J. Nissinen and C. A. Lutken. Renormalization-Group Potential for Quantum Hall Effects. *Phys. Rev. B*, 85:155123, 2012.

[85] Bjørn Jahren. Geometric structures in dimension two. http://www.uio.no/studier/emner/matnat/math/MAT4510/h13/gs.pdf, Accessed Aug. 2014.

[86] Caroline Series. Hyperbolic Geometry. http://homepages.warwick.ac.uk/~masbb/Papers/MA448.pdf, Accessed 26th April 2017.

[87] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 1. Publish or Perish Inc., 3rd edition, 2005.

[88] Alvaro Lozano-Robledo. *Elliptic Curves, Modular Forms, and Their L-Functions*. American Mathematical Society, 2011.

[89] Fred Diamond and Jerry Shurman. *A First Course in Modular Forms*. Springer, 2016.

[90] B. Schoeneberg. *Elliptic Modular Functions*. Springer-Verlag, 1974.

[91] Robert A. Rankin. *Modular Forms and Functions*. Cambridge University Press, 1977.

[92] John B. Fraleigh. *A First Course in Abstract Algebra*. Pearson Education Limited, 7th edition, 2014.

[93] C. P. Burgess, R. Dib, and B. P. Dolan. Derivation of the Semicircle Law from the Law of Corresponding States. *Phys. Rev. B*, 62(23):15359–15362, 2000.

[94] J. Nissinen and C. A. Lütken. The Quantum Hall Curve. https://arxiv.org/abs/1207.4693, 2012.

[95] S. W. Hwang, H. P. Wei, L. W. Engel, D. C. Tsui, and A. M. M. Pruisken. Scaling in Spin-Degenerate Landau Levels in the Integer Quantum Hall Effect. *Phys. Rev. B*, 48(15):11416–11419, 1993.

[96] R. B. Laughlin, M. L. Cohen, J. M. Kosterlitz, H. Levine, S. B. Libby, and A. M. M. Pruisken. Scaling of Conductivities in the Fractional Quantum Hall Effect. *Phys. Rev. B*, 32(2):1311–1314, 1985.

[97] Y. T. Wang, G.-H. Kim, C. F. Huang, S.-T. Lo, W.-J. Chen, J. T. Nicholls, L.-H. Lin, D. A. Ritchie, Y. H. Chang, C.-T. Liang, and B. P. Dolan. Probing Temperature-Driven Flow Lines in a Gate Two-Dimensional Electron Gas with Tunable Spin-Splitting. *J. Phys.: Condens. Matter*, 24:405801, 2012.

[98] R. Willett, J. P. Eisenstein, H. L. Störmer, D. C. Tsui, A. C. Gossard, and J. H. English. Observation of an Even-Denominator Quantum Number in the Fractional Quantum Hall Effect. *Phys. Rev. Lett.*, 59(15), 1987.

[99] W. Pan, J.-S. Xia, V. Shvarts, D. E. Adams, H. L. Störmer, D. C. Tsui, L. N. Pfeiffer, K. W. Baldwin, and K. W. West. Exact Quantization of the Even-Denominator Fractional Quantum Hall State ν=5/2 Landau Level Filling Factor. *Phys. Rev. Lett.*

[100] Y. W. Suen, L. W. Engel, M. B. Santos, M. Shayegan, and D. C. Tsui. Observation of a $\nu = 1/2$ Fractional Quantum Hall State in a Double-Layer Electron System. *Phys. Rev. Lett.*, 68(9), 1992.

[101] J. Shabani, Y. Liu, M. Shayegan, L. N. Pfeiffer, K. W. West, and K. W. Baldwin. Phase Diagrams for the $\nu = 1/2$ Fractional Quantum Hall Effect in Electron Systems Confined to Symmetric, Wide GaAs Quantum Wells. *Phys. Rev. B*, 88:245413, 2013.

[102] Y. Liu, M. A. Mueed, Md. S. Hossain, S. Hasdermir, L. N. Pfeiffer, K. W. West, K. W. Baldwin, and M. Shayegan. Morphing of Two-Dimensional Hole Systems at $\nu = 3/2$ in Parallel Magnetic Fields: Compressible, Stripe and Fractional Quantum Hall Phases. *Phys. Rev. B*, 94:155312, 2016.

[103] Y. Liu, D. Kamburov, M. Shayegan, L. N. Pfeiffer, K. W. West, and K. W. Baldwin. Evolution of the 7/2 Fractional Quantum Hall State in Two-Subband Systems. *Phys. Rev. Lett.*, 107:266802, 2011.

[104] J. A. Hutasoit, A. C. Balram, S. Mukherjee, Y.-H. Wu, S. S. Mandal, and A. Wojs. The Enigma of the $\nu = 2 + 3/8$ Fractional Quantum Hall Effect. *Phys. Rev. B*, 95:125302, 2017.

[105] D. Jose and A. Datta. Structures and Chemical Properties of Silicene: Unlike Graphene. *Acc. Chem. Res.*, 47(2):593–602, 2014.

[106] M. E. Dávila and G. L. Lay. Few Layer Epitaxial Germanene: A Novel Two-Dimensional Dirac Material. *Nature: Scientific Reports*, 6.

[107] Z. Ni, E. Minamitani, Y. Ando, and S. Watanabe. Germanene and Stanene on Two-Dimensional Substrates: Dirac Cone and Z_2 Invariant. *Phys. Rev. B*, 96:075427, 2017.

[108] C. F. Huang, Y. H. Chang, H. H. Cheng, Z. P. Yang, H. D. Yeh, C. H. Hsu, C.-T. Liang, D. R. Hang, and H. H. Lin. An Experimental Study on $\Gamma(2)$ Modular Symmetry in the Quantum Hall System with a Small Spin Splitting. *J. Phys.: Condens. Matter*, 19(2), 2006.

[109] James R. Munkres. *Topology*. Prentice Hall, 2nd edition, 2000.

[110] William P. Thurston. *Three-Dimensional Geometry and Topology*, volume 1. Princeton University Press, 1997.

[111] C. Clapham, J. Nicholson, C. Chatfield, R. Cheal, J. B. Gavin, J. R. Pulham, and D. P. Thomas. *Concise Dictionary of Mathematics*. Oxford University Press, 5th edition, 2014.

[112] Sean Carroll. *Spacetime and Geometry*. Pearson Education Limited, 2014.

[113] Michael Murray. Line Bundles. http://www.maths.adelaide.edu.au/michael.murray/line_bundles.pdf, Accessed 14th Sept. 2017.

[114] George Svetlichny. Preparation for Gauge Theory. https://arxiv.org/abs/math-ph/9902027, Accessed 15th Sept. 2017.

[115] Kristian Stølevik Olsen. Modular Forms and Universality Classes of Topological Matter. Master's thesis, University of Oslo, 2016.