

Field Theoretic Aspects of Condensed Matter Physics: An Overview[★]

Eduardo Fradkin

*Department of Physics and Institute for Condensed Matter Theory,
University of Illinois at Urbana-Champaign, 1110 West Green St, Urbana Illinois 61801-3080, U.S.A.*

Abstract

In this chapter I discuss the impact of concepts of Quantum Field Theory in modern Condensed Physics. Although the interplay between these two areas is certainly not new, the impact and mutual cross-fertilization has certainly grown enormously with time, and quantum Field Theory has become a central conceptual tool in Condensed Matter Physics. In this chapter I cover how these ideas and tools have influenced our understanding of phase transitions, both classical and quantum, as well as topological phases of matter, and dualities.

Keywords:

Contents

1	Introduction	3
2	Early Years: Feynman Diagrams and Correlation Functions	3
3	Critical Phenomena	4
3.1	Classical Critical Phenomena	4
3.2	Landau-Ginzburg Theory	4
3.3	The Renormalization Group	5
3.3.1	Scaling	5
3.3.2	The Operator Product Expansion	6
3.3.3	Fixed Points	6
3.3.4	Universality	7
3.3.5	RG flows	8
3.3.6	Asymptotic Freedom	9
4	Quantum Criticality	11
4.1	Dynamic Scaling	12
4.2	The Ising Model in a Transverse Field	12
4.3	Quantum Antiferromagnets and Non-Linear Sigma Models	14
4.3.1	Spin coherent states	14
4.3.2	Path integral for a spin- S degree of freedom	15
4.3.3	Quantum Ferromagnet	16
4.3.4	Quantum Antiferromagnet	16
5	Topological Excitations	17
5.1	Topological Excitations: Vortices and Magnetic Monopoles	18
5.1.1	Vortices in two dimensions	18
5.1.2	Magnetic monopoles in compact electrodynamics	22
5.2	Non-Linear Sigma Models and Antiferromagnetic Quantum Spin Chains	25
5.3	Topology and open integer-spin chains	26

[★]Work was supported in part by the US National Science Foundation through grant No. DMR 1725401 at the University of Illinois.

6 Duality in Ising Models	27
6.1 Duality in the 2D Ising Model	27
6.2 The 3D duality: \mathbb{Z}_2 gauge theory	28
7 Bosonization	31
7.1 Dirac fermions in one space dimensions	31
7.2 Chiral symmetry and chiral symmetry breaking	34
7.3 The chiral anomaly	36
7.4 Bosonization, anomalies and duality	38
8 Fractional Charge	42
8.1 Solitons in one dimensions	42
8.2 Polyacetylene	43
8.3 Fractionally charged solitons	43
9 Fractional Statistics	45
9.1 Basics of fractional statistics	46
9.2 What is a topological field theory	46
9.3 Chern-Simons Gauge Theory	47
9.4 BF gauge theory	48
9.5 Quantization of Abelian Chern-Simons Gauge Theory	48
9.6 Vacuum degeneracy a torus	50
9.7 Fractional Statistics and Braids	51
10 Topological Phases of Matter	54
10.1 Topological Insulators	54
10.1.1 Dirac fermions in 2+1 dimensions	54
10.1.2 Dirac Fermions and Topological Insulators	55
10.1.3 Chern invariants	56
10.1.4 The quantum Hall effect on a lattice	58
10.1.5 The Anomalous Quantum Hall Effect	61
10.1.6 The Parity Anomaly	62
10.2 Three-dimensional \mathbb{Z}_2 topological insulators	67
10.2.1 \mathbb{Z}_2 Topological Invariants	67
10.2.2 The Axial Anomaly and the Effective Action	68
10.2.3 Theta terms, and Domain walls: Anomaly and the Callan-Harvey Effect .	71
10.3 Chern-Simons Gauge Theory and The Fractional Quantum Hall Effect	73
10.3.1 Landau levels and the Integer Hall effect	73
10.3.2 The Laughlin Wave Function	77
10.3.3 Quasiholes have fractional charge	78
10.3.4 The Jain States	79
10.3.5 Quasiholes have fractional statistics	80
10.3.6 Hydrodynamic Effective Field Theory	80
10.3.7 Composite Boson Field Theory	82
10.3.8 Composite Fermion Field Theory	85
10.3.9 The Compressible States	90
10.3.10 Fractional Quantum Hall Wave Functions and Conformal Field Theory .	91
10.3.11 Edge States and Chiral Conformal Field Theory	99
11 Particle-Vortex Dualities in 2+1 dimensions	105
11.1 Electromagnetic Duality	105
11.2 Particle-Vortex Duality in 2+1 dimensions	105
11.2.1 The 3D XY Model	105
11.2.2 Scalar QED in 3D	107
11.2.3 The duality mapping	109
11.3 Bosonization in 2+1 dimensions	109
11.3.1 Bosonization of the Dirac theory in 2+1 dimensions	110
11.3.2 Bosonization of the Fermi Surface	114
12 Conclusions	116

1. Introduction

Many (if not most) puzzling problems in Condensed Matter Physics involve systems with a macroscopically large number of degrees of freedom often in regimes of large fluctuations, thermal and/or quantum mechanical. The description of the physics of systems of this type requires the framework provided by Quantum Field Theory. Although quantum field theory has its origins in high-energy physics, notably in the development of Quantum Electrodynamics, it has found a nurturing home in Condensed Matter Physics.

There is a long history of cross-fertilization between both fields. Since the 1950's in many of the most significant developments in Condensed Matter Physics, Quantum Field Theory has played a key role if not in the original development but certainly in the eventual understanding the meaning and the further development of the discoveries. As a result, many of the discoveries and concepts developed in Condensed Matter have had a reciprocal impact in Quantum Field theory. One can already see this interplay in the development of the Bardeen-Cooper-Schrieffer theory of superconductivity [1] and its implications in the theory of dynamical symmetry breaking in particle physics by Nambu [2].

The close and vibrant relationship between both fields has continued to these days, and it is even stronger today than before. Many textbooks have been devoted to teaching these ideas and concepts to new generations of condensed matter physicists (and field theorists as well). The earlier texts focused on Green functions which are computed in perturbation theory using Feynman diagrams [3] [4] [5], while the more modern ones have a broader scope, use path integrals and attack non-perturbative problems [6, 7] [8] [9, 10] [11]. In two recent books I have discussed many aspects of the interrelation between condensed matter physics and quantum field theory in more depth than I can do in this chapter [9, 10].

2. Early Years: Feynman Diagrams and Correlation Functions

Quantum field theory played a key role in the development of the Theory of the Fermi Liquid [12, 13]. The theory of the Fermi liquid was first formulated by Landau using the framework of hydrodynamics and the quantum Boltzmann equation. Landau's ideas were later given a microscopic basis using Green functions and Feynman diagrams [3], including the effects of quantum fluctuations at finite temperature and non-equilibrium behavior [14, 15]. Linear response theory was developed which allowed the computation of response functions (such as electrical conductivities and magnetic susceptibilities) from the computation of correlation functions for a given microscopic theory [16]. In turn, correlation functions can be computed in terms of a set of Feynman diagrams. These concepts and tools borrowed many concepts from field theory including the study of the analytic structure of the generalized susceptibilities and the associated spectral functions (together with the use of dispersion relations). These developments led to the derivation of the fluctuation-dissipation theorem. These ideas were widely applied to metals [13] and superconductors [1], as well as to quantum magnets [17].

The spectrum of an interacting system has low energy excitations characterized by a set of quantum numbers associated with the symmetries of the theory. These low energy excitations are known as *quasiparticles*. In the case of the Landau theory of the Fermi liquid the quasiparticle is a “dressed” electron: it is a low energy excitation with the same quantum numbers (charge and spin) and an electron but with a renormalized effective mass. There are many such quasiparticles in condensed matter physics. The correlation functions (the propagators) of a physical system has a specific analytic structure. In momentum (and frequency) space, the quasiparticle spectrum is given by the poles of the correlators.

The role of symmetries and, in particular of gauge invariance, in the structure of correlation functions was investigated extensively. A direct consequence of symmetries is the existence of Ward identities which must be satisfied by all the correlation functions of the theory. Ward identities are exact relations that relate different correlation functions. Such identities contain a host of important results. For example, in a theory with a globally conserved charge, the Hamiltonian (and the action) have a global U(1) symmetry associated with the transformation of the local field operator $\phi(x)$ (which can be fermionic or bosonic) to a new field $\phi'(x) = e^{i\theta}\phi(x)$ (where θ is a constant phase). Theories with a global continuous symmetry have a locally conserved current (and satisfy a continuity equation). The Ward identity requires the correlators of these currents (and densities) to be transverse (i.e. they should have vanishing divergence). In the absence of so-called quantum anomalies (which we will discuss below) global symmetries can be made local and become gauge symmetries.

In many circumstances a global symmetry can be spontaneously broken. If the global symmetry is continuous, then the Ward identities imply the existence of gapless excitations known as Goldstone bosons. For example in the case of a superfluid, which has a spontaneously broken U(1) symmetry the Goldstone boson is the gapless phase mode. Instead, a the Néel phase of a quantum antiferromagnet has two gapless Goldstone bosons, the magnons of the spontaneously broken SO(3) global symmetry of this state of matter. Another example is the Ward identity of quantum electrodynamics, which relates the electron self-energy to the electron-photon vertex function, which also holds in non-relativistic electron fluids. In addition, these identities implied the existence of sum rules that the spectral functions must satisfy. All of these results became part of the standard toolkit of condensed matter experimentalists in analyzing their data and for theorists to make predictions. Ward identities and sum rules also imply restrictions on the allowed approximations which are often needed to obtain predictions from a microscopic model.

3. Critical Phenomena

3.1. Classical Critical Phenomena

The late 1960s and particularly 1970s brought about an intense back and forth between condensed matter physics and field theory in the context of the problem of classical critical phenomena and phase transitions. This was going to become a profound revolution on the description of macroscopic physical systems with large-scale fluctuations. The problem of continuous (“second order”) phase transitions has a long history going back to the work of Landau [18, 19] who introduced the concept of an *order parameter field*. This turned out to be a powerful concept of broad applicability in many physical systems sometimes quite different from each other at the microscopic level.

A simple example is that of a ferromagnet with uniaxial anisotropy in which the spins of the atoms in a crystal are strongly favored to be aligned (or anti-aligned) along certain directions of the crystal. The simplest microscopic model for this problem is the Ising model, a spin system in which the individual spins are allowed to take only two values, $\sigma = \pm 1$. The partition function of the Ising model (in any dimension) is

$$Z = \sum_{[\sigma]} \exp\left(-\frac{J}{T} \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \sigma(\mathbf{r})\sigma(\mathbf{r}')\right) \quad (1)$$

where J is the exchange coupling constant and T is the temperature (measured in energy units); here $[\sigma]$ denotes the sum over the 2^N spin configurations (for a lattice with N sites), and $\langle \mathbf{r}, \mathbf{r}' \rangle$ are nearest neighbor sites of the lattice. The order parameter of the Ising model is the local magnetization $\langle \sigma(\mathbf{r}) \rangle$ which, in the case of a ferromagnet, is uniform. The partition function of the Ising model can be computed trivially in one dimension. The solution of the two-dimensional Ising model by Onsager constituted a tour-de-force in theoretical physics [20]. Its actual meaning remained obscure for some time. The work of Schultz, Mattis and Lieb [21] evinced a deep connection between Onsager’s solution and the problem if the spectrum of one-dimensional quantum spin chains [22] (specifically the one-dimensional Ising model in a transverse field [23]). One important result was that the Ising model was in fact a theory of (free) fermions which, crudely speaking, represented the configurations of domain walls of the magnet. However, even this simple model cannot be solved exactly in general dimension, and approximate mean field theories of various sorts were devised over time to understand its physics.

3.2. Landau-Ginzburg Theory

Landau’s approach assumed that close enough to a phase transition, the important spin configurations are those for which the local magnetization varies slowly on lattice scales. In this picture the local magnetization, on long enough length scales, becomes an order parameter field that takes values on the real numbers, and can be positive or negative. Thus, the order parameter field is effectively the average of local magnetizations on some scale large compared to the lattice scale, which we will denote by a real field $\phi(x)$. The thermodynamics properties of a system of this type in d dimensions can be described in terms of a free energy

$$F[\phi] = \int d^d x \left[\frac{\kappa}{2} (\nabla \phi(x))^2 + a(T - T_c)\phi^2(x) + u\phi^4(x) + \dots \right] \quad (2)$$

which is known as the Ginzburg-Landau free energy. Here κ is the stiffness of the order parameter field, T_c is the (mean-field) critical temperature; a and u are two (positive) constants. This expression make sense if the transition is continuous and hence that the order parameter is small near the transition. The energy of the Ising model is invariant under the global symmetry $[\sigma] \mapsto [-\sigma]$. This is the symmetry of the group \mathbb{Z}_2 . Likewise, the Ginzburg-Landau free energy has the global (discrete) symmetry $[\phi(x)] \mapsto [-\phi(x)]$, and also has a \mathbb{Z}_2 global symmetry.

In Landau's approach, which was a mean field theory, the equilibrium state is the global minimum of this free energy. The nature of the equilibrium state depends on whether $T > T_c$ or $T < T_c$: for $T > T_c$ the global minimum is the trivial configuration, $\bar{\phi}(x) = 0$ (this is the paramagnetic state), whereas for $T < T_c$ the equilibrium state is two fold degenerate, $\bar{\phi}(x) = \pm(a(T_c - T)/2)^\beta$, with the two degenerate states being related by the \mathbb{Z}_2 symmetry (this is the ferromagnetic state). In the Landau theory the critical exponent of the magnetization is $\beta = 1/2$ and the critical exponent of the correlation length is $\nu = 1/2$. However, in the case of the 2D Ising model the order parameter exponent is $\beta = 1/8$ [24] and the correlation length exponent is $\nu = 1$. These (and other) apparent discrepancies led many theorists for much of the 1960s believe that each model was different and that these behaviors reflected microscopic differences. In addition, Landau's theory was regarded as phenomenological and believed to be of questionable validity.

3.3. The Renormalization Group

This situation was to change with the development of the Renormalization Group, due primarily to the work of Leo P. Kadanoff [25, 26, 27] and Kenneth G. Wilson [28, 29, 30, 31, 32, 33]. The renormalization Group was going to have (and still has) a profound effect both in Condensed Matter Physics and in Quantum Field Theory (and beyond).

3.3.1. Scaling

Several phenomenological theories were proposed in the 1960s to describe the singular behavior of physical observables near a continuous phase transition [34, 35, 36]. These early works argued that in order to explain the singular behavior of the observables the free energy density had to have a singular part which should be a homogeneous function of the temperature, magnetic field, etc. A function $f(x)$ is homogeneous if it satisfies the property that it transforms irreducibly under dilations, i.e. $f(\lambda x) = \lambda^k f(x)$, where λ is a real positive number (a scale) and k is called the degree. These heuristic ideas then implied that the critical exponents should obey several identities. In 1966 Kadanoff wrote and insightful paper in which he showed that the homogeneity hypothesis implied that in that regime these systems should obey scaling. He showed that this can be justified by performing a sequence of block-spin transformations in which the configurations that vary rapidly at the lattice scale a become averaged at the scale of a larger sized block of length scale $ba > a$ which resulted in a renormalization of the coupling constants from $\{K\}$ at scale a to $\{K'\}$ at the new scale ba [25, 27]. In other words, the block spin transformation amounts to a scale transformation and a renormalization of the couplings (and operators). From this condensed matter/statistical physics perspective the important physics is in the long distance (“infrared”) behavior.

A significant consequence of these ideas was that close enough to a critical point, if the distance $|x - y|$ between two local observables $O(x)$ and $O(y)$ is large compared to the lattice spacing a but small compared to the correlation length ξ , their correlation function takes the form of a power law

$$\langle O(x)O(y) \rangle \sim \frac{\text{const.}}{|x - y|^{2\Delta_O}} \quad (3)$$

where Δ_O is a positive real number known as the scaling dimension of the operator O [34, 25, 37]. These conjectures were known to be satisfied in the non-trivial case of the 2D Ising model [26], as well as in the Landau-Ginzburg theory once the effects of Gaussian fluctuations were included [37]. For example, in the 2D Ising model, the scaling dimensions of the local magnetization σ is $\Delta_\sigma = 1/8$ and of the energy density ε is $\Delta_\varepsilon = 1$, which were sufficient to explain all the singular behaviors known at that time.

The concept of renormalization actually originated earlier in quantum field theory as part of the development of Quantum Electrodynamics (QED). In QED the notion of renormalization was used to “hide” the short distance (“ultraviolet”) divergencies of the Feynman diagrams needed to compute physical processes involving electrons (and positrons) and photons, i.e. their strong, divergent, dependence of an artificially introduced short-distance cutoff or regulator. In particular the sum of the leading diagrams that enter in the electron-photon vertex amounted to

a redefinition (renormalization) of the coupling constant. It was observed by Murray Gell-Mann and Francis Low that this renormalization was equivalent to the solution of a first order differential equation that governed the infinitesimal change of the coupling, the fine structure constant $\alpha = e^2/4\pi$, under an infinitesimal change of the UV cutoff Λ [38]

$$\Lambda \frac{d\alpha}{d\Lambda} \equiv \beta(\alpha) = \frac{2}{3\pi} \alpha^2 + O(\alpha^3) \quad (4)$$

where $\beta(\alpha)$ is the Gell-Mann-Low beta function. Except for the work by Nikolai Bogoliubov and coworkers [39], this reinterpretation by Gell-Mann and Low was not actively pursued, partly because it predicted that the renormalized coupling became very large at short distances, $\alpha \rightarrow \infty$, and, conversely, it vanished in the deep long distance regime, $\alpha \rightarrow 0$ (if the electron bare mass is zero). In other terms, QED is strongly coupled in the UV and trivial in the IR. The same behavior was found in the case of the theory of a scalar field $\phi(x)$ with an ϕ^4 interaction which is relevant in the theory of phase transitions. In addition to these puzzles, the 1960s saw the experimental development of the physics of hadrons which involve strong interactions. For these reasons, for much of that decade most high-energy theorists had largely abandoned the use of quantum field theory, and explored other, phenomenologically motivated, approaches (which led to an early version of string theory.) At any rate the notion that the physics may depend on the scale was present as was the notion that in some regimes field theories may exhibit scale-invariance at least in an approximate form.

3.3.2. The Operator Product Expansion

The next stage of the development of these ideas was the concept of the operator product expansion (OPE). If we denote by $\{O_j(x)\}$ the set of all possible local operators in a theory (a field theory or a statistical mechanical system near criticality), then the product of two observables on this list closer to each other than to any other observable (and to the correlation length ξ) obeys the expansion

$$\lim_{x \rightarrow y} O_j(x) O_k(y) = \lim_{x \rightarrow y} \sum_l \frac{C_{jkl}}{|x - y|^{\Delta_j + \Delta_k - \Delta_l}} O_l\left(\frac{x + y}{2}\right) \quad (5)$$

where this equation should be understood as a weak identity, valid inside an expectation value. Remarkably, this concept was derived independently and simultaneously by Leo Kadanoff [40] (who was working in critical phenomena), by Kenneth Wilson [41] (who was interested in the short distance singularities arising in Feynman diagrams), and by Alexander Polyakov [42, 43] (also working in critical phenomena). In Eq.(5) $\{\Delta_j\}$ are the scaling dimensions of the operators $\{O_j\}$. The coefficients C_{jkl} are (like the dimensions) universal numbers. In a follow up paper Polyakov showed that if the theory has conformal invariance, i.e. scale invariance augmented by conformal transformations which preserve angles, then, provided the operators O_j are suitably normalized, the coefficients C_{jkl} of the OPE are determined by a three point correlator

$$\langle O_j(x) O_k(y) O_l(z) \rangle = \frac{C_{jkl}}{|x - y|^{\Delta_{jk}} |y - z|^{\Delta_{kl}} |z - x|^{\Delta_{lj}}} \quad (6)$$

where $\Delta_{jk} = \Delta_j + \Delta_k - \Delta_l$. These results constitute the beginnings of Conformal Field Theory. In a nontrivial check, Kadanoff and Ceva showed that the OPE holds for the local observables of the 2D Ising model [44].

3.3.3. Fixed Points

The next and crucial step in the development of the renormalization Group was made by Kenneth Wilson. Wilson was a high-energy theorist who wanted to know how to properly define a quantum field theory and the physical meaning of renormalization.

In a Lorentz invariant quantum field theory one is interested in the computation of the expectation value of time-ordered operators. In the case of a self-interacting scalar field $\phi(x)$ in D -dimensional Euclidean space-time, obtained by analytic continuation from Minkowski space-time to imaginary time, the observables are computed from the functional (or path) integral by functional differentiation of the partition function

$$Z = \int \mathcal{D}\phi \exp\left(-S(\phi, \partial_\mu \phi) + \int d^D x J(x)\phi(x)\right) \quad (7)$$

with respect to the local sources $J(x)$. For a scalar field the Euclidean action is

$$S = \int d^D x \left[\frac{1}{2} (\partial_\mu \phi(x))^2 + \frac{m^2}{2} \phi^2(x) + \frac{\lambda}{4!} \phi^4(x) \right] \quad (8)$$

which has the same form as the free energy of the Landau-Ginzburg theory of phase transitions shown in Eq.(2). It is apparent that the Landau-Ginzburg theory is the classical limit of the theory of the scalar field whose partition function is a sum over all histories of the field. It is easy to see that an expansion of the partition function (or of a correlator) in powers of the coupling constant λ can be cast in the form of a sum of Feynman diagrams. To lowest order in λ a typical Feynman diagram involves a one-loop integral in momentum space of the form

$$I(p) = \int \frac{d^D q}{(2\pi)^D} \frac{1}{(q^2 + m^2)((q - p)^2 + m^2)} \quad (9)$$

As noted by Wilson in his Nobel Lecture [33], this integral has large contributions from the IR region of small momenta $q \sim 0$, but for any dimension $D \geq 4$ has a much larger contribution from the UV region of large momenta, which requires the introduction of a UV cutoff Λ in momentum space (or a lattice spacing a in real space by defining the theory on a hypercubic lattice). In quantum field theory one then has to require that somehow one takes the limit $a \rightarrow 0$ (or $\Lambda \rightarrow \infty$). To take this limit in the field theory is very much analogous to the definition of a conventional integral in terms of a limit of a Riemann sum. The difference is that this is a *functional integral*. While for a function of bounded variation in a finite interval (a, b) the limit of a partition of the interval into N steps each of length Δx , such that $N\Delta x = b - a$, exists and defines the integral of the function

$$\lim_{\Delta x \rightarrow 0} \lim_{N \rightarrow \infty} \sum_{j=1}^N f(x_j) \Delta x_j = \int_a^b dx f(x), \quad (10)$$

the analogous statement does not obviously exist in general for a *functional integral*, i.e. an integral over a space of functions which is what is required. In fact, although thousands of integrals of a function are known to exist, there are extremely few examples for a functional integral. Moreover, in order to take the *continuum limit* the lattice spacing must approach zero, $a \rightarrow 0$. This means that physical scales, such as the correlation length ξ , must diverge in lattice units so that they can be fixed in physical units. But to do that one has to be asymptotically close to a continuous phase transition! Hence, the problem of defining a quantum field theory is *equivalent* to the problem of critical phenomena at a continuous phase transition!

Wilson gave a systematic formulation to the Renormalization Group by generalizing the earlier ideas introduced by Kadanoff and the earlier work by Gell-Mann and Low. Wilson's key contribution was the introduction of the concept of a *fixed point* of the Renormalization Group transformation [28, 29, 30, 31]. As we saw, the block-spin transformation is a procedure for coarse graining the degrees of freedom of a physical system resulting in a renormalization of the coupling constants. Upon the repeated action of the RG transformation its effect can be pictured as a *flow* in the space of coupling constants. However, in addition of integrating-out short distance degrees of freedom one needs to restore the units of length which have changed under that process. This requires a rescaling of lengths. Once this is done, Wilson showed that the resulting RG flows necessarily have *fixed points*, special values of the couplings which are invariant (fixed) under the action of the RG transformation. He then deduced that at a fixed point the theory has no scales, aside from the linear size L of the system and the microscopic UV cutoff (the lattice spacing a in a spin system).

This analysis means that for length scales long compared to $a \rightarrow 0$ but short compared to $L \rightarrow \infty$ the theory acquired a new, emergent, symmetry: scale invariance. Therefore, at a fixed point the correlators of all local observables must be homogeneous functions (hence, must scale).

3.3.4. Universality

A crucial consequence of the concept of the fixed point is that phase transitions can be classified into *universality classes*. Universality means that a large class of physical systems with different microscopic properties have fixed points with the same properties, i.e. the same scaling dimensions, operator product expansions and correlation functions at long distances independent on how they are defined microscopically. Although the renormalization group transformation is a transformation scheme that we define and, because of that the location in coupling constant

space of the fixed point itself does depends on the scheme we choose, its universal properties are the same. Thus, universality classes depend only on features such as the space (and spacetime) dimension and the global symmetries of the system. But the systems themselves may be quite different. This we speak if the Ising universality class in 2D, on the superfluid (or XY) transition class in 3D, etc. This concept, which originated in the theory of phase transition, has been adopted and generalized in the development of conformal field theory.

3.3.5. RG flows

Combined with the condition that the correlators decay at long separations, homogeneity implies that the correlators must have the form of Eq.(3). In addition, this equation also implies that at a fixed point the operators (the local observables) have certain scaling dimensions. Let us consider a theory close to a fixed point whose action we will denote by S^* . Let $\{O_j(x)\}$ be a complete set of local observables whose scaling dimensions are $\{\Delta_j\}$. The total action of the theory close to the fixed point then can be expanded as a linear combination of the operators with dimensionless coupling constants $\{g_j\}$

$$S = S^* + \int d^D x \sum_j g_j a^{\Delta_j - D} O_j(x) \quad (11)$$

Under a change of length scale $x \rightarrow x' = bx$, with $b > 1$, the operators (which must transform homogeneously) change as $O_j(bx) = b^{-\Delta_j} O_j(x)$. Since the phase space changes as $d^D x' = b^D d^D x$, we can keep the form of the action provided the coupling constants also change to compensate for these changes as $g'_j = b^{D-\Delta_j} g_j$. Let $b = |x'|/|x| = 1 + da/a$, where da is an infinitesimal change of the UV cutoff a . Then, if we integrate-out the degrees of freedom in the range $a < |x| < a + da$, the rate of change of the coupling constants $\{g_j\}$ under this rescaling is

$$a \frac{dg_j}{da} \equiv \beta(g_j) = (D - \Delta_j)g_j + \dots \quad (12)$$

which we recognize as a Gell-Mann Low beta function for each coupling constant.

This result says that if the scaling dimension $\Delta_j < D$, then the renormalized coupling will *increase* as we increase the length scale, $g'_j > g_j$, and along this direction in coupling constant space the RG flows *away* from the fixed point. Conversely, if $\Delta_j > D$ the renormalized coupling flows to smaller values, $g'_j < g_j$ and the RG flows *into* the fixed point. We then say that an operator is *relevant* if its scaling dimension satisfies $\Delta_j < D$, and that it is *irrelevant* if $\Delta_j > D$. If $\Delta_j = D$ then we say that operator is *marginal*.

To go beyond this simple dimensional analysis one has to include the effects of fluctuations. To lowest orders in the couplings one finds [45]

$$a \frac{dg_j}{da} = (D - \Delta_j)g_j + \sum_{k,l} C_{jkl} g_k g_l + \dots \quad (13)$$

where $\{C_{jkl}\}$ are the coefficients of the OPE shown in Eq.(6). This expression is the general form of a perturbative renormalization group and it is valid close enough to a fixed point.

Wilson and Fisher [30] used a similar approach to analyze how fluctuations alter the results of the Landau-Ginzburg theory. They considered the partition function of Eq.(7) with the action of Eq.(8). Instead of working in real space they considered the problem in momentum space and partitioned the field configurations into slow and fast modes

$$\phi(x) = \phi_<(x) + \phi_>(x) \quad (14)$$

where $\phi_>(x)$ are configurations whose Fourier components have momenta in the range $b\Lambda < |p| < \Lambda$, where Λ is a UV momentum cutoff and $b < 1$. Hence, if we choose $b \rightarrow 1$, the fast modes $\phi_>(x)$ have components in a thin momentum shell near the UV cutoff Λ . Conversely, the slow modes $\phi_<(x)$ have momenta in the range $0 \leq |p| < b\Lambda$. One can then use Feynman diagrams to integrate out the fast modes and derive an effective low-energy action with renormalized couplings. In the case of a free field theory (with $\lambda = 0$) the scaling dimension of the ϕ^4 operator is $\Delta_4 = 2(D - 2)$ whereas the ϕ^2 operator has dimension $\Delta_2 = D - 2$. Upon defining a dimensionless mass and coupling constant by $m^2 = t\Lambda^2$ and $\lambda = g\Lambda^{4-D}$, the beta functions are found to be [10]

$$\beta(t) = -\Lambda \frac{dt}{d\Lambda} = 2t + \frac{g}{2} - \frac{gt}{2} + \dots \quad (15)$$

$$\beta(g) = -\Lambda \frac{dg}{d\Lambda} = (4 - D)g - \frac{3}{2}g^2 + \dots \quad (16)$$

(where we absorbed an uninteresting factor if the definition of the coupling constant).

The RG flows of Eq.(16) show that the free-field (Gaussian) fixed point at $g = 0$ is stable for $D > 4$ and the asymptotic IR behavior is the same as predicted by the Landau-Ginzburg theory. However, for $D < 4$, the free-field fixed point becomes unstable and a new fixed point arises at $g^* = \frac{2}{3}\epsilon + O(\epsilon^2)$, where we have set $\epsilon = 4 - D$. This is the Wilson-Fisher fixed point. At this fixed point the correlation length diverges with an exponent $\nu = \frac{1}{2} + \frac{\epsilon}{12} + O(\epsilon^2)$, which deviates from the predictions of the Landau-Ginzburg theory. The small parameter of this expansion is ϵ , and this is known as the ϵ expansion.

The Wilson-Fisher (WF) fixed point is an example of a non-trivial fixed point at which the correlation length is divergent. It has only one relevant operator, the mass term, which in the IR flows into the symmetric phase for $t > 0$ and flows to the broken symmetry for $t < 0$. Conversely, in the UV it flows into the WF fixed point. For these reasons condensed matter physicists say that this is an IR unstable (or critical) fixed point while high-energy physicists say that it is the UV fixed point. At this fixed point a non-trivial field theory can be defined with non-trivial interactions. UV fixed points also define examples of what in high-energy physics are called renormalizable field theories and can be used to define a continuum field theory.

The $D = 4$ -dimensional theory is special in that the ϕ^4 operator is marginal. As can be seen in Eq.(16), at $D = 4$ the beta function for the dimensionless coupling constant g does not have a linear term and is quadratic in g . In this case the operator is marginally irrelevant, and its beta function has the same behavior as the beta function of Gell-Mann and Low for QED. Such theories are said to have a “triviality problem” since, up to logarithmic corrections to scaling, there are no interactions in the IR and, conversely, become large in the UV.

There are also fixed points at which the correlation length $\xi \rightarrow 0$. These fixed points are IR stable (and in a sense trivial). These stable fixed points are sinks of the IR RG flows. Such fixed points define stable phases of matter, e.g. the broken symmetry state, the symmetric (or unbroken state), etc. However in the UV they are unstable and in high-energy physics such fixed points correspond to non-renormalizable field theories.

More sophisticated methods are needed to go beyond the lowest order beta functions of Eq.(13), and the computation of critical exponents beyond the leading non-trivial order. Possibly the record high-precision calculations have been done for ϕ^4 theory for which the beta function is known to $O(\epsilon^5)$. This has been achieved using the method of dimensional regularization [46, 47, 48] (with minimal subtraction). Special resummation methods (Borel-Padé) have been used to do these calculations in $D = 3$ dimensions [49]. Remarkably, these results are so precise that in the case of the superfluid transition, which is well described by a ϕ^4 theory with a complex field, the results could only be tested in the microgravity environment of the International Space Station!

3.3.6. Asymptotic Freedom

There are several physical systems whose beta function has the form

$$\beta(g) = Ag^2 + \dots \quad (17)$$

The coupling constant has a different interpretation in each theory and the constant $A > 0$, opposite to the sign of the beta function found in QED and ϕ^4 theory in $D = 4$ dimensions. This beta function means that while the associated operator is marginal, with this sign it is actually marginally relevant. This also means that the fixed point is unstable in the IR but the departure from the fixed point is logarithmically small. Conversely, in the UV the RG flows into the fixed point and the effective constant is weak at short distances. This is the origin of the term asymptotic freedom [50]. The paradigmatic examples of theories with a beta function of this form are the Kondo problem, the 2D non-linear sigma model, and the $D = 4$ dimensional Yang-Mills non-abelian gauge theory.

The Kondo problem is the theory of a localized spin-1/2 degree of freedom in a metal. The electrons of the metal couple to this quantum impurity through an exchange interaction of the impurity and the magnetic moment density of the mobile electrons in the metal with coupling constant J . This problem is actually one dimensional since only the s wave channel of the mobile (conduction band) electrons actually couple to the localized impurity. In 1970 Philip Anderson developed a theory of the Kondo problem in terms of the renormalization of the Kondo coupling constant g as a function of the energy scale [51]. Anderson used perturbation theory in J to progressively integrated out the modes of the conduction electrons close to an effective

bandwidth E_c and found that the beta function has the form of Eq.(17)

$$E_c \frac{dJ}{dE_c} = -\rho J^2 + \dots \quad (18)$$

where ρ is the density of states at the Fermi energy of the conduction electrons. This work implied that the free-impurity fixed point is IR unstable and that the effective coupling constant J increases as the energy cutoff E_c is lowered. He argued that at some energy scale, the Kondo scale, perturbation theory breaks down and that there is a crossover to a strong coupling regime which is not accessible in perturbation theory.

Shortly thereafter, in 1973 Wilson developed a numerical renormalization group approach which showed that the Kondo problem is indeed a crossover from the free impurity fixed point to the “renormalized” Fermi liquid [32]. In addition, Wilson used the numerical renormalization group to examine the approach to the strong coupling fixed point and showed that it is characterized by a Wilson ratio, a universal number obtained from the low temperature specific heat and the impurity magnetic susceptibility (in suitable units). Wilson’s numerical RG predicted a number close to 2π for this ratio. In 1980 N. Andrei and P. Wiegmann showed (independently) that the Kondo problem is an example of an integrable field theory that can be solved by the Bethe ansatz [52, 53]. Their exact result was consistent with Wilson’s RG, including the numerical value of the Wilson ratio.

In 1972 Gerard ’t Hooft and Martinus Veltman showed that Yang-Mills gauge theory is renormalizable [46]. This groundbreaking result opened the door to use quantum field theory to develop the theory of strong interactions in particle physics known as Quantum Chromodynamics (QCD). In 1973 David Gross and Frank Wilczek [50] and, independently, David Politzer [54] computed the renormalization group beta function of Yang-Mills theory with gauge group G and found it to be of the same form as Eq.(17),

$$\Lambda \frac{dg}{d\Lambda} = -\frac{g^3}{16\pi^2} \frac{11}{3} C_2(G) + \dots \quad (19)$$

where g is the Yang-Mills coupling constant and Λ is a UV momentum scale. Here $C_2(G)$ is the quadratic Casimir for a gauge group G . For $SU(3)$, the case of physical interest, $C_2(SU(3)) = 3$. This result implies that under the RG at large momenta (short distances) the Yang-Mills coupling constant flows to zero (up to logarithmic corrections). This result holds in the presence of quarks provided the number of quark flavors is less than a critical value. Hence at short distances the effective coupling is weak. Gross and Wilczek called this phenomenon asymptotic freedom. This behavior was consistent with the observation of weakly coupled quarks in deep inelastic scattering experiments. However, the flip side of asymptotic freedom is that at low energies (long distances) the coupling constant grows without limit, which implies that at low energies perturbation theory is not applicable. This strong infrared behavior suggested that in QCD quarks are permanently confined in color neutral bound states (hadrons). However, unlike the Kondo problem we just discussed, QCD is not an integrable theory (so far as we know) and to show that it confines has required the development of Lattice Gauge Theory [55, 56]. To this date the best evidence for quark confinement has been obtained using large-scale Monte Carlo simulations in Lattice Gauge Theory [57].

We close this subsection with a discussion of an important case: the non-linear sigma models. The $O(N)$ non-linear sigma model is the continuum limit of the classical Heisenberg model for a spin with N components. Historically, the non-linear sigma model is the effective field theory for pions in particle physics. We will discuss its role in the theory of quantum antiferromagnets in subsection 4.3 and especially in the case of quantum antiferromagnetic spin chains in subsection 5.2.

The simplest non-linear sigma model is a theory of an N -component scalar field $\mathbf{n}(x)$ which satisfies the unit length local constraint, $\mathbf{n}^2(x) = 1$. The Euclidean Lagrangian is

$$\mathcal{L} = \frac{1}{2g} (\partial_\mu \mathbf{n}(x))^2 \quad (20)$$

where g is the coupling constant (the temperature in the classical Heisenberg model). At the classical level, i.e. in the broken symmetry phase, where $\langle \mathbf{n} \rangle \neq 0$, this model describes the $N - 1$ massless modes (Goldstone bosons) of the spontaneously broken $O(N)$ symmetry. Dimensional analysis shows that the coupling constant has units of ℓ^{D-2} . Hence, we expect to find marginal behavior at $D = 2$. In 1975 Alexander Polyakov used a momentum shell renormalization group

in $D = 2$ dimensions and showed that the beta function of this model is (here a is the short-distance cutoff) [58]

$$\beta(g) = a \frac{dg}{da} = \frac{N-2}{2\pi} g^2 + O(g^3) \quad (21)$$

Hence, in $D = 2$ dimensions also this theory is asymptotically free. As in the other examples we just discussed, asymptotic freedom here also implies that the coupling constant g grows to large values in the low-energy (long-distance) regime. In close analogy with Yang-Mills theory in $D = 4$ dimensions, Polyakov conjectured that the $O(N)$ non-linear sigma model also undergoes dynamical dimensional transmutation [50], that the global $O(N)$ symmetry is restored and that for all values of the coupling constant g the theory is in a massive with a finite correlation length $\xi \sim \exp((N-2)/2\pi g)$. Extensive numerical simulations, used to construct a renormalization group using Monte Carlo simulations [59], showed that there is indeed a smooth crossover from the weak coupling (low temperature) regime to the high temperature regime where the correlation length is finite. The non-linear sigma model is a renormalizable field theory in $D = 2$ dimensions [60]. For $D > 2$ dimensions it can be studied using the $2 + \epsilon$ expansion [60, 49], which predicts the existence of a nontrivial UV fixed point and a phase transition from a Goldstone phase to a symmetric phase.

It turns out that there is a significant number of asymptotically free non-linear sigma models in $D = 2$ dimensions, many of physical interest [61], in particular non-linear sigma models whose target manifold is a coset space, a quotient of a group G and a subgroup H . The $O(N)$ non linear sigma model is an example since the broken symmetry space leaves the $O(N-1)$ subgroup unbroken (the manifold of the Goldstone bosons). In that case the quotient is $O(N)/O(N-1)$ which is isomorphic to the $N-1$ dimensional sphere S_{N-1} . In later sections we will discuss other examples in which more general non-linear sigma models play an important role.

Models on coset spaces arise in the theory of Anderson localization in $D = 2$ dimensions. Anderson localization is the problem of a fermion (an electron) in a disordered system in which the electron experiences a random electrostatic potential. In the limit of strong disorder Philip Anderson showed that all one-particle states are exponentially localized and the diffusion constant (and the conductivity) vanishes[62]. There was still the question of when it is possible for the electron to have a finite diffusion constant (and conductivity). In $D = 2$ dimensions the conductivity is a dimensionless number which suggests that this may be the critical dimension for diffusion. Abrahams, Anderson, Licciardello and Ramakrishnan used a weak disorder calculation to construct a scaling theory that implied that in $D = 2$ dimensions the RG flow of the conductivity at long distances (large samples) flows to zero and all states are localized [63]. Shortly thereafter Wegner gave strong arguments that showed that the existence of diffusion implied that there are low-energy “diffuson” modes which behaved as Goldstone modes of a non-linear sigma model on the quotient manifold $O(N_+ + N_-)/O(N_+) \times O(N_-)$ in the “replica limit” $N_\pm \rightarrow 0$ [64]. A field theory approach to this non-linear sigma model was developed by McKane and Stone [65] and by Hikami [66].

4. Quantum Criticality

Quantum criticality is the theory of a phase transition of a system (e.g. a magnetic system) at zero temperature that occurs as a coupling constant (or parameter) is varied continuously. Although not necessarily under that name, this question has existed as a conceptual problem for a long time, In particular, already in 1973 Wilson considered the problem of the behavior of quantum field theories below four spacetime dimensions and their phase transitions [67].

The modern interest in condensed matter physics stems from discoveries made since the late 1980s. Since that time the behavior of condensed matter systems at a *quantum* critical point has emerged as a major focus in the field. There were several motivations for this problem. One was (and still is) to understand the behavior of quantum antiferromagnets in the presence of frustrating interactions. Frustrating interactions are interactions which favor incompatible types of antiferromagnetic orders. The result is the presence of intermediate non-magnetic “valence bond” phases that favor the formation of spin singlets between nearby spins. These phases typically either break the point group symmetry of the lattice or are spin liquids (which will be discussed below). Another motivation is that doped quantum antiferromagnets typically harbor superconducting phases (among others) whose high-temperature behavior is a “strange” metal that violates the basic assumptions (and behaviors) of Fermi liquids. The most studied version of this problem is the case of the copper oxide high temperature superconductors. It was conjectured that there is a quantum critical point inside the superconducting phase at which the

antiferromagnetic order (or other orders) disappears and which may be the reason for the strange metal behavior above the superconducting critical temperature. Many of these questions are discussed in depth by Sondhi and coworkers [68] and in the textbook by Sachdev [69].

4.1. Dynamic Scaling

We will consider a general quantum phase transition and assume that it is scale invariance. However, except for the case of relativistic quantum field theories, in condensed matter systems space and time do not need to scale in the same way. Let us assume that the system of interest has just one coupling constant g and that the system of interest has a quantum phase transition (at zero temperature) at some critical value g_c between two phases, for instance one with a spontaneously broken symmetry and a symmetric phase. If the quantum phase transition is continuous then the correlation length ξ will diverge at g_c and so will the correlation time ξ_t . However these two scales are in general different and do not necessarily diverge at the same rate. So, in general, if some physical quantity is measured at the quantum critical point at some momentum \mathbf{p} and frequency ω , the length scale of the measurement is $2\pi/|\mathbf{p}|$ and at a frequency is $\omega \sim |\mathbf{p}|^z$, where z is the dynamic critical exponent. Let us say that we measure the observable \mathcal{O} at momentum \mathbf{p} and frequency ω at g_c . Scale invariance in both space and time means that at g_c the observable $\mathcal{O}(\mathbf{p}, \omega)$ at momentum \mathbf{p} and frequency ω must scale as

$$\mathcal{O}(\mathbf{p}, \omega) = |\mathbf{p}|^{-\Delta_O} \tilde{\mathcal{O}}(|\mathbf{p}|^z / \omega) \quad (22)$$

where Δ_O is the scaling dimension of the observable \mathcal{O} .

The situation changes at finite temperature T . A quantum field theory at temperature T is described by a path integral on a manifold which along the imaginary time direction τ is finite of length $2\pi/T$ and periodic for a theory bosonic fields and anti-periodic for fermionic fields [10]. Since the imaginary time direction is finite, the behavior for correlation times $\xi_\tau < 2\pi/T$ and $\xi_\tau > 2\pi/T$ must be different. Indeed, in the first regime the behavior is essentially the same as at $T = 0$, while in the second it should be given by the classical theory in the same space dimension. At the quantum critical point g_c there is only one time scale $\xi_\tau \sim 2\pi/T$ and only one length scale $\xi \sim (2\pi/T)^{1/z}$.

4.2. The Ising Model in a Transverse Field

The prototype of the quantum phase transition is the Ising model in a transverse field. This model describes a system of spin-1/2 degrees of freedom with ferromagnetic interactions with uniaxial anisotropy in the presence of a transverse uniform magnetic field. The Hamiltonian is

$$H = -J \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \sigma_3(\mathbf{r}) \sigma_3(\mathbf{r}') - h \sum_{\mathbf{r}} \sigma_1(\mathbf{r}) \quad (23)$$

where J and h are the exchange coupling constant and the strength of the transverse field, respectively. Here σ_1 and σ_3 are the two Pauli matrices defined on the sites $\{\mathbf{r}\}$ of a lattice with ferromagnetic interactions between spins on nearest neighboring sites. The Hilbert space is the tensor product of the states of the spins at each site of the lattice. At each site there are two natural bases of states: the eigenstates of σ_3 , which we denote by $|\uparrow\rangle$ and $|\downarrow\rangle$ (whose eigenvalues are ± 1), and the eigenstates of σ_1 , which we denote by $|\pm\rangle$ (whose eigenvalues are also ± 1).

It is well known that the Ising Model in a Transverse Field on a hypercubic lattice in D dimensions is equivalent to a classical Ising model in $D + 1$ dimensions [70, 71]. These two models are related through the transfer matrix. Indeed, a classical Ising model can be regarded as a path integral representation of the quantum model in one dimension less. For simplicity we will see how this work for the 2D the classical ferromagnetic Ising model of Eq.(1), but the construction is general. We will regard the configuration of spins on a row of the 2D lattice as a state of a quantum system, and the set of states on all rows as the evolution of the state along the perpendicular direction that we will regard as a discretized imaginary time. The contribution from two adjacent rows to the partition function defines the matrix element of a matrix between two arbitrary configurations. In Statistical Physics this matrix is known as the Transfer Matrix \hat{T} and the full partition function (with periodic boundary conditions) is

$$Z = \text{tr} \hat{T}^{N_\tau} \quad (24)$$

where N_τ is the number of rows. For the case of the ferromagnetic Ising model (actually, for any unfrustrated model) the transfer matrix can always be constructed to be hermitian. This property

holds in fact for any theory that satisfies a property known as reflection positivity which requires that all (suitably defined) correlation functions be positive. For theories of this type, and the Ising model is an example, the matrix elements of the transfer matrix can be identified with the matrix element of the evolution operator of a quantum theory for a small imaginary time step [70]. Also, the positivity of the correlators is equivalent to the condition of positivity of the norm of states in the quantum theory.

For the classical models that satisfy these properties, all directions of the lattice are equivalent. Moreover, asymptotically close to the critical point, the behavior of all the correlators becomes isotropic, i.e. invariant under the symmetries of Euclidean space. This means that the arbitrary choice of the direction for the transfer matrix is irrelevant. Consequently, in the quantum model its equal-time correlators behave the same way as its correlation functions in imaginary time. In other words space and time behave in the same way and the quantum theory is relativistically invariant. This implies at the quantum critical point the energy $\varepsilon(\mathbf{p})$ of its massless excitations should behave as $\varepsilon(\mathbf{p}) = v |\mathbf{p}|$. In a relativistic theory the dynamical critical exponent must be $z = 1$ and the coefficient v is the speed of the excitations (the “speed of light”). We should note that this is not necessarily always the case. There are in fact classical systems, e.g. liquid crystals [72], which are spatially anisotropic and map onto quantum mechanical theories in one less dimension for which the dynamical critical exponent $z \neq 1$. One such example are the Lifshitz transitions of nematic liquid crystals in three dimensions and the associated quantum Lifshitz model in $D = 2$ dimensions, for which the dynamical exponent is $z = 2$ [73].

Just as in the classical counterpart in $D + 1$ dimensions, the quantum model in $D \geq 1$ has two phases: a broken symmetry ferromagnetic phase for $J \gg h$ and a symmetric paramagnetic phase for $h \gg J$. In the symmetric phase the ground state is unique (asymptotically it is the eigenstate of σ_1 with eigenvalue +1), while in the broken symmetry phase the ground state is doubly degenerate (and asymptotically it is an eigenstate of σ_3) and there is a non-vanishing expectation value of the local order parameter $\langle \sigma_3(\mathbf{r}) \rangle \neq 0$. In the symmetric phase the correlation function of the local order parameter decays exponentially with distance with a correlation length ξ , as does the connected correlation function in the broken symmetry phase. The model has a continuous quantum phase transition at a critical value of the ratio h/J . For general space dimensions $D > 1$ this model is not exactly solvable and much of what we know about it is due to large-scale numerical simulations.

This problem was solved exactly in one-dimension [23] using the Jordan-Wigner transformation that maps a one dimensional quantum spin system to a theory of free fermions [22]. The fermion operators at site j are

$$\chi_1(j) = K(j-1)\sigma_3(j), \quad \chi_2(j) = i\hat{K}(j)\sigma_3(j) \quad (25)$$

where $K(j)$ is the kink creation operator (i.e. the operator that creates a domain wall between sites j and $j+1$ [70], and is given by

$$K(j) = \prod_{n \leq j} \sigma_1(n) \quad (26)$$

The operators $\chi_1(j)$ and $\chi_2(j)$ are hermitian, $\chi_j^\dagger(n) = \chi_j(n)$, and obey the *anticommutation* algebra

$$\{\chi_j(n), \chi_{j'}(n')\} = 2\delta_{jj'}\delta_{nn'} \quad (27)$$

Hence, they are fermionic operators are hermitian, anti-commute with each other and square to the identity. Operators of this type are called *Majorana fermions*.

Alternatively, we can use the more conventional (Dirac) fermion operators $c(n)$ and its adjoint $c^\dagger(n)$ which are related to the Majorana fermions as

$$c(n) = \chi_1(n) + i\chi_2(n), \quad c^\dagger(n) = \chi_1(n) - i\chi_2(n) \quad (28)$$

which obey the standard anticommutation algebra

$$\{c(n), c(n')\} = \{c^\dagger(n), c^\dagger(n')\} = 0, \quad \{c(n), c^\dagger(n')\} = \delta - nn' \quad (29)$$

In this sense, a Majorana fermion is half of a Dirac fermion.

In terms of the Majorana operators the Hamiltonian of Eq.(23) becomes

$$H = i \sum_j \chi_1(j)\chi_2(j) + ig \sum_j \chi_2(j)\chi_1(j+1) \quad (30)$$

where we have rescaled the Hamiltonian by a factor of h and the coupling constant is $g = J/h$. Here we have not specified the boundary conditions (which depend on the fermion parity). Qualitatively, the Majorana fermions can be identified with the domain walls of the classical models. In the Ising model the number of domain walls on each row is not conserved but their parity is. Likewise, the number of Majorana fermions N_F is not conserved either but the fermion parity, $(-1)^{N_F}$, is conserved.

It is an elementary exercise to show that the spectrum of this theory has a gap $G(g)$ which vanishes at $g_c = 1$ as $G(g) \sim |g - g_c|^\nu$, with an exponent $\nu = 1$. Since the Hamiltonian of Eq.(30) is quadratic in the Majorana operators, these operators obey linear equations of motion. In the scaling regime we take the limit of the lattice spacing $a \rightarrow 0$ and the coupling constant $g \rightarrow g_c = 1$, while keeping the quantity $m = (g - g_c)/a$ fixed. In this regime, the two-component hermitian spinor field $\chi = (\chi_1, \chi_1)$, and $\chi^\dagger = \chi$, satisfies a Dirac equation

$$(i\partial - m)\chi = 0 \quad (31)$$

where we set the speed $v = 1$, and where defined the 2×2 Dirac gamma-matrices $\gamma_0 = \sigma_2$, $\gamma_1 = i\sigma_3$, and $\gamma_5 = \sigma_1$. Upon defining $\bar{\chi} = \chi^T \gamma_0$, we find that the Lagrangian of this field theory is

$$\mathcal{L} = \bar{\chi} i\partial \chi - \frac{1}{2} m \bar{\chi} \chi \quad (32)$$

which indeed becomes massless at the quantum phase transition of the Ising spin chain. For these considerations, we say that the phase transition of the Ising model (2D classical or 1D quantum) is in the universality class of massless Majorana fermions where $m \rightarrow 0$. In Eq.(32) we have used the standard Feynman slash notation, $\partial = \gamma_\mu a^\mu$, where a^μ is a vector.

4.3. Quantum Antiferromagnets and Non-Linear Sigma Models

As we noted above, the discovery of high temperature superconductors in the copper oxide compounds prompted the study of the behavior of these strongly correlated materials at low temperatures and of possible quantum phase transitions which they may host. The prototypical cuprate material La_2CuO_4 is a quasi two-dimensional Mott insulator which exhibits long-range antiferromagnetic order below a critical temperature T_c . A simple microscopic model is a spin-S quantum Heisenberg antiferromagnet on the 2D square lattice of the Cu atoms, whose Hamiltonian is

$$H = \frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} J(|\mathbf{r} - \mathbf{r}'|) \mathbf{S}(\mathbf{r}) \cdot \mathbf{S}(\mathbf{r}') \quad (33)$$

where \mathbf{S} are the spin-S angular momentum operators. We will consider the case where the exchange interaction for nearest neighbors J is dominant and a weaker $J' \ll J$ for next nearest neighbors. In this section we do not consider the regime $J' \simeq J$ in which the interactions compete for incompatible ground states due to frustration effects.

4.3.1. Spin coherent states

The simplest way to see the physics of this antiferromagnet is to construct a path-integral representation for a spin-S system using spin coherent states [74, 75, 76]. For details see Ref.[9] which we follow here. A coherent state of the $(2S + 1$ -dimensional) spin-S representation of $SU(2)$ is the state $|\mathbf{n}\rangle$, labeled by the spin polarization unit vector \mathbf{n}

$$|\mathbf{n}\rangle = e^{i\theta(\mathbf{n}_0 \times \mathbf{n}) \cdot \mathbf{S}} |S, S\rangle \quad (34)$$

where $\mathbf{n}^2 = 1$. The states of the spin-S representation are spanned by the eigenstates of S_3 and S^2 ,

$$S_3 |S, M\rangle = M |S, M\rangle, \quad S^2 |S, M\rangle = S(S+1) |S, M\rangle \quad (35)$$

and $|S, S\rangle$ is the highest weight state with eigenvalues S and $S(S+1)$. In Eq.(34) \mathbf{n}_0 is a unit vector along the axis of quantization (the direction \mathbf{e}_3), and θ is the colatitude, such that $\mathbf{n} \cdot \mathbf{n}_0 = \cos \theta$. Two spin coherent states, $|\mathbf{n}_1\rangle$ and $|\mathbf{n}_2\rangle$, are not orthonormal,

$$\langle \mathbf{n}_1 | \mathbf{n}_2 \rangle = e^{i\Phi(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_0) S} \left(\frac{1 + \mathbf{n}_1 \cdot \mathbf{n}_2}{2} \right)^S \quad (36)$$

where $\Phi(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_0)$ is the area of the spherical triangle of the unit sphere spanned by the unit vectors \mathbf{n}_1 , \mathbf{n}_2 and \mathbf{n}_0 . However, there is an ambiguity in the definition of the area of the spherical

triangle since the sphere is a 2-manifold without boundaries: if the “inside” triangle has spherical area Φ , the complement (“outside”) triangle has area $4\pi - \Phi$. Thus, the ambiguity of the phase prefactor of Eq.(36) is

$$e^{i4\pi S} = 1 \quad (37)$$

since S is an integer or a half-integer. So, the quantization of the representations of $SU(2)$ makes the ambiguity unobservable. In addition, the spin coherent states $|\mathbf{n}\rangle$ satisfy the resolution of the identity

$$I = \int |\mathbf{n}\rangle\langle\mathbf{n}| \left(\frac{2S+1}{4\pi} \right) \delta(\mathbf{n}^2 - 1) d^3n \quad (38)$$

and

$$\langle \mathbf{n}|S|\mathbf{n}\rangle = S \mathbf{n} \quad (39)$$

4.3.2. Path integral for a spin- S degree of freedom

As an example consider problem of a spin- S degree of freedom coupled to an external magnetic field $\mathbf{B}(t)$ that varies slowly in time. The (time-dependent) Hamiltonian is given by the Zeeman coupling

$$H(t) = \mathbf{B}(t) \cdot \mathbf{S} \quad (40)$$

As usual, the path-integral is obtained by inserting the (over-complete) set of coherent states at a large number of intermediate times. The resulting path integral is a sum of the histories of the spin polarization vector $\mathbf{n}(t)$

$$Z = \text{tr } \exp\left(i \int_0^T dt H(t)\right) = \int \mathcal{D}\mathbf{n} \exp(iS[\mathbf{n}]) \prod_t \delta(\mathbf{n}^2(t) - 1) \quad (41)$$

where the action is

$$S = S S_{WZ}[\mathbf{n}] - S \int_0^T dt \mathbf{B}(t) \cdot \mathbf{n}(t) \quad (42)$$

where $S_{WZ}[\mathbf{n}]$ is the Wess-Zumino action

$$S_{WZ}[\mathbf{n}] = \int_0^T dt \mathbf{A}[\mathbf{n}] \cdot \partial_t \mathbf{n} \quad (43)$$

$$= \int_0^1 d\tau \int_0^T dt \mathbf{n}(t, \tau) \cdot \partial_t \mathbf{n}(t, \tau) \times \partial_\tau \mathbf{n}(t, \tau) \quad (44)$$

where $\mathbf{A}[\mathbf{n}]$ is the vector potential of a Dirac magnetic monopole (of unit magnetic charge) at the center of the unit sphere. The vector potential $\mathbf{A}[\mathbf{n}]$ has a singularity associated with the Dirac string of the monopole. We can write an equivalent expression which is singularity-free using Stokes Theorem. We did this in the second line of Eq.(44) which required to extend the circulation of \mathbf{A} on the closed path described by $\mathbf{n}(t)$ to the flux of the vector potential through the submanifold Σ of the unit sphere S_2 whose boundary is the history $\mathbf{n}(t)$, i.e. the area of Σ . The smooth (and arbitrary) extension of configuration $\mathbf{n}(t)$ to the interior of Σ is done by defining $\mathbf{n}(t, \tau)$ such that $\mathbf{n}(t, 1) = \mathbf{n}_0$, $\mathbf{n}(t, 0) = \mathbf{n}(t)$, and $\mathbf{n}(0, \tau) = \mathbf{n}(T, \tau)$. Since S_{WZ} is the area of the submanifold Σ of the unit sphere S_2 , just as in Eq.(37), here too there is an ambiguity of 4π in the definition of the area. Here too, this ambiguity is invisible since the spin S is an integer or a half-integer.

The path integral of Eq.(41) was derived first by Michael Berry [77] (and extended by Barry Simon [78]). The first term (which we called Wess-Zumino by analogy with its field theoretic versions) is called the *Berry Phase*. The role of this term, which is first order in time derivatives, is to govern the quantum dynamics of the spin which, in presence of a uniform magnetic field, executes a precessional motion of the (Bloch) sphere. It is also apparent from this expression that in the large- S limit, the path integral can be evaluated by means of a semiclassical approximation.

The coherent-state construction shows that this problem is equivalent to the path integral of a formally massless non-relativistic particle of unit electric charge on the surface of the unit sphere with a magnetic monopole of magnetic charge S in its interior! This is not surprising since the Hilbert space of a non-relativistic particle moving on the surface of a sphere with a radial magnetic field (the field of a magnetic monopole) has a Landau level type spectrum with a degeneracy given by the flux. The condition of a massless particle means that only the lowest Landau level survives and all other levels have an infinite energy gap.

The coherent state approach has been used to derive a path integral formulation for ferromagnets and antiferromagnets. A detailed derivation can be found in Ref. [9].

4.3.3. Quantum Ferromagnet

We will consider first the simpler case of a quantum ferromagnet and in Eq.(33) we will set $J = -|J| < 0$ for nearest neighbors and zero otherwise. The action for the path-integral for the spin- S quantum Heisenberg ferromagnet on a hypercubic lattice is

$$\mathcal{S} = S \sum_{\mathbf{r}} \mathcal{S}_{WZ}[\mathbf{n}(\mathbf{r}, t)] - \frac{|J|S^2}{2} \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \int_0^T dt (\mathbf{n}(\mathbf{r}, t) - \mathbf{n}(\mathbf{r}', t))^2 \quad (45)$$

where we have subtracted the classical ground state energy. The order parameter for this theory is the expectation value of the local magnetization, $\mathbf{n} = \langle \mathbf{n}(\mathbf{r}) \rangle$, which is constant in space but points in an arbitrary direction in spin space.

In the low energy regime the important configurations are slowly varying in space and we can simply approximate the action of Eq.(45) by its continuum version in d space dimensions

$$\mathcal{S} = \frac{S}{a_0^d} \int d^d x \mathcal{S}_{WZ}[\mathbf{n}(x, t)] - \frac{|J|S^2}{2a_0^d} \int d^d x \int_0^T dt (\nabla \mathbf{n}(x, t))^2 \quad (46)$$

where a_0 is the lattice spacing. As before, the path integral; is done for a field which satisfies everywhere in space-time the constraint $\mathbf{n}^2(x, t) = 1$. This action can be regarded as non-relativistic non-linear sigma model.

It is straightforward to show that the classical equations of motion for this theory are the Landau-Lifshitz equations

$$\partial_t \mathbf{n} = |J|S a_0^2 \mathbf{n} \times \nabla^2 \mathbf{n} \quad (47)$$

subject to the constraint $\mathbf{n}^2 = 1$. Due to the constraint, the Landau-Lifshitz equation is non-linear. We will decompose the field into a longitudinal and two transverse components, σ and π , respectively

$$\mathbf{n} = \begin{pmatrix} \sigma \\ \pi \end{pmatrix} \quad (48)$$

subject to the constraint $\sigma^2 + \pi^2 = 1$. The linearized Landau-Lifshitz equations become (to linear order in π)

$$\partial_t \pi_1 \simeq -|J|S a_0^2 \nabla^2 \pi_2, \quad \partial_t \pi_2 \simeq +|J|S a_0^2 \nabla^2 \pi_1 \quad (49)$$

The solution to these equations are ferromagnetic spin waves (magnons or Bloch waves) which satisfy the dispersion relation

$$\omega(\mathbf{p}) \simeq |J|S a_0^2 \mathbf{p}^2 + O(\mathbf{p}^4) \quad (50)$$

which shows that the dynamic exponent for a ferromagnet is $z = 2$. Notice that in this case the two transverse components are not independent (they are effectively a dynamical pair). These are the Goldstone bosons of a ferromagnet.

4.3.4. Quantum Antiferromagnet

Formally, the quantum antiferromagnet has a coherent state path integral whose action is

$$\mathcal{S} = S \sum_{\mathbf{r}} \mathcal{S}_{WZ}[\mathbf{n}(\mathbf{r}, t)] - \frac{JS^2}{2} \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \int_0^T dt \mathbf{n}(\mathbf{r}, t) \cdot \mathbf{n}(\mathbf{r}', t) \quad (51)$$

with $J > 0$. For a bipartite lattice, e.g. the 1D chain, and the square and cubic lattices, the classical ground state is an antiferromagnet with a Néel order parameter, the staggered magnetization. Let $\mathbf{m}(\mathbf{r})$ be the expectation value of the local magnetization. A bipartite lattice is the union of two interpenetrating sublattices, and the local magnetization is staggered, i.e. it takes values with opposite signs (with equal values) on the two sublattices. Thus, we make the change of variables, $\mathbf{n}(\mathbf{r}, t) \rightarrow (-1)^r \mathbf{n}(\mathbf{r}, t)$ in Eq.(51) and find

$$\mathcal{S} = S \sum_{\mathbf{r}} (-1)^r \mathcal{S}_{WZ}[\mathbf{n}(\mathbf{r}, t)] - \frac{JS^2}{2} \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \int_0^T dt (\mathbf{n}(\mathbf{r}, t) - \mathbf{n}(\mathbf{r}', t))^2 \quad (52)$$

We want to obtain the low energy effective action for the field $\mathbf{n}(\mathbf{r}, t)$. To this end, we decompose this field into a slowly varying part, that we will call $\mathbf{m}(\mathbf{r}, t)$, and a small rapidly varying part $\mathbf{l}(\mathbf{r}, t)$ (which represents ferromagnetic fluctuations)

$$\mathbf{n}(\mathbf{r}, t) = \mathbf{m}(\mathbf{r}, t) + (-1)^r a_0 \mathbf{l}(\mathbf{r}, t) \quad (53)$$

Since $\mathbf{n}^2(\mathbf{r}, t) = 1$, we will demand that the slowly varying part also obeys the constraint, $\mathbf{m}^2(\mathbf{r}, t) = 1$, and require that the two components be orthogonal to each other, $\mathbf{m} \cdot \mathbf{l} = 0$.

Due to the behavior of the staggered Wess-Zumino terms of Eq.(52), the resulting continuum field theory turns out to have subtle but important differences between one dimension and higher dimensions. Here we will state the results for two and higher dimensions. We will discuss in detail the one-dimensional below when we discuss the role of topology.

It turns out that if the dimension $d > 1$, the contribution of the staggered Wess-Zumino terms for *smooth* field configurations is [75, 79, 80]

$$\lim_{a_0 \rightarrow 0} S \sum_{\mathbf{r}} (-1)^r \mathcal{S}_{WZ}[\mathbf{n}(\mathbf{r}, t)] = S \int d^3x \mathbf{l}(\mathbf{x}, t) \cdot \mathbf{m}(\mathbf{x}, t) \times \partial_t \mathbf{m}(\mathbf{x}, t) \quad (54)$$

The continuum limit of the second term of Eq.(52) in the case of a two-dimensional system is

$$\lim_{a_0 \rightarrow 0} \frac{JS^2}{2} \sum_{\langle \mathbf{r}, \mathbf{r}' \rangle} \int_0^T dt (\mathbf{n}(\mathbf{r}, t) - \mathbf{n}(\mathbf{r}', t))^2 = a_0 \frac{JS^2}{2} \int d^3x ((\nabla \mathbf{m}(\mathbf{x}, t))^2 + 4\mathbf{l}^2(\mathbf{x}, t)) \quad (55)$$

The massive field $\mathbf{l}[\mathbf{x}, t]$ represents ferromagnetic fluctuations. Since this is a massive field it can be integrated-out leading to an effective field theory for the *antiferromagnetic* fluctuations $\mathbf{m}(\mathbf{x}, t)$ whose Lagrangian is that of a non-linear sigma model

$$\mathcal{L} = \frac{1}{2g} \left(\frac{1}{v_s} (\partial_t \mathbf{m}(\mathbf{x}, t) - v_s (\nabla \mathbf{m}(\mathbf{x}, t))^2) \right) \quad (56)$$

where the coupling constant is $g = 2/S$ and the spin-wave velocity is $v_s = 4a_0JS$. If we to allow for a weak next-nearest-neighbor interaction $J' > 0$, the coupling constant g and the spin wave velocity v_s become renormalized to $g' \simeq g/\sqrt{1 - 2J'/J}$ and $v'_s \simeq v_s \sqrt{1 - 2J'/J}$.

We conclude that that the quantum fluctuations about a Néel state are well described by a non-linear sigma model. Provided the frustration effects of the next-nearest-neighbor interactions are weak enough, the long-range antiferromagnetic Néel order should extend up to a critical value of the coupling constant g_c where the RG beta function has a non-trivial zero, which signals a quantum phase transition to a strong coupling phase without long-range antiferromagnetic order.

Motivated by the discovery of high temperature superconductivity in the strongly correlated quantum antiferromagnet La₂CuO₄ (at finite hole doping) in 1988 Chakravarty, Halperin and Nelson [81] utilized a quantum non-linear sigma model to analyze this system and its quantum phase transition. La₂CuO₄ is a quasi-two-dimensional material and so it exhibits strong quantum and thermal fluctuations. The upshot of this analysis is that while at $T = 0$ the non-linear sigma model has a quantum phase transition, at $T > 0$ the long range order is absent in a strictly 2D system but present in the actual material due to the weak-three-dimensional interaction. So, in the strict 2D case there is no phase transition but two different *crossover* regimes: a renormalized classical regime (without long range order), a quantum disordered regime and a quantum critical regime. La₂CuO₄ has long range Néel (antiferromagnetic) order at $T = 0$ and is in the renormalized classical regime (with long range order due to the weak 3D interaction).

The non-linear sigma model does not describe the nature of the ground state for $g > g_c$ beyond saying that there is no long range order. The problem is that, unlike the Ising model in a transverse field, the microscopic tuning parameter is the next nearest neighbor antiferromagnetic coupling J' , and to reach the regime $g \simeq g_c$ one has to make $J' \simeq J$. This is the regime in which frustration effects become strong. In this regime the assumption that the important configurations are smooth and close to the classical Néel state is incorrect. The nature of the ground state turns out to depend on the value of S .

5. Topological Excitations

Topology has come to play a crucial role both in Condensed Matter Physics and in Quantum Field Theory. Topological concepts have been used to classify topological excitations such as vortices and dislocations and to provide a mechanism for phase transitions, quantum number fractionalization, tunneling processes in field theories, and nonperturbative construction of vacuum states. Here we will discuss a few representative cases of what has become a very vast subject.

5.1. Topological Excitations: Vortices and Magnetic Monopoles

In Condensed Matter Physics topological excitations play a central role in the description of topological defects and on their role in phase transitions. Here topology integers in the classification of the configuration space into equivalence classes characterized by topological invariants [82]. The most studied example are vortices. Vortices play a key role in the mixed phase of type II superconductors in a uniform magnetic field [83]. Vortices also play a key role in the Statistical Mechanics of 2D superfluids and the the 2D classical XY model [84, 85] [86, 87] [88]. Dislocations and disclinations play an analogous role in the theory of classical melting [84, 89, 90], and 2D and 3D classical liquid crystals [91, 92, 72].

A similar problem occurs in Quantum Field Theory. Theories with global symmetries, such as the two-dimensional $O(3)$ non-linear sigma model discussed above, when formulated in Euclidean space-time have *instantons*. Typically instantons are finite Euclidean action configurations, which are also classified into equivalence classes (associated with homotopy groups) labeled by topological invariants [93, 94]. Instantons play a central role in understanding the non-perturbative structure of gauge theories. Gauge theories with a compact gauge group coupled to matter fields have non-trivial vortex [95] and monopole [96, 97, 98] configurations, as do non-abelian Yang-Mills gauge theories [99]. Instantons have also played a central role in Condensed Matter Physics as well, notably in Haldane's work on 1D quantum antiferromagnets (discussed below), and in the problem of macroscopic quantum tunneling and coherence [100, 101].

5.1.1. Vortices in two dimensions

In this section I will focus on the the problem of the superfluid transition in 2D and the closely related problem of the phase transition of a magnet with an easy-plane anisotropy, the classical XY model. A superfluid is described by an order parameter that is a one-component complex field $\phi(\mathbf{x})$. If electromagnetic fluctuations are ignored, this description also applies to a superconductor. The complex field can be written in terms of an amplitude $|\phi(\mathbf{x})|$, whose square represents the local superfluid density, and a phase $\theta(\mathbf{x}) = \arg(\phi(\mathbf{x}))$. Deep in the superfluid phase the amplitude is essentially constant, that we will set to be a real positive number ϕ_0 , while the phase field $\theta(\mathbf{x})$ is periodic with period 2π and can fluctuate. Similarly, an easy-plane ferromagnet is described by a two-component real order parameter field $\mathbf{M}(\mathbf{x}) = (M_1(\mathbf{x}), M_2(\mathbf{x})) = |\mathbf{M}(\mathbf{x})|(\cos \theta(\mathbf{x}), \sin \theta(\mathbf{x}))$. Deep in the ferromagnetic phase the amplitude $|\mathbf{M}|$ is essentially constant but the phase field $\theta(\mathbf{x})$ can fluctuate.

We will assume that we are in a regime where the local superfluid density $|\phi_0|^2$ is well formed (or, equivalently that $|\mathbf{M}|$ is locally well formed) but that the phase field is fluctuating. In this regime the problem at hand is an $O(2) \simeq U(1)$ non-linear sigma model, and its partition function takes the form

$$Z = \int \mathcal{D}\theta \exp\left(- \int d^2x \frac{1}{2g} (\partial_\mu \theta(\mathbf{x}))^2\right) \quad (57)$$

where we defined the coupling constant $g = T/J|\phi_0|^2$, where T is the temperature, J is an interaction strength, and $|\phi_0|^2$ is the magnitude (squared) of the amplitude of the order parameter, which we will take to be constant; $\kappa = J|\phi_0|^2$ is the phase stiffness.

Except for the requirement that the phase field be *locally* periodic, $\theta \simeq \theta + 2\pi$, superficially this seems to be a trivial free (Gaussian) field theory. We will see that the periodicity (or, *compactification*) condition makes this theory non-trivial. Indeed, configurations of the phase field that are weak enough that they do not see the periodicity condition, for all practical purposes, can be regarded as being non-compact and ranging from $-\infty$ to $+\infty$. However there are many configurations for which the periodicity condition is essential. Such configurations are called *vortices*.

Even in the absence of vortices, the periodic (compact) nature of the phase field is essential to the physics of this problem. In fact the only allowed observables must be invariant under local periodic shifts of the phase field. This implies that the phase field θ itself is not a physical observable but that exponentials of the phase of the form $\exp(in\theta(\mathbf{x}))$ are physical. This operator is just the order parameter field of the XY model. In Conformal Field theory operators of this type are called vertex operators [102, 103]. We will see below that this theory has a dual field ϑ , associated with vortices, and that there are vertex operators of the dual field. In String Theory the model of a compactified scalar is known as the compactified boson and represents the coordinate of a string on a compactified space, in this case a circle S_1 [104].

To picture a vortex consider a large closed curve C on the 2D plane. Hence, topologically a closed curve is isomorphic to a circle, $C \simeq S_1$. The phase field $\theta(\mathbf{x})$ is equivalent to a unit

circle S_1 . Therefore the configuration space are maps of S_1 (the large circle) onto S_1 (the unit circle of the order parameter space). The configurations can be classified by the number of times the phase winds on the large circle C . The winding number is an integer called the topological charge of the configuration, the vorticity. Thus, a vortex is a configuration of the phase field $\theta(\mathbf{x})$ that winds by $2\pi m$ (where m is an integer):

$$\frac{(\Delta\theta)_C}{2\pi} = \frac{1}{2\pi} \oint_C d\mathbf{x} \cdot \nabla\theta(\mathbf{x}) \equiv i \oint_0^{2\pi} \frac{d\varphi}{2\pi} e^{i\theta(\varphi)} \partial_\varphi e^{-i\theta(\varphi)} = m \quad (58)$$

where $\varphi \in (0, 2\pi]$ is the azimuthal angle for a vector at the center of the large circle C . Here n is the vorticity or *winding number* of the configuration; $m > 0$ is a vortex and $m < 0$ is an anti-vortex. The vorticity is a *topological invariant* of the field the configuration $\theta(\mathbf{x})$ which does not change under smooth changes.

The winding number of a vortex is a topological invariant that classifies the configurations of the phase field as continuous maps of a large circle S_1 onto the unit circle S_1 defined by the phase field. In Topology such continuous maps are called homotopies. The winding number classifies these maps into a discrete set of equivalence classes, which form a *homotopy group* under the composition of two configurations. In this case the homotopy group is called $\Pi_1(S_1)$. Since the equivalence classes are classified by a topological invariant that takes integer values, the homotopy group $\Pi_1(S_1)$ is isomorphic to the group of integers, \mathbb{Z} [82].

The field $j_\mu(\mathbf{x}) = \partial_\mu\theta(\mathbf{x})$ is the superfluid current, and the *vorticity* $\omega(\mathbf{x})$ is the curl of the current, i.e.

$$\omega(\mathbf{x}) = \epsilon_{\mu\nu}\partial_\mu j_\nu(\mathbf{x}) = \epsilon_{\mu\nu}\partial_\mu\partial_\nu\theta(\mathbf{x}) \quad (59)$$

which vanishes unless $\theta(\mathbf{x})$ has a branch-cut singularity across which the phase field jumps by $2\pi n$. Let $\omega(\mathbf{x})$ be the vorticity field with singularities at the locations $\{\mathbf{x}_j\}$ of vortices with topological charge m_j

$$\omega(\mathbf{x}) = 2\pi \sum_j m_j \delta^2(\mathbf{x} - \mathbf{x}_j) \quad (60)$$

which is satisfied by the phase field configuration

$$\theta(\mathbf{x}) = \sum_j 2\pi m_j \text{Im} \ln(z - z_j) \quad (61)$$

where we have used the complex coordinates $z = x_1 + ix_2$. Away from the singularities $\{\mathbf{x}_j\}$, this configuration obeys the Laplace equation. Hence, it has a Cauchy-Riemann dual field $\vartheta(\mathbf{x})$ which satisfies the Cauchy-Riemann equation

$$\partial_\mu\vartheta = \epsilon_{\mu\nu}\partial_\nu\theta \quad (62)$$

which satisfies the Poisson equation

$$-\nabla^2\vartheta(\mathbf{x}) = \omega(\mathbf{x}) \quad (63)$$

whose solution is

$$\vartheta(\mathbf{x}) = \int d^2y G(|\mathbf{x} - \mathbf{y}|) \omega(\mathbf{y}) \quad (64)$$

where $G(|\mathbf{x} - \mathbf{y}|)$ is the Green function of the 2D Laplacian

$$-\nabla^2 G(|\mathbf{x} - \mathbf{y}|) = \delta^2(\mathbf{x} - \mathbf{y}) \quad (65)$$

In 2D this Green function is

$$G(|\mathbf{x} - \mathbf{y}|) = \frac{1}{2\pi} \ln\left(\frac{a}{|\mathbf{x} - \mathbf{y}|}\right) \quad (66)$$

where a is a short distance cutoff (a lattice spacing). In what follows we will assume that the Green function of Eq.(66) has been cutoff so that $G(|\mathbf{x} - \mathbf{y}|) = 0$ for $|\mathbf{x} - \mathbf{y}| \leq a$.

The energy of a configuration of vortices $\{n_j\}$ with vanishing total vorticity, $\sum_j m_j = 0$, is

$$\begin{aligned} E[\theta] &= \frac{J\phi_0^2}{2} \int d^2x (\partial_\mu\theta)^2 \\ &= \frac{J\phi_0^2}{2} \int d^2x \int d^2y \omega(\mathbf{x}) G(|\mathbf{x} - \mathbf{y}|) \omega(\mathbf{y}) = 2\pi J\phi_0^2 \sum_{j>k} m_j m_k \ln\left(\frac{a}{|\mathbf{x}_j - \mathbf{x}_k|}\right) \end{aligned} \quad (67)$$

where we used that configurations with non-vanishing vorticity do not contribute to the partition function since they have infinite energy in the thermodynamic limit. We conclude that, up to an unimportant prefactor, that the partition function of Eq.(57) is the same as the partition function a gas of charges $\{m_j\}$ (the vortices) with total vanishing vorticity, $\sum_j m_j = 0$,

$$Z_{2DCG} = \sum_{\{\{m_j\}\}} \exp \left(-2\pi \frac{J\phi_0^2}{T} \sum_{j < k} m_j m_k \ln \left(\frac{|\mathbf{x}_j - \mathbf{x}_k|}{a} \right) \right) \quad (68)$$

which is known as the two-dimensional (neutral) Coulomb gas [84].

Kosterlitz and Thouless argued that the neutral two dimensional Coulomb gas has a phase transition a critical temperature T_{KT} between a low temperature phase dielectric phase in which vortices and anti-vortices are bound into neutral pairs and a high temperature phase in which pairs unbind, the vortex charges are screened and the vortex gas is in a plasma phase. A simple estimate of the critical temperature is found by computing the free energy of a single vortex $F_{\text{vortex}} = E_{\text{vortex}} - TS_{\text{vortex}}$, where E_{vortex} of a single vortex and S_{vortex} is the vortex entropy. In 2D the energy of a single vortex of unit topological charge, $n = 1$, is $\pi J\phi_0^2 \ln(L/a)$, where L is the linear size of the system, and the vortex entropy is $S_{\text{vortex}} = \ln(L/a)^2$, where $(L/a)^2$ is the number of places where a vortex can be located, i.e. the area of the system in units of the spacing a . Vortices unbind (proliferate) when the entropy beats the energy. Hence, the critical point occurs at a temperature T_{KT} where the vortex free energy vanishes, $F_{\text{vortex}}[T_{KT}] = 0$. This yields the estimate

$$T_{KT} = \frac{\pi}{2} J\phi_0^2 \quad (69)$$

which is the Kosterlitz-Thouless critical temperature.

There is an alternative, equivalent way to see this physics. Let us consider the theory if Eq.(57) in the background of a U(1) gauge field $A_\mu(x)$

$$Z[A] = \int \mathcal{D}\theta \exp \left(-\frac{1}{2g} \int d^2x (\partial_\mu \theta - A_\mu)^2 \right) \quad (70)$$

where the curl of the background gauge field A_μ is chosen to be equal to the local vorticity of Eq.(59),

$$\epsilon_{\mu\nu} \partial_\mu A_\nu(x) = \omega(x) \quad (71)$$

In other words, the gauge A_μ field forces the phase field to have vortices at the locations $\{x_j\}$.

Next we will use a Hubbard-Stratonovich transformation (a Gaussian identity) to write Eq.(70) in the equivalent form

$$Z[A] = \int \mathcal{D}\theta \int \mathcal{D}a_\mu \exp \left(-\frac{g}{2} \int d^2x a_\mu^2 + i \int d^2x a_\mu(x) (\partial_\mu \theta(x) - A_\mu(x)) \right) \quad (72)$$

where we neglected an unimportant prefactor. In Eq.(72) the phase field can be integrated-out (after an integration by parts) to yield the constraint $\partial_\mu a_\mu = 0$. This constraint is solved by introducing the dual field $\vartheta(x)$ such that

$$a_\mu(x) = \epsilon_{\mu\nu} \partial_\nu \vartheta(x) \quad (73)$$

Hence, the partition function $Z[A]$ becomes

$$Z[A] = \int \mathcal{D}\vartheta \exp \left(-\frac{g}{2} \int d^2x (\partial_\mu \vartheta)^2 + i \int d^2x \vartheta(x) \omega(x) \right) \quad (74)$$

where we have assumed that the total vorticity is zero, $\int d^2x \omega(x) = 2\pi \sum_j m_j = 0$. This identity shows that the quantization of the vortex topological charge requires that the dual field ϑ should also be periodic (compact) with period 1.

It is elementary to see that after performing the gaussian integral over the field ϑ in Eq.(74) and summing over all vortex configurations (with total vanishing vorticity) we reproduce the partition function of the 2D neutral Coulomb gas [105, 106]. Thus, the theory of Eq.(70) defined in terms of the phase field θ and the theory of Eq.(74) are equivalent. This is an example of a *duality transformation* which amounts to the exchange of the fields

$$\theta \leftrightarrow \vartheta, \quad \text{and} \quad g \leftrightarrow 1/g \quad (75)$$

A more careful analysis done on the lattice shows that the dual field ϑ is defined on the dual of the square lattice (which is also a square lattice) [88, 105]. In addition, by differentiating both the partition function of the phase field of Eq.(70) and the partition function of the dual field ϑ of Eq.(74) with respect to the background field A_μ we obtain the duality identification of the currents

$$\frac{1}{g} \partial_\mu \theta \longleftrightarrow g \epsilon_{\mu\nu} \partial_\nu \vartheta \quad (76)$$

as an operator identity. This identity has the same form as the duality of forms in geometry. In this case, the dual of the current $\partial_\mu \theta$, which is a 1-form field, is another 1-form field, $\partial_\mu \vartheta$. In both cases the field θ and its dual ϑ are associated with global symmetries. Later we will encounter similar duality identifications but their content will be different in different dimensions. Moreover, running these identities backwards it is easy to see that the insertion of the order parameter operator $\exp(ip\theta(x))$ (with p and integer), which acts as a *source* on the phase field θ , in the partition function is equivalent to minimally couple the dual field ϑ to a singular gauge field $B_\mu(x)$ whose curl is p at the location x . In other words, *charges are dual to vortices*.

We will now use the identity we derived in Eq.(74) to compute the full partition function which is a sum over all vortex configurations. Now we will assume that the vortices have a large core energy (i.e. the energy associated with the short distance behavior of the Green function of Eq.(66)). For a vortex with unit topological charge we will call this energy u , and for a vortex with charge m the core energy will be um^2 . The effects of the core energy can be accounted for in terms of a vortex fugacity $z = \exp(-un^2/T)$. The total partition function now is

$$Z = \sum_{\{m_j\}} \int \mathcal{D}\vartheta \exp \left(-\frac{g}{2} \int d^2x (\partial_\mu \vartheta)^2 + i \sum_j 2\pi m_j \vartheta(x_j) - \frac{u}{T} \sum_j m_j^2 \right) \quad (77)$$

where, once again, we assumed global charge neutrality.

In the limit in which $z = \exp(-u/T) \ll 1$, only the vortices with the lowest topological charges $m = \pm 1$ contribute and, furthermore, they are dilute. In this limit we can perform over the configurations with the lowest topological charge and derive the effective field theory of the dual field ϑ :

$$Z_{\text{SG}} = \int \mathcal{D}\vartheta \exp \left(- \int d^2x \left(\frac{g}{2} (\partial_\mu \vartheta)^2 - v \cos(2\pi\vartheta) \right) \right) \quad (78)$$

which is known as the sine-Gordon field theory. The coupling constant of the sine-Gordon theory is $v = 2 \exp(-u/T)/a^2$, where a is the core size (the lattice spacing). Hence, the cosine operator of the sine-Gordon theory represents the effects of the vortices.

We can now analyze the role of vortices by looking at the renormalization group of the sine-Gordon theory. The first question is what is the scaling dimension of the cosine operator. This can be computed by looking at the vortex operator correlator in the theory with $v = 0$:

$$\langle \exp(i2\pi\vartheta(x)) \exp(-i2\pi\vartheta(y)) \rangle = \frac{\text{const.}}{|x-y|^{2\pi/g}} \quad (79)$$

which implies that the vortex scaling dimension is

$$\Delta_{\text{vortex}} = \frac{\pi}{g} \quad (80)$$

Eq.(12) implies that in $D = 2$ dimensions the operator is irrelevant if $\Delta > 2$, relevant for $\Delta < 2$ and marginal for $\Delta = 2$. Hence, vortices are marginal if $\Delta_{\text{vortex}} = 2$ which happens if $g = \pi/2$. This is the same as to say that the system is at the Kosterlitz-Thouless critical temperature $T = T_{KT}$. Hence, vortices are irrelevant if $T < T_{KT}$ and relevant for $T > T_{KT}$.

This RG analysis tells us that $T > T_{KT}$, when vortices proliferate, the coupling constant v flows to strong coupling to a regime where ϑ is pinned to an integer value and the theory is in a massive phase. In this phase the connected vortex correlator decays exponentially with distance which is the same as to say that the vortex charge is screened. This is why in this phase the vortices proliferate. It can be shown that in this phase the correlator of the spins of the XY model, i.e. the correlator of the vertex operator $\exp(i\theta(x))$, decays exponentially with distance and the systems is in its disordered phase.

There is still the question of the nature of the phase with $T < T_{KT}$. The Mermin-Wagner Theorem [107] (and its generalizations by Hohenberg [108] and Coleman [109]) states that classical

statistical mechanical systems with a global continuous symmetry group cannot undergo spontaneous symmetry breaking in space dimensions $D \leq 2$ (and quantum systems with space-time dimensions $D \leq 2$). Does this theory violate the Mermin-Wagner Theorem? The answer is no. It is easy to see that in the phase in which the vortices are irrelevant, i.e. for $T < T_{KT}$, the correlator of the order parameter operator is always a power law of the distance $|x - y|$,

$$\langle \exp(i\theta(x)) \exp(-i\theta(y)) \rangle = \frac{\text{const.}}{|x - y|^{g/2\pi}} \quad (81)$$

with an exponent that depends on temperature and satisfies

$$\frac{g}{2\pi} = \frac{T}{2\pi J\phi_0^2} \leq \frac{1}{4} \quad (82)$$

In other words, the entire low temperature phase is not an ordered phase of matter since the correlator is not constant at long distance. On the other hand, the correlators of the order parameter $\exp(i\theta)$ and of the vortices $\exp(i\theta)$ have a power law behavior for $T < T_{KT}$, we conclude that in this temperature range the system is scale invariant and that it has a line of critical points.

In summary, we succeeded in expressing the partition function as a sum over configurations of the topological excitations, the vortices. We succeeded in doing that because vortices are labeled by their coordinates on the plane. In addition, we found a non-trivial phase transition since the entropy and the energy both scale logarithmically with the linear size of the system or, equivalently, that vortices became marginally relevant at a critical temperature. This is the mechanism behind the Kosterlitz-Thouless transition.

One may ask if this construction is generic and the answer is no. As an example consider the Abelian Higgs model in $D = 2$. This model has a complex scalar field minimally coupled to a Maxwell gauge field and, hence, its gauge group is U(1). In the classical spontaneously broken phase, the gauge field becomes massive. In this phase the long range coherence of the phase field of the vortices of the scalar field is screened at the scale of the penetration depth. As a result the Euclidean action of the vortices is now finite. Furthermore, on longer scales the interaction energy between vortices becomes short ranged. Hence, instead of a 2D Coulomb gas now one has a gas of particles (vortices and anti-vortices) with short range interactions. In this case the entropy always dominates and the vortices proliferate [110]. This behavior is quite analogous to the restoration of symmetry by proliferation of domain walls in one-dimensional classical spin chains [19] and to the analogous problem of tunneling in the path integral formulation of quantum mechanical double-well potentials [93]. As a caveat, we should note that, in spite of the obvious similarities, the 2D Abelian Higgs model does not describe correctly a 2D superconductor (i.e. a superconducting film) since the electromagnetic field is not confined to the film and this renders the electromagnetic action nonlocal.

5.1.2. Magnetic monopoles in compact electrodynamics

Instantons are of great interest in Quantum Field Theory since they provide for a mechanism to understand the non-perturbative structure of these theories. For this reasons they have been used to understand the mechanisms of quark confinement and the role of quantum anomalies in non-Abelian gauge theories [99, 97, 98, 96, 111]. Instantons in non-Abelian gauge theories are magnetic monopoles and the condensation of monopoles have long been argued to be the mechanism behind quark confinement.

The simplest non-Abelian gauge theory that has magnetic monopoles is the Georgi-Glashow [112]. This model has a three-component real field ϕ and an SU(2) Yang-Mills gauge field A_μ . In its Higgs phase the scalar field ϕ acquires an expectation value which breaks the gauge symmetry group SU(2) down to its diagonal U(1) subgroup. Since $U(1) \subset SU(2)$, this Abelian gauge group is compact, meaning that its magnetic fluxes are quantized. Polyakov [113] and 't Hooft [111] showed that in 2+1 Euclidean dimensions have non-singular instanton solutions which at long distances resemble the magnetic monopole originally proposed by Dirac in 1931 [114]

$$B_i(\mathbf{x}) = \frac{q}{2} \frac{x_i}{|\mathbf{x}|^2} - 2\pi q \delta_{i,3} \delta(x_1) \delta(x_2) \theta(-x_3) \quad (83)$$

The first term in Eq.(83) is the magnetic radial field of a monopole of magnetic charge q . The second term represents an infinitely long infinitesimally thin solenoid ending at the location of the monopole, $\mathbf{x} = \mathbf{0}$, that supplies the quantized magnetic flux $2\pi q$. This singular term is known as the Dirac string. The string itself (and its orientation) is physically unobservable to any

electrically charged particle that obeys the Dirac quantization condition, $qe = 2\pi$ (in units where $\hbar = c = 1$).

In the language of a lattice gauge theory [32], a theory with a compact (i.e. periodic) U(1) gauge fields on a $D = 3$ cubic lattice, describing a compact gauge field in 2 + 1 dimensions. This theory should have instantons that resemble magnetic monopoles much in the same way as a theory with a compact global U(1) symmetry has vortices. The simplest example is Polyakov's compact electrodynamics [97] whose partition function is

$$Z = \prod_{x,\mu} \int_0^{2\pi} \frac{dA_\mu(x)}{2\pi} \exp\left(\frac{1}{4e^2} \sum_{x,\mu,\nu} \cos(F_{\mu\nu}(x))\right) \quad (84)$$

where $F_{\mu\nu}(x) = \Delta_\mu A_\nu(x) - \Delta_\nu A_\mu(x) \equiv \sum_\mu A_\mu$ is the magnetic flux through the elementary plaquette labeled by a site x and a pair of directions, μ and ν , with $\mu = 1, 2, 3$. This theory is invariant under local gauge transformations $A_\mu(x) \rightarrow A_\mu(x) + \Delta_\mu \Phi(x)$ and it is also invariant under local periodic shifts of the gauge fields $A_\mu(x) \rightarrow A_\mu(x) + 2\pi\ell_\mu(x)$, where $\ell_\mu(x) \in \mathbb{Z}$. The plaquette flux operator satisfies the lattice version of the Bianchi identity that the product of exponentials of the flux on the faces of every elementary cube of the lattice is

$$\prod_{\text{cubefaces}} e^{iF_{\mu\nu}(x)} = 1 \quad (85)$$

which says that the theory can have magnetic monopoles of integer magnetic charge.

We will analyze this theory following an approach analogous to what we used for vortices in section 5.1.1. To this end we will consider the partition function

$$Z[B_{\mu\nu}] = \int \mathcal{D}A_\mu \exp\left(-\frac{1}{4e^2} \int d^3x (F_{\mu\nu}(x) - B_{\mu\nu}(x))^2\right) \quad (86)$$

where $F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength of the abelian U(1) Maxwell gauge field A_μ . Here $B_{\mu\nu}(x)$ is an (anti-symmetric) two-form background gauge field. The coupling constant of this theory is e^2 . Since A_μ is a connection it has units of length $^{-1}$, and $F_{\mu\nu}^2$ is a dimension 4 field. Then, in $D = 3$ dimensions, e^2 has units of length $^{-1}$.

The theory is invariant under two local transformations, namely the usual invariance under gauge transformations

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \Phi(x), \quad B_{\mu\nu}(x) \rightarrow B_{\mu\nu}(x) \quad (87)$$

where $\Phi(x)$ is an arbitrary smooth function of x . The presence of the background two-form field $B_{\mu\nu}$ now requires invariance under one-form gauge transformations

$$A_\mu(x) \rightarrow A_\mu(x) + a_\mu(x), \quad B_{\mu\nu}(x) \rightarrow B_{\mu\nu}(x) + \partial_\mu a_\nu - \partial_\nu a_\mu \quad (88)$$

The two-form gauge field $B_{\mu\nu}$ essentially represents the magnetic monopoles. Let $\{m_j\}$ be a configuration of monopoles of charges m_j with coordinates $\{x_j\}$, with total vanishing monopole charge, $\sum_j m_j = 0$. Let $\mathcal{M}(x)$ be the magnetic monopole density at x ,

$$\mathcal{M}(x) = 2\pi \sum_j m_j \delta^3(x - x_j) \quad (89)$$

which can be expressed as the curl of the two-form gauge field $B_{\mu\nu}$,

$$\mathcal{M}(x) = \frac{1}{2} \epsilon_{\mu\nu\lambda} \partial_\mu B_{\nu\lambda}(x) \quad (90)$$

We will proceed next much in the same way as in Eq.(72) and rewrite the partition function of Eq.(86) in terms of a two-form Hubbard-Stratonovich field $b_{\mu\nu}(x)$ such that

$$\begin{aligned} Z[B] &= \int \mathcal{D}A_\mu \int \mathcal{D}b_{\mu\nu} \exp\left(-\frac{e^2}{4} \int d^3x b_{\mu\nu}^2(x) + i \int d^3x \frac{1}{2} b_{\mu\nu}(x) [F_{\mu\nu}(x) - B_{\mu\nu}(x)]\right) \\ &= \int \mathcal{D}A_\mu \int \mathcal{D}b_{\mu\nu} \exp\left(-\frac{e^2}{4} \int d^3x b_{\mu\nu}^2(x) + i \int d^3x \left[A_\mu(x) \partial_\nu b_{\mu\nu}(x) - \frac{1}{2} b_{\mu\nu}(x) B_{\mu\nu}(x)\right]\right) \end{aligned} \quad (91)$$

Thus, the gauge field A_μ plays the role of a Lagrange multiplier field the enforces the constraint

$$\partial_\nu b_{\mu\nu}(x) = 0 \quad (92)$$

which is solved in terms of a compact scalar field $\vartheta(x)$

$$b_{\mu\nu}(x) = \epsilon_{\mu\nu\lambda} \partial_\lambda \vartheta(x) \quad (93)$$

Using this identity and the definition of the monopole density $\mathcal{M}(x)$ we find that the partition function $Z[B_{\mu\nu}]$ of Eq.(86) becomes

$$\begin{aligned} Z[B] &= \int \mathcal{D}\vartheta \exp \left[- \int d^3x \left(\frac{e^2}{2} (\partial_\mu \vartheta(x))^2 + i\mathcal{M}(x)\vartheta(x) \right) \right] \\ &= \int \mathcal{D}\vartheta \exp \left[- \frac{e^2}{2} \int d^3x (\partial_\mu \vartheta(x))^2 + 2\pi i \sum_j m_j \vartheta(x_j) \right] \end{aligned} \quad (94)$$

which requires that the field ϑ obeys the compactification condition $\vartheta \rightarrow \vartheta + n$, where n is an arbitrary integer. Eq.(94) says that the magnetic monopole instantons of the compact U(1) gauge theory are dual to charges of the dual phase field ϑ , which has a compact U(1) global symmetry.

The full partition function is obtained by summing over all monopole configurations satisfying the total neutrality condition, $\sum_j m_j = 0$. As in section section 5.1.1, we will weigh the configurations with a coupling u and find

$$Z = \sum_{\{m_j\}} Z\{m_j\} \int \mathcal{D}\vartheta \exp \left(- \frac{e^2}{2} \int d^3x (\partial_\mu \vartheta)^2 + \sum_j 2\pi i m_j \vartheta(x_j) - u \sum_j m_j^2 \right) \quad (95)$$

which is the same theory we found in Eq.(77) except that now we are in 3D. Moreover, summing only over dilute configurations of monopoles and anti-monopoles we find, once again the sine-Gordon theory but now in $D = 3$ dimensions:

$$Z = \int \mathcal{D}\vartheta \exp \left(- \int d^3x \left(\frac{e^2}{2} (\partial_\mu \vartheta)^2 - v \cos(2\pi \vartheta) \right) \right) \quad (96)$$

with $v = 2 \exp(-u)/a^3$.

In spite of the similarities between Eq.(96) and the sine-Gordon theory in 2D, Eq.(78), the physics is very different. It is straightforward to see that, in the limit $v = 0$, the monopole operator correlator is

$$\langle \exp(2\pi i \vartheta(x)) \exp(-2\pi i \vartheta(y)) \rangle = \exp \left(\frac{4\pi^2}{e^2} [G(|x-y|) - G(0)] \right) \simeq \exp \left(\frac{\pi}{2e^2} \left[\frac{1}{R} - \frac{1}{a} \right] \right) \quad (97)$$

where a is the short-distance cutoff. Unlike the behavior of the correlator of the vortex operators in 2D found in Eq.(79), Eq.(97) does not show a power-law behavior. The reason is that at $v = 0$ the compactified field ϑ should be regarded as the Goldstone boson of a spontaneously broken U(1) symmetry. However, the cosine operator is now always relevant and the field ϑ is pinned and its fluctuations are actually massive.

Looking back at the partition function of Eq.(95), we could integrate out the field ϑ and obtain an expression with the same form as the Coulomb gas of Eq.(68) except that now this is the three-dimensional neutral Coulomb gas [97]

$$Z_{3DCG} = \sum_{\{m_j\}} \exp \left(- \frac{\pi}{2e^2} \sum_{j < k} m_j m_k \left[\frac{1}{|\mathbf{x}_j - \mathbf{x}_k|} - \frac{1}{a} \right] \right) \quad (98)$$

Now the energy of an isolated monopole is finite (in the infrared) while the entropy is still logarithmically divergent. The conclusion is that the monopoles proliferate and the 3D neutral Coulomb gas is always in a plasma phase with screened interactions. These results also imply that the expectation value of the Wilson loop operator in the U(1) gauge theory decays exponentially with the *area* of the loop. The conclusion is that in 2+1 dimensions compact electrodynamics is in a confining phase for all values of the coupling constant e^2 [97, 98].

5.2. Non-Linear Sigma Models and Antiferromagnetic Quantum Spin Chains

We will now discuss the one-dimensional case in which topology plays a crucial role. In one space dimension, the sites of the chain are labelled by an integer $j = 1, \dots, N$ (with N even). After some simple manipulations, the continuum limit of the staggered Wess-Zumino terms of Eq.(52) is

$$\lim_{a_0 \rightarrow 0} S \sum_{j=1}^N (-1)^j S_{WZ}[\mathbf{n}(j, t)] = \frac{S}{2} \int d^2x \mathbf{m}(x, t) \cdot \partial_t \mathbf{m}(x, t) \times \partial_x \mathbf{m}(x, t) + S \int d^2x \mathbf{l}(x, t) \cdot \mathbf{m}(x, t) \times \partial_t \mathbf{m}(x, t) \quad (99)$$

The continuum limit of the second term of Eq.(52) is essentially the same as in Eq.(55).

By putting together the results of Eq.(99) and Eq.(55), we find that in the case of a 1D chain the continuum field theory has the Lagrangian density

$$\mathcal{L}[\mathbf{m}, \mathbf{l}] = -2a_0JS^2\mathbf{l}^2 + S\mathbf{l} \cdot \mathbf{m} \times \partial_t \mathbf{m} - \frac{a_0JS^2}{2}(\partial_x \mathbf{m})^2 + \frac{S}{2}\mathbf{m} \cdot \partial_t \mathbf{m} \times \partial_x \mathbf{m} \quad (100)$$

Finally, we integrate-out the ferromagnetic fluctuations represented by the massive field $\mathbf{l}(x, t)$ to find that the effective low-energy Lagrangian has the form

$$\mathcal{L} = \frac{1}{2g} \left(\frac{1}{v_s} (\partial_t \mathbf{m})^2 - v_s (\partial_x \mathbf{m})^2 \right) + \frac{\theta}{4\pi} \mathbf{m} \cdot \partial_x \mathbf{m} \times \partial_t \mathbf{m} \quad (101)$$

which is a non-linear sigma model with $N = 3$ components (and hence has a global $O(3)$ symmetry) with a *topological term* whose coupling constant is θ . This result was first derived by Haldane [115].

In Eq.(101) the effective coupling constant g , the spin-wave velocity v_s , and the θ -angle are given by $g = \frac{2}{S}$, $v_s = 2a_0JS$, and $\theta = 2\pi$. As in the 2D case, since $g = 2/S$, large S implies that the non-linear sigma model is in the weak coupling regime. Provided that $v_s > 0$, one can always rescale the time and space coordinates without affecting the form of the Lagrangian, including the coupling constant or, as we will see, the value of the θ angle. In what follows we will assume that we have done the rescaling in such a way that we set $v_s = 1$ and, that time and space scale as lengths. Thus we will use a relativistic notation and, after an analytic continuation to imaginary time, we label the time coordinate by x_2 and the space direction by x_1 . The partition function now takes the form

$$Z = \int \mathcal{D}\mathbf{m} \exp \left(-\frac{1}{2g} \int_{\Omega} d^2x (\partial_\mu \mathbf{m})^2 + i\theta Q[\mathbf{m}] \right) \quad (102)$$

where Ω is the spacetime manifold. In this notation the first term of the exponent is called the Euclidean action of the non-linear sigma model. In Eq.(102) we denoted by $Q[\mathbf{m}]$ the quantity

$$Q[\mathbf{m}] = \frac{1}{8\pi} \int_{\Omega} d^2x \epsilon_{\mu\nu} \mathbf{m} \cdot \partial_\mu \mathbf{m} \times \partial_\nu \mathbf{m} \quad (103)$$

Here we will consider the case in which the spacetime manifold Ω is closed. In particular we will assume that it is a two-sphere S_2 . The quantity $Q[\mathbf{m}]$ is the integral of a total derivative which counts the number of times the field configuration $\mathbf{m}(x_1, x_2)$ wraps around the sphere S_2 . In other words, it yields a non-vanishing result only for “large” configurations which wind (or wrap) around the sphere S_2 . Since $Q[\mathbf{m}]$ is an integer, it has the same for all *smooth* the field configuration \mathbf{m} that can be smoothly deformed into each other and are homotopically equivalent. If we demand that the field configurations \mathbf{m} have finite Euclidean action, which requires that at infinity the configurations take the same (but arbitrary) value of \mathbf{m} , we have effectively compactified the $x_1 - x_2$ plane into a two sphere S_2 . On the other hand the field \mathbf{m} is restricted by the constraint $\mathbf{m}^2 = 1$ to take values on a two-sphere S_2 . Therefore, the field configurations $\mathbf{m}(x_1, x_2)$ are smooth maps of the S_2 of the coordinate space to the S_2 of the target space of the field \mathbf{m} . Hence, the integer $Q[\mathbf{m}]$ classifies the smooth field configurations into a set of *equivalence classes* each labeled by the integer Q . Under composition homotopies form groups, and the equivalence classes themselves also form a group which, in this case, is isomorphic to the group of integers, \mathbb{Z} . In Topology, these statements are summarized by the notation $\Pi_2(S_2) \simeq \mathbb{Z}$. The

configurations with non-zero values of Q are called *instantons* which, in the quantum problem, represent tunneling processes of the non-linear sigma model.

Another consequence of Q being an integer is that the contribution of its term to the weight of the path integral of Eq.(103) is a periodic function of θ angle. On the other hand, since the only allowed values of the θ are $\theta = 0 \pmod{2\pi}$ for S integer, or $\theta = \pi \pmod{2\pi}$ for S a half-integer, the contribution of the topological invariant Q to the weight of the path integral is

$$\exp(i\theta Q[\mathbf{m}]) = (-1)^{2SQ} \quad (104)$$

Therefore, for spin chains with S integer, the weight is 1 and the topological invariant does not contribute to the path integral. But, if S is a half-integer, the weight is $(-1)^{Q[\mathbf{m}]}$, and it does contribute. Moreover, its contribution is the same for all half-integer values of S .

These results have important consequences for the physics of spin chains which led Haldane to some startling conclusions [115]. In the weak coupling regime, $g \ll 1$ (equivalently, for large S), we can use the perturbative renormalization group and derive the beta function for all these $O(3)$ non-linear sigma models (with or without topological terms) and find that their beta functions are the same as in Eq.(21) with $N = 3$. Hence, for all S , the effective coupling flows to large values. We can make this inference since the topological term yields no contribution for all configurations which are related by smooth deformations. Thus, we infer that all spin chains with S integer are in a massive phase with an exponentially small energy gap $\sim \exp(-2\pi S)$. This result is nowadays known as the Haldane gap.

On the other hand, these results also imply that all spin chain with half-integer spin S are also the same and, in particular, the same as spin-1/2 chains. However, the Hamiltonian of the quantum spin chain with spin-1/2 degrees of freedom is an example of an integrable system and its spectrum is known to be gapless from its Bethe Ansatz solution [116, 117]. However, the spin-1/2 chain is not only gapless but its low energy states are gapless solitons with a relativistic spectrum. The low energy description of a theory of this type must be described by a conformal field theory. Haldane concluded that the RG flow for spin-1/2 chains must have an IR stable fixed point at some finite (and large) value of the coupling constant. This conjecture was confirmed by Affleck and Haldane [118] who showed that the spin-1/2 Heisenberg chains are in the universality class of the $SU(2)_1$ Wess-Zumino-Witten model [119] whose CFT was solved by Knizhnik and Zamolodchikov [120].

5.3. Topology and open integer-spin chains

In the preceding section we showed that integer spin chains have a Haldane gap. We did that by showing that in that case the topological term is absent. However, the derivation is correct provided the spacetime manifold is closed, e.g. a sphere, a torus, etc. What happens if the system has a boundary? Let us denote the boundary of Ω by $\Gamma = \partial\Omega$. For example we will take Γ to be along the imaginary time direction and hence that it is a circle of circumference $1/T$, where T is the temperature. The topological term has the same form as the Berry phase term of the path integral for spin, the “Wess-Zumino” term of Eq.(44), but its prefactor is 1/2 as big. Thus, for a system with an open boundary the topological term yields a net contribution equal to a Berry phase with a net prefactor of $S/2$.

In other words, the boundary of the integer spin chain behaves as a localized degree of freedom whose spin is $1/2 \pmod{\text{an integer}}$. From the periodicity requirement we also see that if the spin chain is made of *odd-integer* degrees of freedom, there should be a spin 1/2 degree of freedom localized at the open boundary!. Conversely, if the chain is made of *even-integer* spins, there is no boundary degree of freedom! This line of argument implies that an antiferromagnetic chain with odd-integer spins must have a spin-1/2 degree of freedom at the boundary whereas a chain of even-integer spins should not. Notice that the “bulk” behavior is the same for both odd and even integer spin chains. The difference is whether or not they have a non-trivial “zero-mode” state at the boundary. In particular, the existence of this state is robust, i.e. it cannot be removed by making smooth changes to the quantum Hamiltonian or, what is the same, the boundary state is *topologically protected*.

A simple system that displays a protected spin-1/2 zero mode at the boundary is a generalized $S = 1$ spin chain with Hamiltonian

$$H = \alpha \sum_{j=1}^N \mathbf{S}(j) \cdot \mathbf{S}(j+1) + \beta \sum_{j=1}^N (\mathbf{S}(j) \cdot \mathbf{S}(j+1))^2 \quad (105)$$

where $S(j) = (S_x(j), S_y(j), S_z(j))$ are the spin 1 matrices at each lattice site j . This problem was examined in great detail by Affleck, Kennedy, Lieb and Tasaki [121] who showed that at the special value of the parameters $\alpha = 1/2$ and $\beta = 1/6$ this Hamiltonian takes the form of a sum of projection operators

$$H = \sum_j P_2(S(j) + S(j+1)) \quad (106)$$

where P_2 is an operator that projects out the spin 2 states. These authors constructed the exact ground state, known as the AKLT state, of this Hamiltonian by writing each spin 1 degree of freedom of two spin-1/2 degrees of freedom at each site. They showed that the ground state is a projected product state in which the ‘‘constituent’’ spin-1/2 degrees of freedom (each labeled by + and – respectively) on nearby sites j and $j+1$ are in a valence bond singlet state of the form $\frac{1}{\sqrt{2}}(|\uparrow_{j,+}, \downarrow_{j+1,-}\rangle - |\downarrow_{j,+}, \uparrow_{j+1,-}\rangle)$ (and symmetrizing at each site to project onto a spin 1 state). This state of the spin-1 chain is translation invariant and gapped and hence agrees with Haldane’s result. Moreover, it has a spin-1/2 degree of freedom at each open boundary [122].

The arguments discussed above show that the AKLT state is a gapped topological state. The gapless spin-1/2 boundary states are an example of *spin fractionalization*. The spin-1/2 boundary degrees of freedom are an example of an *edge state* which are present in many, thought not all, topological phases of matter. One may ask what symmetry protects the gaplessness of these edge degrees of freedom. The only perturbation that would give a finite energy gap to the spin-1/2 edge states is an external magnetic field. However, this perturbation would break the global SU(2) symmetry of the Hamiltonian as well as time reversal invariance.

6. Duality in Ising Models

Duality plays a significant role of our understanding of statistical physics and of quantum field theory. Many seemingly unrelated correspondences between different theories have come to be called *dualities*.

6.1. Duality in the 2D Ising Model

One of the earliest versions of duality transformations was used to relate the high-temperature expansion of the 2D classical Ising model and its low-temperature expansion [123]. In all dimensions, for concreteness we will think of a hypercubic lattice, the high temperature expansion is a representation of the partition function as a sum of contributions of closed loops on the lattice. At temperature T , a configuration of loops γ contributes with a weight which in the Ising model has the form $C(\gamma) \tanh^{L(\gamma)}(\beta)$, where $\beta = 1/T$ and $L(\gamma)$ is the length of the loop γ , i.e. the number of links on the loop, and $C(\gamma)$ is an entropic factor that counts the number of allowed loops with fixed perimeter $L(\gamma)$. However not all loop configurations are allowed as in the Ising model they satisfy constraints such as being non-overlapping, etc.

In section 4.2 we noted that the partition function of the classical Ising model (in any dimension) can be interpreted as the path-integral of a quantum spin model in one dimension less on a lattice with a discretized imaginary time. In this picture, the loops γ can be regarded as processes in which pairs of particles are created at some initial (imaginary) time, evolve and eventually are annihilated at a later (imaginary) time. In other words, the high temperature phase is a theory of a massive scalar field. The restrictions on the allowed loop configurations represents interactions among these particles. In the temperature range in which the expansion in loops is convergent the particles are massive as the loops are small. As the radius of convergence of the expansion is approached, longer and increasingly fractal-like loops begin to dominate the partition function, and concomitantly the mass of the particles decreases. This process is the signal of the approach to a continuous phase transition where the particles become massless. In fact, right at the critical point the particle interpretation is lost as the associated fields acquire anomalous dimensions.

Returning to the 2D classical Ising model, Kramers and Wannier also considered the low temperature expansion. This is an expansion around one of the broken symmetry state, e.g. the state with all spins up. In this low temperature regime the partition function is a sum of configurations of flipped spins. A typical configuration is a set of clusters of flipped spins. In the absence of a uniform field, a configuration of flipped spins has a energy cost only on links of the lattice with oppositely aligned spins (“broken bonds”). Thus a cluster of flipped spins costs an energy equal to 2 (I assumed that I set $J = 1$) for each broken bond at the *boundary* of the cluster. This boundary is *domain wall* which is a closed loop on the *dual lattice*. In 2D the dual of the square lattice is the square lattice of the dual sites (the centers of the elementary plaquettes of the

2D lattice). In 2D the links of the direct lattice pierce the links of the dual lattice. We can see that this analogous to the the geometric duality of forms: sites (“0-forms”) are dual to plaquettes (“2-forms”) and links (“1-forms”) are dual to links (also “1-forms”).

Thus, in 2D the low temperature expansion is an expansion in the loops γ^* of the dual lattice that represent the domain walls. We can also regard the domain walls (the dual loops) as the histories of pairs of particles on the dual lattice. However, the weight of each dual loop is $\exp(-2\beta)$ per link of γ^* . Except for that, the counting and restrictions on the dual loops γ^* are the same as those of γ . This means that there is a one-to-one correspondence between the two expansions with the replacement $\tanh\beta \leftrightarrow \exp(-2\beta^*)$. This mapping means that the dual of the 2D classical Ising model (with global \mathbb{Z}_2 symmetry) at inverse temperature β is a dual Ising model (also with global \mathbb{Z}_2 symmetry) on the dual lattice at inverse temperature β^* . This correspondence is in close analogy with what we discussed in section 5.1.1.

In particular if one assumes that there is a transition at $\beta_c = 1/T_c$ then there should also be a transition at $\beta_c^* = -1/2 \ln \tanh\beta_c$. Moreover, if one further assumes (as Kramers and Wannier did [123]) that there is a unique transition (correct in the Ising model but not in other cases), then the critical point must be such that $\exp(-2\beta_c) = \tanh\beta_c$, which yields the value $T_c = 2/\ln(\sqrt{2} + 1)$, which agrees with the Onsager result [20].

In section 4.2 we showed that the classical 2D Ising model is equivalent to the one-dimensional Ising model in a transverse field whose Hamiltonian is given in Eq.(23). The 1D Ising model on a transverse field has spin degrees of freedom defined on the sites of a one-dimensional chain labeled by an integer-valued variable n . The 1D Hamiltonian is expressed in terms of local operators, the Pauli matrices $\sigma_3(n)$ and $\sigma_1(n)$. The 1D chain has a dual lattice whose sites are the midpoints of the chain. Thus in 1D sites (“0-forms”) are dual to links (“1-forms”) and viceversa.

We will now see that there is a Hamiltonian version of the Kramers-Wannier duality [70]. In Eq.(26) we introduced the kink creation operator which flips are σ_3 operators to the left and including site j . For clarity we will denote the kink creation operator operator as $\tau_3(\tilde{n})$, where \tilde{n} is the site of the dual lattice between the sites n and $n + 1$ of the original lattice. We will now define an operator $\tau_1(\tilde{n}) = \sigma_3(n)\sigma_3(n+1)$. The operators $\tau_1(\tilde{n})$ and $\tau_3(\tilde{n})$ satisfy the same algebra os the Pauli operators $\sigma_1(n)$ and $\sigma_3(n)$. Furthermore, we readily find the Hamiltonian of the dual theory is the same (up to boundary conditions) as the original Hamiltonian of Eq.(23) except that the dual coupling constant is $\lambda^* = 1/\lambda$. So, once again, if we assume that there is a single (quantum) phase transition we require $\lambda_c^* = \lambda_c$ which is only satisfied by $\lambda_c = 1$. In this language the kink creation operator plays the role of a disorder operator [70].

6.2. The 3D duality: \mathbb{Z}_2 gauge theory

We will now discuss the role of duality in the 3D Ising model on a cubic lattice. This case, and its generalizations to higher dimensions, was considered first by Franz Wegner [124].

In section 6.1 we discussed the loop representation of the high temperature expansion and showed that it has the form form in all dimensions. Hence, in 3D the high temperature expansion is an expansion in closed loops γ with weight of $\tanh\beta$ per unit link of loop. Hence, in 3D as well, the high temperature phase can be regarded as field theory of a massive scalar field. As in the 2D case, the weight of a loop configuration is $\tanh\beta$ for each link of the loop γ .

However, the low temperature expansion has a radically different form and physical interpretation. Much as in 2D, the low temperature expansion is an expansion in clusters of flipped spins. Here too, in the absence of an external field, the only energy cost resides at the boundary of the clusters of overturned spins, the domain walls. But in 3D the clusters occupy volumes whose boundaries are closed surfaces Σ^* . In 3D a link (bond) is dual to a plaquette (expressing the fact that in 3D a 1-form is dual to a 2-form). Hence the closed domain walls of overturned spins is dual to a closed surface Σ^* on the dual lattice. The weight of each configuration of closed surfaces is $\exp(-2\beta)$ for each plaquette of the surface Σ^* .

These facts mean that the dual of the 3D Ising model is a theory on the dual lattice with coupling constant β^* , with $\exp(-2\beta^*) = \tanh\beta$, such that its expansion for small β^* is a sum over configurations of closed surfaces σ^* , and for large β^* is a sum of configurations of closed loops γ^* on the dual lattice. The dual theory is the naturally defined on (dual) plaquettes, not on links. To this end, let us define a set of Ising-like degrees of freedom $\sigma_\mu(x) = \pm 1$ located on the links (x, μ) of the dual lattice. These degrees of freedom are coupled on each plaquette of the lattice. Since each plaquette has four links, the coupling involves the degrees of freedom on all four links

of each plaquette. The partition function of the dual theory is

$$Z = \sum_{\{\sigma_\mu(x)\}} \exp \left(\beta^* \sum_{\text{plaquettes}} \sigma_\mu(x) \sigma_\nu(x + e_\mu) \sigma_\mu(x + e_\mu + e_\nu) \sigma_\nu(x) \right) \quad (107)$$

where the sum in the exponent (the negative of the “action”) runs on all the plaquettes of the dual lattice.

The action of the theory of Eq.(107) is invariant under the reversal of all six Ising degrees of freedom on links sharing a given site x . This is a *local symmetry*. Unlike the 3D Ising model, which has a global \mathbb{Z}_2 symmetry of flipping all spins simultaneously, this theory has *local* (or *gauge*) \mathbb{Z}_2 symmetry. This is the simplest example of a lattice gauge theory in which the degrees of freedom are gauge fields that take values on the \mathbb{Z}_2 gauge group [124, 32, 71].

We saw that in the spin model the high temperature expansion (i.e. the expansion in powers of $\tanh \beta$) is a sum over loop configurations which can be interpreted in terms of processes in which pairs of particles are created, evolve (in imaginary time) and then are destroyed (also in pairs). The analogous interpretation of the expansion of the \mathbb{Z}_2 gauge theory in powers of $\tanh \beta^* = \exp(-2\beta)$ as a sum over the configurations of closed surfaces is not in terms of the histories of *particles* but in terms of the histories of *closed strings* which, as they evolve, sweep the closed surfaces. The physics is, however, more complex as the sum over surfaces runs over all surfaces of arbitrary topology with arbitrary number of handles (or *genus*). Thus, over (imaginary) time a closed string is created, evolves, splits into two closed strings, etc.

Therefore, the 3D Ising model, which has a \mathbb{Z}_2 global symmetry is dual to a \mathbb{Z}_2 gauge theory which has a \mathbb{Z}_2 local symmetry. The duality maps the high temperature (disordered) phase of the Ising model to the strong coupling (small β^*) phase of the gauge theory. In the gauge theory language this is the *confining* phase. This can be seen by computing the expectation value of the Wilson loop operator on the closed loop Γ , which here reads $\langle \prod_{(x,\mu) \in \Gamma} \sigma_\mu(x) \rangle$. In the small β^* phase this expectation value decays exponentially with the size of the minimal surface bounded by the loop Γ . This behavior of the area law of the Wilson loop. Under duality the insertion of this operator is equivalent to an Ising model with a domain wall terminating on the loop Γ , and the area law of the Wilson loop is the consequence of the fact that if the Ising model has long range order, the free energy cost of the domain wall scales with its area. Moreover, in this phase of the gauge theory the closed strings are small meaning that the string tension (the energy per unit length of string) is finite. In the Ising model language, the string tension becomes surface tension of the domain wall.

In the preceding subsection we discussed a Hamiltonian version of the duality. We will briefly do the same in the case of the 2+1 dimensional Ising model. The hamiltonian of the Ising model in a transverse field on a square lattice is

$$H_{2D-TFIM} = - \sum_{\mathbf{r}} \sigma_1(\mathbf{r}) - \lambda \sum_{\mathbf{r}, j=1,2} \sigma_3(\mathbf{r}) \sigma_3(\mathbf{r} + \mathbf{e}_j) \quad (108)$$

where \mathbf{r} labels the sites of the square lattice and \mathbf{e}_j (with $j = 1, 2$) are the two (orthonormal) primitive unit vectors of the lattice. Here too, at each site we have a two-level system (the spins), σ_1 and σ_3 are Pauli matrices acting on these states at each site and λ is the coupling constant. Just as in the 1D case, this system has two phases: an ordered phase for $\lambda > \lambda_c$ and a disordered phase for $\lambda < \lambda_c$, where λ_c is a critical coupling. This Hamiltonian is invariant under the global \mathbb{Z}_2 symmetry generated by the global spin flip operator $\mathcal{Q} = \prod_{\mathbf{r}} \sigma_1(\mathbf{r})$ which commutes with the Hamiltonian, $[\mathcal{Q}, H_{2D-TFIM}] = 0$.

Let us consider now a \mathbb{Z}_2 gauge theory on the dual of the square lattice. We will now define a two-dimensional Hilbert space on each link, denoted by $(\tilde{\mathbf{r}}, j)$ (with $j = 1, 2$) of the dual lattice with sites labelled by $\tilde{\mathbf{r}}$. We will further define at each link $(\tilde{\mathbf{r}}, j)$ the Pauli operators $\sigma_j^1(\tilde{\mathbf{r}})$ and $\sigma_j^3(\tilde{\mathbf{r}})$. The Hamiltonian of the \mathbb{Z}_2 gauge theory is

$$H_{\mathbb{Z}_2 \text{gauge}} = - \sum_{\tilde{\mathbf{r}}, j} \sigma_j^1(\tilde{\mathbf{r}}) - g \sum_{\tilde{\mathbf{r}}} \sigma_1^3(\tilde{\mathbf{r}}) \sigma_2^3(\tilde{\mathbf{r}} + \tilde{\mathbf{e}}_1) \sigma_1^3(\tilde{\mathbf{r}} + \tilde{\mathbf{e}}_1 + \tilde{\mathbf{e}}_2) \sigma_2^3(\tilde{\mathbf{r}}) \quad (109)$$

where $\tilde{\mathbf{e}}_j = \epsilon_{jk} \mathbf{e}_k$ (with $j, k = 1, 2$).

Unlike the Hamiltonian of Eq.(108), the Hamiltonian of Eq.(109) has a local symmetry. Let $\tilde{\mathcal{Q}}(\tilde{\mathbf{r}})$ be the operator that flips the \mathbb{Z}_2 gauge degrees of freedom on the four links that share the site $\tilde{\mathbf{r}}$,

$$\tilde{\mathcal{Q}}(\tilde{\mathbf{r}}) = \prod_{j=1,2} \sigma_j^1(\tilde{\mathbf{r}}) \sigma_j^1(\tilde{\mathbf{r}} - \tilde{\mathbf{e}}_j) \quad (110)$$

These operators commute with each other, $[\tilde{Q}(\tilde{\mathbf{r}}), \tilde{Q}(\tilde{\mathbf{r}}')] = 0$, and commute with the Hamiltonian, $[H_{\mathbb{Z}_2\text{gauge}}, \tilde{Q}(\tilde{\mathbf{r}})] = 0$. The Hilbert space of gauge-invariant states are eigenstates of $\tilde{Q}(\tilde{\mathbf{r}})$ with unit eigenvalue,

$$\tilde{Q}(\tilde{\mathbf{r}})|\text{Phys}\rangle = |\text{Phys}\rangle \quad (111)$$

(for all $\tilde{\mathbf{r}}$). This constraint is the the \mathbb{Z}_2 Gauss Law. The Hamiltonian $H_{\mathbb{Z}_2\text{gauge}}$ has two phases: a *confining phase* for $g < g_c$, and a *deconfined phase* for $g > g_c$ (where g_c is a critical coupling).

The duality transformation is defined so that

$$\sigma_1(\mathbf{r}) = \prod_{\text{plaquette}(\mathbf{r})} \sigma_j^3(\tilde{\mathbf{r}}) \quad (112)$$

is the product of the σ_j^3 operators of the gauge theory on a dual plaquette centered at the site \mathbf{r} , and

$$\sigma_1^1(\tilde{\mathbf{r}}) = \sigma_3(\mathbf{r})\sigma_3(\mathbf{r} + \mathbf{e}_2), \quad \sigma_2^1(\tilde{\mathbf{r}}) = \sigma_3(\mathbf{r})\sigma_3(\mathbf{r} + \mathbf{e}_1) \quad (113)$$

These identities imply that the gauge theory constraint of the \mathbb{Z}_2 Gauss Law, Eq.(111), is satisfied. This also means that duality is a mapping of the *gauge-invariant* sector of the \mathbb{Z}_2 gauge theory to the Hilbert space of the Ising model in a transverse field.

Eq.(112) and Eq.(113) imply that the Hamiltonians $H_{\text{2D-TFM}}$ and $H_{\mathbb{Z}_2\text{gauge}}$ are equal to each other with the identification of the coupling constants $g = 1/\lambda$. Hence, the *ordered phase* of the Ising model, $\lambda > \lambda_c$, maps onto the *confining phase* of the \mathbb{Z}_2 gauge theory and, conversely, the *disordered phase* of the Ising model maps onto the *deconfined phase* of the gauge theory.

Eq.(113) also implies that the operator $\sigma_3(\mathbf{r})$ of the Ising model can be identified with the operator

$$\sigma_3(\mathbf{r}) = \prod_{\gamma(\mathbf{r})} \sigma_j^1(\tilde{\mathbf{r}}) \quad (114)$$

where the product is on the links of the dual lattice pierced by the path $\gamma(\mathbf{r})$ on the direct lattice ending at the site \mathbf{r} .

While $\sigma_3(\mathbf{r})$ is of course just the order parameter of the Ising model, the dual operator defined by Eq.(114) anti-commutes with the plaquette term of the Hamiltonian of Eq.(109) and creates an \mathbb{Z}_2 flux excitation at the plaquette. With some abuse of language, this operator can be regarded as creating a \mathbb{Z}_2 “magnetic monopole”. Since in the confining phase this operator has an expectation value, we can regard this phase as a magnetic condensate.

We will now discuss the role of boundary conditions which is different in both theories. Let us examine the behavior of the \mathbb{Z}_2 gauge theory with periodic boundary conditions. Periodic boundary conditions means that the 2D space is topologically a 2-torus. It is straightforward to show that the ground state in the confining phase is unique and insensitive to boundary conditions. The physics of the deconfined phase is more subtle. We will now see that on a 2-torus it has a four-fold degenerate ground state and that the degeneracy is not due to the spontaneous breaking of a global symmetry.

Let γ_1 and γ_2 be two non-contractible loops on the torus along the directions 1 and 2 respectively. Let us consider the (“electric”) Wilson loop operators along γ_1 and γ_2 , $W[\gamma_1] = \prod_{(r,j) \in \gamma_1} \sigma_j^3(\mathbf{r})$ and similarly for γ_2 . Similarly let us consider the (“magnetic”) ’t Hooft operators $\tilde{W}[\tilde{\gamma}_1]$ and $\tilde{W}[\tilde{\gamma}_2]$ defined on the non-contractible closed paths of the dual lattice $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, such that $\tilde{W}[\tilde{\gamma}_1] = \prod_{\tilde{\gamma}_1} \sigma_2^1(\tilde{\mathbf{r}})$ and $\tilde{W}[\tilde{\gamma}_2] = \prod_{\tilde{\gamma}_2} \sigma_1^1(\tilde{\mathbf{r}})$. It is easy to see that the Wilson and ’t Hooft operators satisfy that $W[\gamma_i]^2 = \tilde{W}[\tilde{\gamma}_i]^2 = I$, and that

$$[W[\gamma_i], W[\gamma_j]] = [\tilde{W}[\tilde{\gamma}_i], \tilde{W}[\tilde{\gamma}_j]] = 0, \quad \{W[\gamma_1], \tilde{W}[\tilde{\gamma}_2]\} = \{W[\gamma_2], \tilde{W}[\tilde{\gamma}_1]\} = 0 \quad (115)$$

Let us consider the special limit of $g \rightarrow \infty$. In this limit the Wilson loop operators $W[\gamma_i]$ on the two non-contractible loops of the 2-torus commute with the Hamiltonian $H_{\mathbb{Z}_2\text{gauge}}$ of Eq.(109). Therefore the eigenstates of the Hamiltonian can be chosen to be the eigenstates of these two Wilson loops. Since the Wilson loop operators are hermitian and obey $W[\gamma_i]^2 = I$, the spectrum of each loop is two-dimensional $|\pm 1\rangle_i$ ($i = 1, 2$) with eigenvalues ± 1 , respectively. At $g \rightarrow \infty$ these states are also eigenstates of $H_{\mathbb{Z}_2\text{gauge}}$. Thus, the ground state of $H_{\mathbb{Z}_2\text{gauge}}$ is four dimensional. The ’t Hooft loops $\tilde{W}[\tilde{\gamma}_i]$ act as ladder operators in this restricted Hilbert space.

The conclusion is that, on a 2-torus, at least in the limit $g \rightarrow \infty$, the ground state of $H_{\mathbb{Z}_2\text{gauge}}$ is degenerate. However, this degeneracy is not due to the spontaneous breaking of a global symmetry. Rather, it reflects the topological character of the theory. To see this, one can extend this analysis to a theory to be on a more general two-dimensional and instead of a 2-torus we can

consider a surface with g handles (not to be confused with the coupling constant!), e.g. $g = 0$ for a sphere (or disk), $g = 1$ for a 2-torus, $g = 2$ for a pretzel, etc. For each of these two-dimensional surfaces the number different number of non-contractible loops is g , and the Wilson loops defined on them commute with each other (and with $H_{\mathbb{Z}_2\text{gauge}}$). Hence, in the limit $g \rightarrow \infty$ $H_{\mathbb{Z}_2\text{gauge}}$ has a ground state degeneracy of 2^g which, clearly, depends only on the topology of the surface. We conclude that, at least as $g \rightarrow \infty$, the Hamiltonian $H_{\mathbb{Z}_2\text{gauge}}$ is in a *topological phase*.

However, is the exact degeneracy found in the limit $g \rightarrow \infty$ a property of the entire deconfined phase? For a finite system the expansion in powers of $1/g$ is convergent. On the other hand, in the thermodynamic limit, $L \rightarrow \infty$, the expansion has a finite radius of convergence with g_c being an upper bound. Let us consider the ground state at $g = \infty$ with eigenvalue +1 for the Wilson loop $W[\gamma_1]$. The degenerate state with eigenvalue -1 is created by the 't Hooft loop $\tilde{W}[\tilde{\gamma}_2]$, i.e. $| -1 \rangle_1 = \tilde{W}[\tilde{\gamma}_2] | +1 \rangle_1$. The 't Hooft operator is a product of σ^1 operators on links along the direction 1 crossed by the path $\tilde{\gamma}_2$ of the dual lattice, which involves (at least) L links. Then, it takes L orders in the expansion in powers of $1/g$ to mix the state $| +1 \rangle_1$ with the state $| -1 \rangle_1$, and this amplitude is of order $1/g^L$, which is exponentially small. The same argument applies to the mixing between all four states. For a finite but large system of linear size L , the degeneracy is lifted but the energy splitting is exponentially small in the system size. Hence, the topological protection is a feature of the deconfined phase in the thermodynamic limit $L \rightarrow \infty$.

We conclude that the deconfined phase of the \mathbb{Z}_2 lattice gauge theory is a topological phase. This result extends to the case of a gauge theory with discrete gauge group \mathbb{Z}_k , the cyclic group of k elements. In general spacetime dimension $D > 2$ the \mathbb{Z}_k gauge theory also has a confined and a deconfined phase and if $k \gtrsim 4$ it also has an intermediate “Coulomb phase”[125, 126]. In section 9.4 we will see that the low-energy (IR) regime of the deconfined phase is described by a topological quantum field theory known as the level k BF theory.

7. Bosonization

The behavior of one-dimensional electronic systems is of great interest in Condensed Matter Physics for many reasons. One is that, even for infinitesimally weak interactions, these one-dimensional metals violate the basic principles of the Landau Theory of the Fermi Liquid [13]. A central assumption of Fermi Liquid theory is that at low energies the excitation energy of the fermion (electron) quasiparticle is always much larger than its width. Hence, at asymptotically low energies the quasiparticle excitations become increasingly sharp.

A manifestation of this feature is that the quasiparticle propagator (the “Green function”) has a pole on the real frequency axis with a finite residue Z . In one-dimensional metals this assumption always fails since the residue Z vanishes, the Fermi field acquires a non-trivial anomalous dimension, and the pole is replaced by a branch cut. To this date this is the best example of what is called a “non-Fermi liquid”. In one-dimensional metals this non-Fermi liquid is often called a “Luttinger Liquid” [127]. A detailed analysis can be found in chapter 6 of Ref. [9], and in other books.

Bosonization provides for a powerful tool to understand the physics of these non-trivial systems. Bosonization is a duality between a massless Dirac field in 1+1 dimensions and a (also massless) relativistic Bose (scalar) field [128, 129, 130, 131]. In this context, bosonization is a set of operator identities relating observables between two different (dual) continuum field theories. These identities have a close resemblance to the Jordan-Wigner transformation which relates operators of a theory of spinless (Dirac) fermions on a one-dimensional lattice to a theory of bosons with hard cores (i.e. spins) on the same lattice [22].

7.1. Dirac fermions in one space dimensions

To understand how these operator identities come about and what these fields mean in the Condensed Matter context we will consider the simple problem of a system of non-interacting spinless fermions $c(n)$ on a one-dimensional chain of length L (with $L \rightarrow \infty$), for simplicity with periodic boundary conditions. The Hamiltonian is

$$H_0 = -t \sum_{n=1}^L c(n)^\dagger c(n+1) + \text{h.c.} \quad (116)$$

In momentum space, and in the thermodynamic limit $L \rightarrow \infty$, the Hamiltonian becomes

$$H_0 = \int_{-\pi}^{\pi} \frac{dp}{2\pi} (\varepsilon_0(p) - \mu) c^\dagger(p) c(p) \quad (117)$$

where $-\pi \leq p \leq \pi$, μ is the chemical potential (which fixes the fermion number) and $\varepsilon_0(p)$ is the (free) quasiparticle energy which, in this case, is

$$\varepsilon_0(p) = -2t \cos p \quad (118)$$

This simple system has a global internal symmetry

$$c'(n) = e^{i\alpha} c(n), \quad c'(n)^\dagger = e^{-i\alpha} c(n) \quad (119)$$

where α is a constant phase (with period 2π). This global symmetry reflects the conservation of fermion number

$$N_F = \sum_n c^\dagger(n)c(n) \quad (120)$$

The free fermion Hamiltonian of Eq.(116) is also invariant under lattice translations.

The ground state of this system is obtained by occupying all single particle states with energy below the chemical potential, $E \leq \mu \equiv E_F$, which defines the Fermi energy E_F . In what follows we will redefine the zero of the energy at the Fermi energy. In the thermodynamic limit the one-particle states are labeled by momenta defined in the first Brillouin zone $[-\pi, \pi]$. The ground state $|G\rangle$ of this system is obtained by filling up all single-particle states with momentum p in the range $[-p_F, p_F]$, where p_F is the Fermi momentum. Hence, the occupied states have energy $\varepsilon_0(p) \leq E_F$. The ground state is

$$|G\rangle = \prod_{|p| \leq p_F} c^\dagger(p)|0\rangle \quad (121)$$

and is called the *filled Fermi sea*. We will further assume that the fermionic system is dense in the sense that the occupied states are a finite fraction of the available states, i.e. we assume that $|E_F|$ is a *finite* fraction of the bandwidth $W = 4t$.

The single-particle excitations have low energy if $|\varepsilon_0(p) - E_F| \ll |E_F|$. This range can only be accessed by quasiparticles with momenta p close to $\pm p_F$. For states with p close to p_F we can a linearized approximation and write

$$\varepsilon_0(p) \simeq v_F(p - p_F) + \dots \quad (122)$$

and similarly for the states near $-p_F$. Here $v_F = \frac{\partial \varepsilon_0}{\partial p}|_{p_F} = 2t \sin p_F$ is the Fermi velocity. Let q being the momentum measured from p_F (or $-p_F$ in the other case). This approximation is correct in a range of momenta $-\Lambda \leq q \leq \Lambda$, where $\frac{2\pi}{L} \ll \Lambda \ll \pi$. In this regime we can approximate the lattice fields $c(n)$ with two continuum right moving $\psi_R(x)$ and left moving $\psi_L(x)$ Fermi fields, such that (with $x = na_0$)

$$c(n) \sim e^{ip_F x} \psi_R(x) + e^{-ip_F x} \psi_L(x) \quad (123)$$

in terms of which the Hamiltonian takes the continuum form

$$\begin{aligned} H_{\text{continuum}} &= \int dx \left(\psi_R^\dagger(x)(-iv_F)\partial_x \psi_R(x) - \psi_L^\dagger(x)(-iv_F)\partial_x \psi_L(x) \right) \\ &= \int \frac{dq}{2\pi} q v_F \left(\psi_R^\dagger(q)\psi_R(q) + \psi_L^\dagger(q)\psi_L(q) \right) \end{aligned} \quad (124)$$

In what follows I will rescale time and space in such a way as to set $v_F = 1$.

Eq.(124) is the Hamiltonian of a massless Dirac field in 1+1 dimensions. It describes the effective low energy behavior of the excitations of the fermionic system defined on a lattice by the Hamiltonian of Eq.(116). Here low energy means asymptotically close to the Fermi energy (which we have set to zero) and momenta close to $\pm p_F$. We will see that this effective relativistic field theory gives a complete description of the universal low energy physics encoded in the microscopic model of Eq.(116) up to some subtle issues associated with what are called quantum anomalies.

Let us denote by $\psi(x)$ the bi-spinor field $\psi(x) = (\psi_R(x), \psi_L(x))$ and the 2×2 Dirac gamma-matrices in terms of the three Pauli matrices are

$$\gamma_0 = \sigma_1, \quad \gamma_1 = -i\sigma_2, \quad \gamma_5 = \sigma_3 \quad (125)$$

which obey the Dirac algebra

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu} \quad (126)$$

where $g_{\mu\nu}$ is the metric tensor in 1+1-dimensional Minkowski spacetime

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (127)$$

Using the standard notation $\bar{\psi}(x) = \psi^\dagger(x)\gamma_0$ (where I left the spinor index $\alpha = 1, 2$ implicit) the Lagrangian density of the free massless Dirac fermion is

$$\mathcal{L} = \bar{\psi} i\partial^\mu \psi \quad (128)$$

where we have used the Feynman slash notation which denotes the contraction of a vector field, say A_μ with the Dirac gamma matrices with a slash, $A_\mu \gamma^\mu \equiv A$.

Formally, the Dirac lagrangian of Eq.(128) (and the Dirac Hamiltonian of Eq.(124)) are invariant under two separate global transformations. It has a global U(1) (gauge) symmetry transformation

$$\psi'(x) = e^{i\theta} \psi(x) \quad (129)$$

(where θ is constant) under which the two components of the bi-spinor ψ transform in the same way

$$\psi'_R(x) = e^{i\theta} \psi_R(x), \quad \psi'_L(x) = e^{i\theta} \psi_L(x) \quad (130)$$

The massless Dirac theory is also invariant under a global gauge U(1) chiral transformation

$$\psi'(x) = e^{i\theta\gamma_5} \psi(x) \quad (131)$$

which in components it reads

$$\psi'_R(x) = e^{i\theta} \psi_R(x), \quad \psi'_L(x) = e^{-i\theta} \psi_L(x) \quad (132)$$

In general, if a system has a global continuous symmetry it should have a locally conserved current and a globally conserved charge. This is the content of Noether's Theorem. Since the massless Dirac theory has these two global symmetries one would expect that it should have two separately conserved currents.

The global U(1) symmetry has an associated current $j_\mu = (j_0, j_1)$ given by

$$j_\mu = \bar{\psi} \gamma_\mu \psi \quad (133)$$

which is invariant under the global U(1) symmetry. In terms of the right and left moving Dirac fields, ψ_R and ψ_L , the components of the current are

$$j_0 = \psi_R^\dagger \psi_R + \psi_L^\dagger \psi_L, \quad j_1 = \psi_R^\dagger \psi_R - \psi_L^\dagger \psi_L \quad (134)$$

This current is locally conserved and satisfies the continuity equation

$$\partial_\mu j^\mu = 0 \quad (135)$$

It has an associated conserved global charge, which we will call fermion number

$$Q = \int_{-\infty}^{\infty} dx j_0(x) = \int_{-\infty}^{\infty} dx (\psi_R^\dagger \psi_R + \psi_L^\dagger \psi_L) \quad (136)$$

This is the continuum version of the conservation of fermion number N_F of the free fermion lattice model discussed above.

Since this current is conserved it can be coupled to an electromagnetic field through a term in the Lagrangian

$$\mathcal{L}_{\text{int}} = -e j_\mu A^\mu \quad (137)$$

where A^μ is the electromagnetic vector potential, which in 1+1 dimensions has only two components, $A^\mu = (A_0, A_1)$. Here e is a coupling constant which is interpreted as the electric charge. This coupling amounts to making the U(1) symmetry a local gauge symmetry under which the fields transform as follows

$$\psi'(x) = e^{i\theta(x)} \psi(x), \quad A'_\mu(x) = A_\mu(x) + \frac{1}{e} \partial_\mu \theta(x) \quad (138)$$

Now the conservation of fermion number becomes the conservation of the total electric charge

$$Q_e = -eQ \quad (139)$$

In a continuum non-relativistic model the Fermi field is $\psi(x)$ (ignoring spin), e.g. an electron gas in one dimension such as a quantum wire, the analog of decomposition shown in Eq.(123) of the electron field $\psi(x)$ becomes

$$\psi(x) = e^{ip_F x} \psi_R(x) + e^{-ip_F x} \psi_L(x) \quad (140)$$

The electron field $\psi(x)$ transforms as $\psi'(x) = e^{i\alpha} \psi(x)$ under the global U(1) gauge symmetry of Eq.(130). In particular, the local density operator $\rho(x) = \psi^\dagger(x)\psi(x)$ is invariant under this symmetry. Under both decompositions of Eqs. (123) and (140), the local density operator becomes

$$\rho(x) = \psi_R^\dagger(x)\psi_R(x) + \psi_L^\dagger(x)\psi_L(x) + e^{i2p_F x} \psi_R^\dagger(x)\psi_L(x) + e^{-i2p_F x} \psi_L^\dagger(x)\psi_R(x) \quad (141)$$

Clearly the non-relativistic density operator $\rho(x)$ is invariant under the global U(1) gauge symmetry. The same considerations applies to the lattice version of the density operator (the local occupation number).

Thus, the density operator $\rho(x)$ can be decomposed into a slowly varying part (the first two terms in Eq.(141)) and the last two terms which oscillate with wave vectors $Q = \pm 2p_F$. These observations imply that we can express $\rho(x)$ in the form

$$\rho(x) = \bar{\rho} + j_0(x) + e^{iQx} \psi_Q(x) + e^{-iQx} \psi_{-Q}(x) \quad (142)$$

This decomposition can be interpreted as a Fourier expansion of the density operator in term of newly defined slowly-varying fields. In Eq.(142) $Q = 2p_F$ and $\bar{\rho}$ is the average density, where we assumed that the Dirac density operator $j_0(x)$ has vanishing expectation value (i.e. it is normal-ordered). In Eq.(142) we defined the (bosonic) operators

$$\psi_Q(x) = \psi_R^\dagger(x)\psi_L(x), \quad \psi_{-Q}(x) = \psi_L^\dagger(x)\psi_R(x) \quad (143)$$

which characterize the oscillatory component of the density. The operators $\psi_{\pm Q}(x)$ are the order parameters of a charge density wave state in one dimension and Q is the ordering wavevector.

Repulsive interactions between the electrons can cause scattering process between the right and left moving components to become relevant (in the Renormalization Group sense) leading to the spontaneous breaking of translation invariance. The resulting state is known as a charge-density-wave (CDW). In this state the operators $\psi_{\pm Q}(x)$ (or a linear combination of them) acquire a non-vanishing expectation value, and the expectation value of the density operator $\rho(x)$ has a (static) modulated component, and the Dirac Hamiltonian density becomes

$$\mathcal{H} = \psi_R^\dagger(x)(-i)\partial_x \psi_R(x) - \psi_L^\dagger(x)(-i)\partial_x \psi_L(x) + m \left(\psi_R^\dagger(x)\psi_L(x) + \psi_L^\dagger(x)\psi_R(x) \right) \quad (144)$$

where m is the Dirac mass. The operator in the second term of Eq.(144) mixes the right and left moving components of the Dirac spinor. This hermitian operator is known as the Dirac mass term. In relativistic notation this operator is written

$$\bar{\psi}\psi = \psi_R^\dagger(x)\psi_L(x) + \psi_L^\dagger(x)\psi_R(x) \quad (145)$$

When $m \neq 0$, i.e. when $\langle \bar{\psi}\psi \rangle \neq 0$, the fermionic spectrum has a mass gap, and the electronic states with momenta $\pm p_F$ have an energy gap. Alternatively we could have considered a state with the in which the (hermitian) operator that has an expectation value is

$$i\bar{\psi}\gamma_5\psi = i \left(\psi_R^\dagger(x)\psi_L(x) - \psi_L^\dagger(x)\psi_R(x) \right) \quad (146)$$

which is known as the γ_5 mass term. In a more general CDW state both mass terms can be present.

7.2. Chiral symmetry and chiral symmetry breaking

The CDW states we introduced have different symmetries. To see this let us observe that the operators $\psi_{\pm Q}(x)$ transform non-trivially under a global U(1) chiral transformation

$$\psi'_{\pm Q}(x) = e^{\mp 2i\theta} \psi_{\pm Q}(x) \quad (147)$$

Upon substituting this transformation in the expansion of the density $\rho(x)$ of Eq.(142), we see that a chiral transformation is equivalent to a uniform displacement of the density operator by θ/p_F ,

$$\rho(x) \rightarrow \rho(x + \theta/p_F) \quad (148)$$

Under a chiral transformation with arbitrary angle θ , the Dirac and γ_5 mass term operators transform as an orthogonal transformation of a two-component vector. In particular for a chiral transformation with $\theta = \pi/4$,

$$\bar{\psi}\psi \mapsto i\bar{\psi}\gamma_5\psi, \quad i\bar{\psi}\gamma_5\psi \mapsto -\bar{\psi}\psi \quad (149)$$

which is equivalent to a displacement of the density by 1/4 of the period of the CDW. In this sense these two states are equivalent, as is the state with a more general linear combination.

The microscopic lattice model (and the continuum non-relativistic model) are invariant under the spatial inversion symmetry $x \leftrightarrow -x$. This also implies a symmetry under the exchange of right and left moving components of the Dirac spinor, $\psi_R \leftrightarrow \psi_L$. In the massless Dirac theory this operation is equivalent to the multiplication of the spinor by the Pauli matrix σ_1 . In the massless theory the multiplication by σ_2 (followed by a chiral transformation with $\theta = \pi/2$) has the same effect.

These symmetries have an important effect on the fermionic spectrum. To understand what they do let us consider the one-particle Dirac Hamiltonian with a Dirac mass m (in momentum space)

$$h = \begin{pmatrix} p & m \\ m & -p \end{pmatrix} \quad (150)$$

This operator anti-commutes with the Pauli matrix σ_2 , $\{h, \sigma_2\} = 0$. Let $|E\rangle$ be an eigenstate of the Hamiltonian h with energy eigenvalue $E = \sqrt{p^2 + m^2}$. Let us consider the state $\sigma_2|E\rangle$. It is also an eigenstate of h but with energy $-E$, i.e.

$$h\sigma_2|E\rangle = -\sigma_2h|E\rangle = -E|E\rangle, \quad \Rightarrow \sigma_2|E\rangle = |-E\rangle \quad (151)$$

Hence if $|E\rangle$ is an eigenstate of energy E , then $\sigma_2|E\rangle$ is an eigenstate of energy $-E$. This means that the spectrum is invariant under charge conjugation symmetry. Notice that under this operation the spinor transforms as

$$\sigma_2 \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} -i\psi_L \\ i\psi_R \end{pmatrix} = e^{-iy_5\pi/2} \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad (152)$$

In other words, this theory is invariant under charge conjugation C and parity \mathcal{P} which, combined, it implies that it is invariant under time-reversal \mathcal{T} .

The same consideration applies in the case of a γ_5 mass term in which case the one-particle Dirac Hamiltonian now is

$$h = \begin{pmatrix} p & -im_5 \\ im_5 & -p \end{pmatrix} \quad (153)$$

This Hamiltonian now anti-commutes with the Pauli matrix σ_1 which also implies that the same charge conjugation symmetry C , $|E\rangle \leftrightarrow |-E\rangle$, is present in the spectrum of the case of a γ_5 mass. We can also repeat the argument on parity invariance \mathcal{P} , which is now multiplication by σ_1 . Thus the theory is invariant under time-reversal \mathcal{T} .

However, if the theory has both a Dirac mass m and a γ_5 mass m_5 these symmetries are broken. Indeed, the one-particle Dirac Hamiltonian now is

$$h = \begin{pmatrix} p & m - im_5 \\ m + im_5 & -p \end{pmatrix} = p\sigma_3 + m\sigma_1 + m_5\sigma_2 \quad (154)$$

which no longer has a spectral symmetry. In this case $C\mathcal{P}$ is broken and, hence, so it \mathcal{T} since $C\mathcal{PT}$ remains unbroken (as it should).

In a lattice system this is a symmetry transformation only if $\theta = \pi n/p_F$ is a lattice displacement, which restricts the allowed values of the chiral angle to be discrete. Although this is true interactions play a significant role in the actual behavior. In fact, there are physical situations in which an effective continuous symmetry actually emerges in this he infrared (long-distance) limit. This is what happens when the CDW is incommensurate and, as we will see in the next subsection, it slides under the action of an electric field.

In the case of a half-filled system (with only nearest-neighbor hopping matrix elements) the Fermi wave vectors are $\pm\pi/2$. In this case the allowed discrete chiral transformation has a chiral

angle $\pi/2$ under which $\bar{\psi}\psi \mapsto -\bar{\psi}\psi$ (and similarly for $i\gamma_5\psi$) corresponding by a translation by one lattice spacing. The allowed four fermion operator is $(\bar{\psi}\psi)^2$. If the lattice fermions are spinless this operator reduces to

$$(\bar{\psi}\psi)^2 = -2j_R(x)j_L(x) + \lim_{y \rightarrow x} \psi_R^\dagger(x)\psi_L^\dagger(x)\psi_L(y)\psi_R(y) + \text{h.c.} \quad (155)$$

where introduced the right and left moving (chiral) components of the current operator

$$j_R(x) = \frac{1}{2}(j_0(x) + j_1(x)) = \psi_R^\dagger(x)\psi_R(x), \quad j_L(x) = \frac{1}{2}(j_0(x) - j_1(x)) = \psi_L^\dagger(x)\psi_L(x) \quad (156)$$

The first term in Eq.(155) is known as the backscattering interaction and has scaling dimension 2. Hence, it is a marginal operator. The theory with only the first term is known in Condensed Matter Physics as the Luttinger model and in High-Energy Physics at the (massless) Thirring model. The second operator in Eq.(155) formally violates momentum conservation as its total momentum is $4p_F = 2\pi$ which is a reciprocal lattice vector and, as such, it is equivalent to zero (mod 2π). Such an operator is not formally allowed in a (naive) continuum theory. This operator is due to a lattice Umklapp process and breaks the formal continuous chiral symmetry to a discrete \mathbb{Z}_2 subgroup. Although the naive scaling dimension of this operator is 2 (and hence it is formally marginal). If the fermions are spinless, the leading operator actually vanishes and the leading non-vanishing operator actually has dimension 4, which is irrelevant. However, as shown above, backscattering processes of the form $j_R(x)j_L(x)$ are part of this operator and are exactly marginal.

If the interaction is strong enough the backscattering interaction can make the Umklapp operator relevant. When this happens the fermionic system has a quantum phase transition to an insulating state with a period 2 (commensurate) CDW state. On the other hand, for spin 1/2 fermions the operator $(\bar{\psi}\psi)^2$ is allowed and is marginally relevant. If the interactions are repulsive the resulting state is an antiferromagnetic Néel state at quantum criticality, while for attractive interactions it is a period 2 CDW. This is what happens in the 1D Hubbard model (see, e.g. Ref. [9] for a detailed analysis).

There are two theories in relativistic systems which are closely related to this problem. One if the Gross-Neveu model [132] which is a theory of N species of massless Dirac spinors with Lagrangian density

$$\mathcal{L}_{GN} = \bar{\psi}_a i\partial\psi_a + g(\bar{\psi}_a\psi_a)^2 \quad (157)$$

with $a = 1, \dots, N$ (summation of repeated indices is implied). The spin-1/2 Hubbard model corresponds to the case $N = 2$. This Lagrangian is invariant only under the discrete chiral symmetry $\bar{\psi}_a\psi_a \mapsto -\bar{\psi}_a\psi_a$. This is a discrete, \mathbb{Z}_2 , symmetry and as such it can be spontaneously broken in 1+1 dimensions. For $N \geq 2$, the resulting state has a (dynamically) broken \mathbb{Z}_2 chiral symmetry and that there is a chiral condensate $\langle \bar{\psi}_a\psi_a \rangle \neq 0$ corresponding to a period 2 CDW.

The other theory is known as the chiral Gross-Neveu model whose Lagrangian density is

$$\mathcal{L}_{cGN} = \bar{\psi}_a i\partial\psi_a + g((\bar{\psi}_a\psi_a)^2 - (\bar{\psi}_a\gamma_5\psi_a)^2) \quad (158)$$

which has the full continuous chiral symmetry. For $N = 1$ this theory is equivalent to the Luttinger model (and to the Gross-Neveu model if we ignore the umklapp term). For $N \geq 2$ the chiral symmetry is formally broken. If this were true this theory would violate the Mermin-Wagner theorem. However a detailed study (most easily done using bosonization methods) shows that instead of long range order the correlator of both mass terms decay as a power law as a function of distance, consistent with the requirements by this theorem.

We close this subsection by noting that if the lattice model is not half filled but its density is either incommensurate or has a higher degree of commensurability, say p/q , the chiral symmetry is actually continuous (in the incommensurate case) or effectively continuous since the requisite umklapp terms are strongly irrelevant. However, if the lattice model is not at half filling charge conjugation symmetry C is broken at the lattice scale (in the UV), where it is equivalent to particle-hole symmetry, but it is recovered in the low-energy, IR, regime (up to irrelevant operators). In this sense, both the continuous chiral symmetry and charge conjugation symmetry can be regarded as emergent IR symmetries

7.3. The chiral anomaly

In section 7.2 we showed that the theory of massless Dirac fermions, in addition to a global U(1) gauge gauge symmetry, has a second conservation law which we called a global U(1) chiral

symmetry, shown in Eq.(131). This symmetry implies that there is a locally associated chiral current j_μ^5 , given by

$$j_\mu^5(x) = \bar{\psi}(x)\gamma_\mu\gamma_5\psi(x) \quad (159)$$

which also satisfies a continuity equation

$$\partial^\mu j_\mu^5 = 0 \quad (160)$$

and there is a globally conserved chiral charge Q^5

$$Q^5 = \int_{-\infty}^{\infty} dx j_0^5(x) = \int_{-\infty}^{\infty} dx (\psi_R^\dagger\psi_R - \psi_L^\dagger\psi_L) \quad (161)$$

It is easy to check that the Dirac current j_μ and the chiral current j_μ^5 are related by

$$j_\mu^5 = \epsilon_{\mu\nu} j^\mu \quad (162)$$

where $\epsilon_{\mu\nu}$ is the second rank Levi-Civita tensor.

The simultaneous conservation of both currents j_μ and j_μ^5 in the massless Dirac theory implies that the right and left moving densities j_R and j_L , defined in Eq.(156), should be separately conserved. In fact, if the Dirac theory has a mass term

$$\mathcal{L} = \bar{\psi}i\partial^\mu\psi - m\bar{\psi}\psi \quad (163)$$

it is straightforward to show that

$$\partial^\mu j_\mu^5 = 2m\bar{\psi}\gamma^5\psi \quad (164)$$

which means that in the massive theory the axial current is not conserved. This is easy to understand since the mass term mixes the right and left moving components of the Dirac spinor and, hence, the right and left moving densities are not conserved.

What happens in the *massless* limit, $m \rightarrow 0$, is more subtle. This problem was investigated in the late 1960's in 3+1 dimensions by S. Adler [133] and by J. S. Bell and R. Jackiw [134] who were interested in the anomalous decay of a neutral pion into two photons, $\pi^0 \rightarrow 2\gamma$. This process appears at third order of perturbation theory and it involves the computation of a triangle diagram (a fermion loop). In 3+1 dimensions this process has a UV divergence which needed to be regulated. These authors showed that it is not possible to find a regularization in which both the Dirac (gauge) current $j_\mu = \bar{\psi}\gamma_\mu\psi$ and the axial current $j_\mu^5 = \bar{\psi}\gamma_\mu\gamma_5\psi$ are conserved. In other words, if gauge invariance is preserved then the axial current is not and has an *anomaly* and results in a non-conservation of the axial current, $\partial^\mu j_\mu^5 \neq 0$. On the other hand, at least in the case of the physical gauge-invariant regularization, the obtained expression for the anomaly in the axial current is *universal*, independent of the value of the UV regulator (the cutoff). Sometime later G. 't Hooft showed that in non-abelian gauge theories instant processes also lead to anomalies and, furthermore, the result was also universal [96, 111]. Since the result is universal and, hence, independent of the UV scale, this led to the concept of anomaly matching conditions.

We will examine this problem in 1+1 dimensions (although it plays a key role in the theory of topological insulators in three space dimensions [135]). As we noted above, the lattice model is gauge-invariant and has only one conserved current. The conserved axial current appeared only in the low-energy regime in which the lattice model is described by a theory of massless Dirac fermions. To understand this problem we will consider the theory of fermions in 1+1 dimensions coupled to a U(1) gauge field. In the presence of a background (i.e. not quantized) electromagnetic field in the $A_0 = 0$ gauge the free fermion Hamiltonian of Eq.(116) becomes

$$H[A] = -t \sum_{n=1}^L c^\dagger(n)e^{ieA(n,t)/\hbar}c(n) + \text{h.c.} \quad (165)$$

An uniform and constant electric field is E is represented by a vector potential $A = -cEt$ (with c being the speed of light). The net effect of this gauge field is to shift of the momentum of the fermion quasiparticles $p \rightarrow p + ecEt/\hbar$ or, what is the same, to displace the fermion dispersion relation in momentum by the $ecEt/\hbar$. This means that the Fermi points are also shifted by that amount, $p_F \rightarrow p_F + ecEt/\hbar$ and $-p_F \rightarrow -p_F + ecEt/\hbar$. This means that the single particle states between p_F and $p_F + ecEt/\hbar$ that were empty for $E = 0$ are now occupied and the states between

$-p_F$ and $-p_F + ecEt/\hbar$ that were occupied for $E = 0$ are now empty. This means that number of right-moving fermions is increasing at a rate of ecE/\hbar and that the number of left-moving fermions decreases at the same rate. This results in a net current. Throughout this process the total number of fermions is not changed, gauge invariance is satisfied, but the number of right and left moving fermions are not separately concerned. Notice that this is an effect that involved the entire Fermi sea but the net effect is at low energies.

Let us see now how this plays out in the effective Dirac theory. The massless Dirac Lagrangian density in the background of an unquantized electromagnetic field A_μ is

$$\mathcal{L} = \bar{\psi}(i\partial - eA)\psi \quad (166)$$

Since there is no mass term the Dirac equation still decouples into two equations, for the right and left moving components of the Dirac spinor. In the $A_0 = 0$ gauge they are

$$i\partial_0\psi_R = (-i\partial_1 - A^1)\psi_R, \quad i\partial_0\psi_L = (i\partial_1 - A^1)\psi_L \quad (167)$$

In the temporal gauge, $A_0 = 0$, a uniform electric field $E = \partial_0 A^1$, and A^1 increases linearly with time. As A^1 increases, the Fermi momentum p_F (which is equal to the Fermi energy E_F) also increases at the rate eE . The density of states of a system of length L is $L/(2\pi)$. So, the rate of change of the number of right-moving fermions is

$$\frac{dN_R}{dx_0} = \frac{e}{2\pi}E \quad (168)$$

where we defined $N_R = \int_0^L dx j_R$ and similarly for N_L . If the UV regulator of the theory is compatible with gauge invariance, then the total fermion number must be conserved and the total vacuum charge must remain equal to zero,

$$Q = \int_0^L dx j_0(x) = N_R + N_L = 0 \quad (169)$$

Thus, if N_R increases, then N_L must decrease by the same amount. Or, equivalently, the electric field E creates an equal number of particles N_R and of antiparticles $\bar{N}_L = -N_L$.

On the other hand, the chiral charge $Q^5 = N_R - N_L$ must increase at the rate

$$\frac{dQ^5}{dx_0} = \frac{dN_R}{dx_0} + \frac{d\bar{N}_L}{dx_0} = \frac{e}{\pi}E \quad (170)$$

Again, the details of the UV regularization do not matter, only that it is gauge-invariant. We can also interpret Eq.(170) as the rate of particle-antiparticle pair creation by an electric field.

In a relativistic notation these results are expressed as

$$\partial^\mu j_\mu^5 = \frac{e}{2\pi}\epsilon_{\mu\nu}F^{\mu\nu} \quad (171)$$

Hence, the formally conserved current j_μ^5 has an anomaly and is not conserved due to quantum effects. Since it is not conserved, we cannot gauge the chiral symmetry. In the next subsection we will see that the the anomaly is closely related with bosonization.

7.4. Bosonization, anomalies and duality

We will reexamine the problem at hand from the point of view of the fermionic currents of the Dirac theory j_μ as operators. Since the currents obey the continuity equation, $\partial_\mu j^\mu = 0$ one expects that this would imply that it may be possible to write them as a curl of a scalar field, i.e. $j_\mu(x) = \epsilon_{\mu\nu}\partial^\nu\phi(x)$ where $\phi(x)$ should be a scalar field. Since this should be an operator identity we will need to understand how the currents act on the physical Hilbert space.

The Fermi-Bose mapping in 1D systems is closely related to the chiral anomaly we just discussed. Bosonization of a system of 1+1 dimensional massless Dirac fermions is a set of operator identities understood as matrix elements of the observables in the physical Hilbert space. These identities were first derived by Daniel Mattis and Elliott Lieb [128], based on earlier work by Julian Schwinger [136]. These identities were rediscovered (and their scope greatly expanded) by Alan Luther and Victor Emery [129], by Sidney Coleman [131], by Stanley Mandelstam [130], and by Edward Witten [137]. A non-abelian version of bosonization was subsequently derived by Witten [119].

The physical Hilbert space is defined as follows. Let $|FS\rangle \equiv |0\rangle$ denote the *filled Fermi sea*. In what follows we will assume that the physical system is macroscopically large and that local operators create the physical states by acting *finitely* on the filled Fermi sea. Physical observables, such as the right and left moving densities $j_R(x)$ and $j_L(x)$, need to be *normal-ordered* with respect to the physical vacuum state, the filled Fermi (Dirac) sea $|0\rangle$. The normal ordered densities are : $j_R(x) : \equiv j_R(x) - \langle 0|j_R(x)|0\rangle$ and : $j_L(x) : \equiv j_L(x) - \langle 0|j_L(x)|0\rangle$. Since the densities are products of fermion operators they need to be defined as a limit in which the operators are separated by a short distance η . Crucial to this construction is that the computation of the expectation values be regularized in such a way that the charge (gauge) current $j_\mu(x)$ is locally conserved and satisfies the continuity equation $\partial_\mu j^\mu = 0$ as an operator identity. In what follows all expectation values will refer to the filled Fermi sea state $|0\rangle$.

The propagators of the right and left moving Fermi fields are given by

$$\langle \psi_R^\dagger(x_0, x_1)\psi_R(0, 0) \rangle = \frac{-i}{2\pi(x_0 - x_1 + i\epsilon)}, \quad \langle \psi_L^\dagger(x_0, x_1)\psi_L(0, 0) \rangle = \frac{i}{2\pi(x_0 + x_1 + i\epsilon)} \quad (172)$$

The expectation value of the currents at a space location x_1 are

$$\langle j_R(x_1) \rangle = \lim_{\eta \rightarrow 0} \langle \psi_R^\dagger(x_1 + \eta)\psi_R(x_1 - \eta) \rangle = \frac{i}{4\pi\eta}, \quad \langle j_L(x_1) \rangle = \lim_{\eta \rightarrow 0} \langle \psi_L^\dagger(x_1 + \eta)\psi_L(x_1 - \eta) \rangle = \frac{-i}{4\pi\eta} \quad (173)$$

which are divergent at short distances. It follows that the normal-ordered right and left moving current densities satisfy the equal-time commutation relations

$$[j_R(x_1), j_R(x'_1)] = -\frac{i}{2\pi}\partial_1\delta(x_1 - x'_1), \quad [j_L(x_1), j_L(x'_1)] = +\frac{i}{2\pi}\partial_1\delta(x_1 - x'_1) \quad (174)$$

These identities imply that the normal-ordered space-time components of the current $j_\mu = (j_0, j_1)$ satisfy the equal-time commutation relations

$$[j_0(x_1), j_1(x'_1)] = -\frac{i}{\pi}\partial_1\delta(x_1 - x'_1), \quad [j_0(x_1), j_0(x'_1)] = [j_1(x_1), j_1(x'_1)] = 0 \quad (175)$$

The non-vanishing right-hand sides of these commutators are known as Schwinger terms. These identities define the the U(1) (Kac-Moody) current algebra.

We should note that in a theory of *non-relativistic* Fermi fields $\psi(\mathbf{x}, t)$ in all dimensions, the charge density $\rho(\mathbf{x})$ and the current operators $\mathbf{j}(\mathbf{x}) = \frac{1}{2i}(\psi^\dagger(\mathbf{x})\nabla\psi(\mathbf{x}) - \nabla\psi^\dagger(\mathbf{x})\nabla\psi(\mathbf{x}))$ satisfy a similar expression (also at equal times)

$$[\rho(\mathbf{x}), j_k(\mathbf{x}')] = -i\frac{e^2}{mc^2}\partial_k(\delta(\mathbf{x} - \mathbf{x}')\rho(\mathbf{x})), \quad [\rho(\mathbf{x}), \rho(\mathbf{x}')] = 0, \quad [j_k(\mathbf{x}), j_l(\mathbf{x}')] = 0 \quad (176)$$

In one dimension, and in the regime in which the fermions have a macroscopic density so that $\rho(x) \simeq \langle \rho(x) \rangle$, after a multiplicative rescaling of the operators, the non-relativistic identities of Eq.(176) are equivalent to the U(1) current algebra of Eq.(175).

The U(1) current algebra of Eq.(175) is reminiscent of the equal-time canonical commutation relations of a scalar field. Indeed, if $\phi(x)$ is a scalar field and $\Pi(x) = \partial_0\phi(x)$ is its canonically conjugate momentum, then they obey the equal-time canonical commutation relations

$$[\phi(x_1), \Pi(x'_1)] = i\delta(x_1 - x'_1) \quad (177)$$

We can then identify the charge density j_0 and the current density j_1 with the scalar field operators

$$j_0(x) = \frac{1}{\sqrt{\pi}}\partial_1\phi(x), \quad j_1(x) = -\frac{1}{\sqrt{\pi}}\Pi(x) = -\frac{1}{\sqrt{\pi}}\partial_0\phi(x) \quad (178)$$

which obey the U(1) current algebra of Eq.(175) as a consequence of the canonical commutation relations, Eq. (177). Furthermore, we can rewrite Eq.(177) in the Lorentz covariant form

$$j_\mu(x) = \frac{1}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu\phi(x) \quad (179)$$

which is clearly consistent with the local conservation of the current j_μ ,

$$\partial_\mu j^\mu(x) = 0 \quad (180)$$

Let us examine now the question of the conservation of the chiral current j_μ^5 . In Eq.(162) we showed that the gauge current and the chiral current are related by $j_\mu^5 = \epsilon_{\mu\nu} j^\nu$. Therefore the divergence of the chiral current is

$$\partial^\mu j_\mu^5 = \epsilon_{\mu\nu} \partial^\mu j^\nu = \frac{1}{\sqrt{\pi}} \epsilon_{\mu\nu} \epsilon^{\nu\lambda} \partial^\mu \partial_\lambda \phi = -\frac{1}{\sqrt{\pi}} \partial^2 \phi \quad (181)$$

where we used the identification of the gauge current in terms of the scalar field ϕ , Eq.(179). Therefore we conclude that

$$\partial^\mu j_\mu^5 = 0 \Leftrightarrow \partial^2 \phi = 0 \quad (182)$$

This equation states that the chiral current as an operator identity is conserved if and only if the field ϕ is a free massless scalar field, whose Lagrangian density is

$$\mathcal{L}_B = \frac{1}{2} (\partial_\mu \phi)^2 \quad (183)$$

In Eq.(166) we considered the free massless Dirac Lagrangian coupled to a background (not quantized) gauge field A_μ through the usual minimal coupling which here is $\mathcal{L}_{\text{int}} = -e j_\mu A^\mu$. using the bosonization identity for the gauge current, Eq.(179) we see that the Lagrangian density of the bosonized theory now becomes

$$\begin{aligned} \mathcal{L}_B[A] &= \frac{1}{2} (\partial_\mu \phi)^2 - \frac{e}{\sqrt{\pi}} \epsilon_{\mu\nu} \partial^\nu \phi(x) A^\mu(x) \\ &\equiv \frac{1}{2} (\partial_\mu \phi)^2 + J(x) \phi(x) \end{aligned} \quad (184)$$

where the source $J(x)$ is

$$J(x) = \frac{e}{\sqrt{\pi}} \epsilon_{\mu\nu} \partial^\nu A^\mu(x) = \frac{e}{\sqrt{4\pi}} F^*(x) \quad (185)$$

where $F^*(x) = \epsilon_{\mu\nu} F^{\nu\mu}(x)$ is the (Hodge) dual of the field strength $F_{\mu\nu}$. Hence, F^* is (essentially) the source for the scalar field $\phi(x)$. This implies that the equation of motion of the scalar field must be

$$-\partial^2 \phi(x) = J(x) = \frac{e}{\sqrt{\pi}} \epsilon_{\mu\nu} \partial^\nu A^\mu \quad (186)$$

Retracing our steps we find that the chiral current j_μ^5 obeys

$$\partial^\mu j_\mu^5 = -\frac{1}{\sqrt{\pi}} \partial^2 \phi = \frac{e}{2\pi} \epsilon_{\mu\nu} F^{\mu\nu} \quad (187)$$

which reproduces the the chiral anomaly given in Eq.(170).

These results suggest that the theory of a free massless Dirac spinor must be equivalent to the theory of the free massless scalar field. This statement is known as bosonization. However for this identification to be correct there must be an identification of the Hilbert spaces and of all the operators of each theory. We will not do this detailed analysis here but we will highlight the most significant statements.

Let us begin with the fermion number of the Dirac theory. Consider a system of fermions of total length L . Using the bosonization identity of Eq.(179) we find that the fermion number $N_F \equiv Q$ is given by

$$N_F = \int_0^L dx_1 j_0(x_0, x_1) = \frac{1}{\sqrt{\pi}} \int_0^L dx_1 \partial_1 \phi(x_0, x_1) = \frac{1}{\sqrt{\pi}} (\phi(x_0, L) - \phi(x_0, 0)) \quad (188)$$

Thus, the vacuum sector of the Dirac theory, with $N_F = 0$, corresponds to the theory of the scalar field with periodic boundary conditions, $\phi(x_1 = 0) = \phi(x_1 = L)$. Furthermore, since the fermion number is quantized, $N_F \in \mathbb{Z}$, changing the fermion number is the same as twisting the boundary conditions of the scalar so that

$$\phi(x_1 + L) = \phi(x_1) + \sqrt{\pi} N_F \quad (189)$$

In String Theory [104] the scalar field is interpreted as the coordinate of a string. Compactifying the space where the string lives to be a circle of radius R means that the string coordinate is defined modulo $2\pi R n$, where n is an integer. We see that the condition imposed by Eq.(189)

is equivalent to say that the scalar field is *compactified* and that the compactification radius is $R = 1/\sqrt{4\pi}$. This identification also imposes the restriction that the allowed operators of the scalar field must obey the identification

$$\phi(x) \sim \phi(x) + 2\pi R n \quad (190)$$

as an equivalency condition.

The simplest bosonic operators that obey the compactification condition are the vertex operators $V_\alpha(x)$,

$$V_\alpha(x) = \exp(i\alpha\phi(x)) \quad (191)$$

The compactification condition then requires that the allowed vertex operators should have $\alpha = n/R = \sqrt{4\pi}n$, where n is an integer. Since the propagator of the scalar field in 1+1-dimensional (Euclidean) spacetime is

$$G(x - x') = -\frac{1}{2\pi} \ln\left(\frac{|x - x'|}{a}\right) \quad (192)$$

where a is a short-distance cutoff, we find that the scaling dimension of the vertex operator is $\Delta_\alpha = \alpha^2/(4\pi) = n^2$. We will see shortly that the vertex operator with $\alpha = \sqrt{4\pi}$ is essentially the Dirac mass operator (which has scaling dimension 1).

The free massless scalar field can be decomposed into right and left moving components, ϕ_R and ϕ_L respectively,

$$\phi = \phi_R + \phi_L, \quad \vartheta = -\phi_R + \phi_L \quad (193)$$

where

$$\vartheta(x_0, x_1) = \int_{-\infty}^{x_1} dx'_1 \Pi(x_0, x'_1) \quad (194)$$

is the Cauchy-Riemann dual of the field $\phi(x)$ since they satisfy the Cauchy-Riemann equation

$$\partial_\mu \phi = \epsilon_{\mu\nu} \partial^\nu \vartheta \quad (195)$$

The right and left moving component of the Dirac spinor are found to have the bosonized expression [130]

$$\psi_R(x) = \frac{1}{\sqrt{2\pi a}} : \exp(i\sqrt{4\pi}\phi_R(x)) :, \quad \psi_L(x) = \frac{1}{\sqrt{2\pi a}} : \exp(-i\sqrt{4\pi}\phi_L(x)) : \quad (196)$$

It is easy to check that the propagators of these operators agree with the expressions given in Eq.(172), and that they have scaling dimension 1/2 and spin 1/2.

How does a chiral transformation act on the scalar field? A chiral transformation by an angle θ , c.f. Eq.(132), acts on the right and moving fermions as $\psi'_R = \exp(i\theta)\psi_R$ and $\psi'_L = \exp(-i\theta)\psi_L$. From Eq.(196) we see that the right and left moving components of the scalar field transform as $\phi'_R = \phi_R + \theta/\sqrt{4\pi}$ and $\phi'_L = \phi_L + \theta/\sqrt{4\pi}$. This means that a chiral transformation by an angle θ of the Dirac fermion by an angle θ is equivalent to a translation (a shift) of the scalar field $\phi' = \phi + 2\theta/\sqrt{4\pi}$.

We can use the Operator Product Expansion discussed in section 3.3.2 to show that the fermion mass terms $\bar{\psi}\psi$ and $i\bar{\psi}\gamma^5\psi$ are given by the following identifications

$$\bar{\psi}\psi = \frac{1}{2\pi a} : \cos(\sqrt{4\pi}\phi) :, \quad i\bar{\psi}\gamma^5\psi =: \sin(\sqrt{4\pi}\phi) : \quad (197)$$

These operators have scaling dimension 1 and transform properly under chiral transformations. These identifications imply that a theory of free *massive* Dirac fermions

$$\mathcal{L}_D = \bar{\psi}i\cancel{\partial}\psi - m\bar{\psi}\psi \quad (198)$$

is equivalent to the sine-Gordon field theory [131] whose Lagrangian is

$$\mathcal{L}_{SG} = \frac{1}{2}(\partial_\mu\phi)^2 - g : \cos(\sqrt{4\pi}\phi) : \quad (199)$$

where $g = m/(2\pi a)$.

Given the central role played by the current algebra identities of Eqs. (175) one may wonder if a similar approach might apply in higher dimensions. Schwinger terms in current algebra play an important role in relativistic field theories. However in higher dimensions their structure is

more complex and does not lead to identities of the type we have discussed. The reason at the root of this problem is largely kinematical. The Bose (scalar) field ϕ is qualitatively a bound state, a collective mode in the language of Condensed Matter Physics. In 1+1 dimensions this collective mode exhausts the spectrum at low energies due to the strong kinematical restriction on one spatial dimension.

The equivalency between the theory of free massive Dirac fermions and the sine-Gordon theory is an example of the power of bosonization. On the Dirac side the mapping the theory is free and its spectrum is well understood. But on the sine-Gordon side the theory is non-linear. In fact in the sine-Gordon theory the fermions are essentially solitons, domain walls of the scalar field. For these and many other reasons that we do not have space here bosonization plays a huge role in understanding the non-perturbative behavior of systems both in Condensed Matter Physics and in Quantum Field Theory in 1+1 dimensions. We will see in section 11.3.1 that to an extent some of these ideas can and have been extended to relativistic systems and classical statistical mechanical systems in 2+1 dimensions.

8. Fractional Charge

8.1. Solitons in one dimensions

We begin by returning to the equivalency between the theory of free massive Dirac fermions and sine-Gordon theory, in 1+1 dimensions. In Eq.(188) we showed that the boundary conditions of the compactified scalar field $\phi(x)$ are determined by the fermion number N_F of the dual Dirac theory, and that the vacuum sector of the Dirac theory maps onto the sine-Gordon theory with periodic boundary conditions. We will now examine the sector with one fermion, $N_F = 1$. This sector of the Dirac theory maps onto the sine-Gordon theory with *twisted* boundary conditions,

$$\phi(L) - \phi(0) = \sqrt{\pi} \quad (200)$$

The Hamiltonian of the sine-Gordon theory is

$$H_{\text{SG}} = \int_{-\infty}^{\infty} dx \left[\frac{1}{2} \Pi^2(x) + \frac{1}{2} (\partial_x \phi(x)) + g \cos(\sqrt{4\pi} \phi(x)) \right] \quad (201)$$

In the sector with periodic boundary conditions the classical ground states are static and uniform configurations that minimize the potential energy. Since the potential energy is a periodic functional of $\phi(x)$ the classical minima are at $\phi_n(x) = (n + 1/2) \sqrt{\pi}$, where n is an arbitrary integer. The classical energy of these ground states is extensive and is given by $E_{\text{gnd}} = -gL$ where L is the linear size of the system.

The classical ground state in the twisted sector is a *domain wall* (or *soliton*) which interpolates between the static and uniform ground states $\phi(x) = \pm \sqrt{\pi}/2$. The classical ground state in this sector is the static solution of the Euler-Lagrange equation

$$\frac{d^2\phi}{dx^2} = -2g \sqrt{\pi} \sin(2 \sqrt{\pi} \phi(x)) \quad (202)$$

such that asymptotically satisfies $\lim_{x \rightarrow \pm\infty} \phi(x) = \pm \sqrt{\pi}/2$. The solution is the classical soliton configuration

$$\phi(x) = \frac{2}{\sqrt{\pi}} \tan^{-1} \left(\exp(2 \sqrt{\pi} g (x - x_0)) \right) - \frac{\sqrt{\pi}}{2} \quad (203)$$

The soliton solution represents a *domain wall* between two symmetry-related classical ground states with $\phi = \pm \sqrt{\pi}/2$.

The energy of the soliton (measured from the energy of the ground state in the trivial sector) is finite and is given by

$$E_{\text{soliton}} = 4 \sqrt{\frac{g}{\pi}} \quad (204)$$

where x_0 is a zero mode of the soliton solution and represents its coordinate. By coupling the bosonized theory to a weak electromagnetic field A_μ , as given in Eq.(184), it is easy to check that it has electric charge $-e$ and, in this sense represents the electron. Then the identities of Eq.(196) can be used that as a quantum state it is indeed a fermion.

8.2. Polyacetylene

In section 7.4 we saw that solitons of a scalar field can be regarded as being equivalent to electrons, fermions with charge $-e$. We will now see that in a theory of fermions coupled to a domain wall of a scalar field, the soliton carries fractional charge. This problem has been extensively studied in one-dimensional conductors such as polyacetylene, in particular by the work of Wu-Pei Su, J. Robert Schrieffer and Alan Heeger [138] and by Roman Jackiw and J. Robert Schrieffer [139]. In quantum field theory this problem was first discussed by Roman Jackiw and Claudio Rebbi [140] and by Jeffrey Goldstone and Frank Wilczek [141].

In section 7.1 we showed that the physics of lattice fermions in one dimension at low energies is well described by a theory of massless Dirac fermions. In a one-dimensional conductor, such as polyacetylene, the fermions couple to the lattice vibrations (phonons). Su, Schrieffer and Heeger (SSH) [138] proposed a simple model in which the electrons couple to the lattice vibrations through a modulation of the hopping amplitude between two consecutive sites n and $n+1$, instead of being a constant t , becomes $t_{n,n+1} = t - g(u_{n+1} - u_n)$, where u_n is the displacement of the ion (a CH group in polyacetylene) at site n from its classical equilibrium position and g is the electron-phonon coupling constant. In polyacetylene there number of electrons (which are spin-1/2 fermions) is equal to the number of sites of the lattice and the electronic band is half-filled. At half filling this simple band structure is invariant under a particle-hole transformation. If the coupling to the lattice vibrations is included this symmetry remains respected provided the displacements change sign $u_n \rightarrow -u_n$ for all lattice sites. In polyacetylene the lattice dimerizes (a process known as a Peierls distortion) and the discrete translation symmetry by one lattice spacing is spontaneously broken: the system becomes a period 2 CDW on the bonds of the lattice. The broken symmetry state is still invariant under a particle-hole transformation.

The effective field theory of this system is a theory of two Dirac spinors $\psi_{\alpha,\sigma}(x)$, where $\alpha = 1, 2$ denotes right and left-moving fermions, and $\sigma = \uparrow, \downarrow$ are the two spin polarizations. The Lagrangian density of this system is

$$\mathcal{L} = \bar{\psi}_\sigma i\partial^\mu \psi_\sigma - g\phi(x)\bar{\psi}_\sigma(x)\psi_\sigma(x) - \frac{1}{2}\phi(x)^2 \quad (205)$$

The real scalar field $\phi(x)$ represents the distortion field of the polyacetylene chain. Here we will assume that the chains has spontaneously distorted and we will regard the scalar field as static and classical. This is a good approximation since the masses of the CH complexes is much bigger than the electron mass. This continuum model is due to Takayama, Lin-Liu and Maki [142] and further developed by Campbell and Bishop [143, 144]. Many of the results fund in this (adiabatic) approximation remain qualitatively correct upon taking into account the quantum dynamics of the chain, even in the limit in which the ions are treated as being “light” (provided the spin of the fermions is taken into account) [145, 146].

In the field theory the discrete symmetry of displacements by one lattice spacing becomes the \mathbb{Z}_2 symmetry $\phi \rightarrow -\phi$. This is a symmetry of the electron phonon system once combined with the discrete chiral transformation $\psi \rightarrow \gamma_5 \psi$ under which $\bar{\psi}\psi \rightarrow -\bar{\psi}\psi$. The ground state is two fold degenerate $\pm\phi_0$ with

$$\phi_0 = \frac{2\Lambda v_F}{g} \exp\left(-\frac{\pi v_F}{g^2}\right) \quad (206)$$

and the Dirac fermion (the electron) has a exponentially small mass, $m = g\phi_0$.

8.3. Fractionally charged solitons

Jackiw and Rebbi showed that the 1+1-dimensional classical ϕ^4 theory has the following soliton solution which interpolates between the two classically ordered states at $\pm\phi_0$ [140]

$$\phi(x) = \phi_0 \tanh\left(\frac{x - x_0}{\xi}\right) \quad (207)$$

where ξ is the correlation length of ϕ^4 theory, and x_0 is the (arbitrary) location of the soliton. They further showed that, when coupled to a theory of massless relativistic fermions through a Yukawa coupling, as in Eq.(205), this soliton carries fractional charge. The argument goes as follows. The one-particle Dirac Hamiltonian for a Dirac fermion with a position-dependent mass $m(x)$ is

$$H = -i\sigma_1\partial_x + m(x)\sigma_3 = \begin{pmatrix} m(x) & -i\partial_x \\ -i\partial_x & -m(x) \end{pmatrix} \quad (208)$$

where $m(x) = g\phi(x)$, with $\phi(x)$ being the soliton solution of Eq.(207). This Hamiltonian is hermitian and real. Furthermore, this Hamiltonian anti-commutes with the Pauli matrix σ_2 . This implies that for every positive-energy state $|E\rangle$ with energy $+E$ there is a negative-energy eigenstate with energy $-E$ given by $\sigma_2|E\rangle$. Hence, the spectrum is particle-hole symmetric (or, what is the same, charge-conjugation invariant). In addition, and consistent with charge-conjugation symmetry, the Hamiltonian of Eq.(208) has state with $E = 0$, a zero-mode, with a normalizable spinor wave function

$$\psi_0(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix} \exp\left(-\text{sgn}(m) \int_0^x dx' m(x')\right) \quad (209)$$

which exists for an arbitrary function $m(x)$ which changes sign once at some location (which we took to be $x_0 = 0$). Jackiw and Rebbi further showed that the soliton (anti-soliton) carries fractional charge

$$Q = \mp \frac{e}{2} \quad (210)$$

This result follows from the spectral asymmetry identity of the density of states $\rho_S(E)$ in the presence of the soliton

$$Q = -\frac{e}{2} \int_0^\infty (\rho_S(E) - \rho_S(-E)) = -\frac{e}{2}\eta \quad (211)$$

where η is the spectral asymmetry of the Dirac operator in the soliton background, and it is known as the APS η -invariant of Atiyah-Patodi-Singer [147]. Given the one-to-one correspondence that exists between positive and negative energy states in the spectrum, the spectral asymmetry follows from the condition that the zero mode be half-filled, which is required by normal-ordering or, what is the same, by charge neutrality. Another way to understand this result is that adding (or removing) a fermion of charge $-e$ results in the creation of a soliton-antisoliton pair, with each topological excitation carrying half of the charge of the electron. In other words, in the dimerized phase the electron is fractionalized. In this analysis we ignored the spin of the electron. If we take it into account the spin degree of freedom the soliton is instead a boson with charge $\mp e$.

There is an alternative, complementary, way to think about the charge of the soliton. Goldstone and Wilczek [141] considered a theory in which the (massless) Dirac fermion is coupled to two real scalar fields, φ_1 and φ_2 , with Lagrangian

$$\mathcal{L} = \bar{\psi} i\partial^\mu \psi - g\varphi_1 \bar{\psi} \psi - ig\varphi_2 \bar{\psi} \gamma_5 \psi \equiv \bar{\psi} i\partial^\mu \psi - g|\varphi| \bar{\psi} \exp(i\theta\gamma_5) \psi \quad (212)$$

where $|\varphi|^2 = \varphi_1^2 + \varphi_2^2$ and $\theta = \tan^{-1}(\varphi_2/\varphi_1)$. They considered a soliton in which $g\varphi_1 = m$ is the constant (in space) Dirac mass and φ_2 winds slowly between two values $\pm\varphi_0$ for $x \rightarrow \pm\infty$. In this theory the one-particle Dirac Hamiltonian is

$$H = -i\sigma_1 \partial_x + g\varphi_1 \sigma_3 + g\varphi_2 \sigma_2 \quad (213)$$

which is hermitian and complex and, hence, it violates CP invariance.

A perturbative calculation of the induced (gauge-invariant) current j_μ , which is given by the triangle diagram of a fermion loop with two gauge field insertions and a coupling of the scalar fields, yields the result (with $a = 1, 2$)

$$\langle j_\mu(x) \rangle = \frac{1}{2\pi} \epsilon_{\mu\nu} \epsilon_{ab} \frac{\varphi_a \partial^\nu \varphi_b}{|\varphi|^2} = \frac{1}{2\pi} \epsilon_{\mu\nu} \partial^\nu \theta \quad (214)$$

which is locally conserved. Notice, however, that the induced axial current j_μ^5 is not conserved, $\partial^\mu \langle j_\mu^5 \rangle = -\frac{1}{2\pi} \partial^2 \theta \neq 0$ and, hence, this current is anomalous. We can now compute the total charge accumulated as the soliton is created adiabatically to be given by the Goldstone-Wilczek formula

$$Q = -e \frac{\Delta\theta}{2\pi} \quad (215)$$

where $\Delta\theta = \theta(+\infty) - \theta(-\infty)$. Since $\lim_{x \rightarrow \pm\infty} \varphi_2(x) = \pm\varphi_0$, we obtain the result

$$Q = -\frac{e}{\pi} \tan^{-1}\left(\frac{g\varphi_0}{m}\right) \quad (216)$$

In the limit $m = g\varphi \rightarrow 0$, where CP (or T) invariance is recovered, we get

$$\lim_{m \rightarrow 0} Q = -\frac{e}{2} \quad (217)$$

which is the Jackiw-Rebbi result for the fractional charge of a soliton of Eq.(210).

The results for the fractional charge of the soliton can also be derived using the bosonization identities of section 7.4. Indeed, the bosonized expression for the Lagrangian of Eq.(212) is

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi)^2 - \frac{g|\varphi|}{2\pi a} \cos(\sqrt{4\pi}\phi - \theta) \quad (218)$$

Deep in the phase in which the Bose field ϕ is massive, the non-linear term in Eq.(218) locks this field to the chiral angle θ , i.e.

$$\phi = \frac{1}{\sqrt{4\pi}}\theta \quad (219)$$

However, the bosonization identities also tell us that the gauge current j_μ is given by the curl of the scalar field ϕ . Therefore, in this state the current j_μ is

$$j_\mu = \frac{1}{\sqrt{\pi}} \epsilon_{\mu\nu} \partial^\nu \phi = \frac{1}{2\pi} \epsilon_{\mu\nu} \partial^\nu \theta \quad (220)$$

which is the same as the Goldstone-Wilczek result of Eq.(214). That these two seemingly different approaches yield the same result is not accidental as they both follow from the axial anomaly.

9. Fractional Statistics

A fundamental axiom of Quantum Mechanics is that identical particles are indistinguishable [148, 149]. In non-relativistic Quantum Mechanics this leads to the requirement that the quantum states of a system of identical particles must be eigenstates of the pairwise particle exchange operator. Since two exchanges are equivalent to the identity operation this implies that the states must be even or odd under pairwise exchanges. This result, in turn, implies that particles can be classified as either being bosons (whose states are invariant under pairwise exchanges) or fermions (whose states change sign under pairwise particle exchanges). A consequence is that bosons obey the the Bose-Einstein (and can condense into a single particle state) whereas fermions obey the Fermi-Dirac distribution and must obey the Pauli exclusion principle. It is an implicit assumption of this line of reasoning that all relevant states of a system of identical particles can be efficiently represented by a (suitably symmetrized or antisymmetrized) product state.

This classification is present at an even deeper level in (relativistic) Quantum Field Theory where locality, unitarity and Lorentz invariance require that the fields be classified as representations of the Lorentz group and obey the Spin-Statistics Theorem [150]. The Spin-Statistics Theorem is actually an *axiom* of local relativistic Quantum Field theory which requires that fields that transform with an integer spin representation of the Poincaré group (i.e. scalars, gauge fields, gravitational fields, etc) must be bosons while fields that transform with a half-integer spin representation (i.e. Dirac spinors, etc) must be fermions. This spin-statistics connection is intrinsic to the construction of String Theory [104].

Given these considerations there was a general consensus that fermions and bosons were the only possible types of statistics. Nevertheless several exceptions to this rule were known to exist. One is the construction of the magnetic monopole in 3+1 dimensional gauge theory by Tai-Tsun Wu and Chen-Ning Yang who showed that a scalar coupled to a Dirac magnetic monopole behaves as a Dirac spinor [151]. This was an early example of statistical transmutation by coupling a matter field to a non-trivial configuration of a gauge field. As we will note below, the construction of anyons (particles with fractional statistics) in 2+1 dimensions has a close parentage to the Wu-Yang example. Examples of statistical transmutation were known to exist in 1+1 dimensional theories where a system of hard-core bosons was shown to be equivalent to a theory of free fermions using the Jordan-Wigner transformation [22] which represents a fermion as a composite operator of a hard-core boson and an operator that creates a kink (or soliton). This construction also underlies the fermion-boson mapping in 1+1 dimensional field theories [128, 129, 131, 130] that we discussed in section 7.4. Finally, in the late 1970s it was found that 1+1 dimensional \mathbb{Z}_N spin systems harbor operators known as parafermions which obey the same algebra shortly afterwards found to be obeyed by anyons in 2+1 dimensions [152].

9.1. Basics of fractional statistics

Jon Magne Leinaas and Jan Myrheim wrote an insightful paper in 1977 in which they examined the structure of the configuration space of the histories of a system of N identical particles [153]. Using the Feynman path-integral approach, they showed that if the worldlines of the identical particles are not allowed to cross, then the configuration space is topologically non-trivial. Through a detailed analysis they showed that the three and higher dimensions under a pairwise particle exchange the states must be either even or odd and hence the particles are either bosons or fermions. Leinaas and Myrheim also showed that in one and two space dimensions the wave functions can change by a *phase*, nowadays known as the statistical angle. In retrospect this result could have been anticipated (but was not) in an earlier paper by Michael G. G. Laidlaw and Cécile Morette DeWitt [154] who did a similar analysis of the configuration space of identical particles in the Feynman path integral.

Frank Wilczek generalized the Aharonov-Bohm effect [155] to describe the quantum mechanics of composite objects made of electric charge and magnetic flux in two space dimensions [156]. Wilczek showed that composite objects made of a non-relativistic particle of charge q bound to a magnetic flux of (magnetic) charge Φ behaves as an object with fractional angular momentum $q\Phi/2\pi$. Here Φ is measured in units of the flux quantum 2π (in units in which $\hbar = c = e = 1$). Furthermore, in a subsequent paper Wilczek [157] showed that, upon an adiabatic process in which the two composites exchange positions without their worldlines coinciding, the wave function of two identical flux-charge composites changes by a phase factor $\exp(\pm iq\Phi)$. For instance if $\Phi = \pi$ (i.e. a half-flux quantum) the acquired phase is equal to $\exp(i\pi) = -1$. thus if the particle was a boson, the composite becomes a fermion and viceversa. Wilczek coined the term *anyon* to describe the behavior of an arbitrary charge-flux composite. Furthermore, this construction also implies that not only *fractional statistics* but also *fractional spin*, consistent with a generalization of the Spin-Statistics Theorem. In other terms, “flux-attachment” implies fractional statistics (and fractional spin). Clearly Wilczek’s construction gave an explicit physical grounding to the general arguments of the 1977 paper by Leinaas and Myrheim. It is worth note the close analogy between this construction in 2+1 dimensions and the Wu-Yang construction in 3+1 dimensions (whose consistency with the Spin-Statistics Theorem had been shown earlier on by Alfred Goldhaber [158]).

In this description the statistics of the composites (the anyons) enters in the form of complex weights (phases) given in terms of the *linking numbers* of the worldlines. Hence, the concept of fractional statistics is intimately related to the theory of knots and of the representations of the braid group. These concepts were originally introduced in physics to describe statistical transmutation in the theory of solitons [159] in the context of the Skyrme model [160]. Yong-Shi Wu [161] developed an explicit connection between the work of Leinaas and Myrheim and Wilczek’s work on anyons in terms of operations acting on the worldlines of the anyons and described by the Braid Group. Wu’s work and the somewhat early paper by Wilczek and Zee on the statistics of solitons [162] marked the definite entry of the theory of knots (and of the braid group) in physics in general and in condensed matter physics in particular. The classification of anyons in terms of representations of the braid group labeled by the fractional statistics (determined by linking numbers) as well as of fractional (or topological) spin (determined by the writhing number of the worldlines) leads to a rich set of physical consequences. As it will turn out, Wilczek’s anyons are described one-dimensional representations of the braid group. As we will see below, additional and intriguing (non-abelian) representations will also play a role.

9.2. What is a topological field theory

We will now consider a special class of gauge theories known as topological field theories. These theories often (but not always) arise as the low energy limit of more complex gauge theories. In general, one expects that at low energies the phase of a gauge theory be either confining or deconfined. While confining phases have (from really good reasons!) attracted much attention, deconfined phases are often regarded as trivial, in the sense that the general expectation is that their vacuum states be unique and the spectrum of low lying states is either massive or massless.

Let us consider a gauge theory whose action on a manifold \mathcal{M} with metric tensor tensor $g_{\mu\nu}(x)$ is

$$S = \int_{\mathcal{M}} d^D x \sqrt{g} \mathcal{L}(g, A_\mu) \quad (221)$$

At the classical level, the energy-momentum tensor $T^{\mu\nu}(x)$ is the linear response of the action to an infinitesimal change of the local metric,

$$T^{\mu\nu}(x) \equiv \frac{\delta S}{\delta g_{\mu\nu}(x)} \quad (222)$$

That a theory is topological means that depends only on the topology of the space in which is defined and, consequently, it is independent of the local properties that depend on the metric, e.g. distances, angles, etc. Therefore, at least at the classical level, the energy-momentum tensor of a topological field theory must vanish identically,

$$T^{\mu\nu} = 0 \quad (223)$$

In particular, if the theory is topological, the *energy* (or Hamiltonian) is also zero. Furthermore, if the theory is independent of the metric, it is invariant under *arbitrary* coordinate transformations. Thus, if the theory is a gauge theory, the expectation values of Wilson loops will be independent of the size and shape of the loops. Whether or not a theory of this type can be consistently defined at the quantum level is a subtle problem which we will briefly touch on below.

It turns out that, due to the non-local nature of the observables of a gauge theory, the low energy regime of a theory in its deconfined phase can have non-trivial global properties. In what follows, we will say that a gauge theory is topological if all local excitations are massive (and in fact we will send their mass gaps to infinity). The remaining Hilbert space of states is determined by global properties of the theory, including the topology of the manifold of their space-time. In several cases, the effective action of a topological field theory does not depend on the metric of the space-time, at least at the classical level. In all cases, the observables are non-local objects, Wilson loops and their generalization.

9.3. Chern-Simons Gauge Theory

Gauge theories play a key role in physics. In 2+1 dimensions it is possible to define a special gauge theory which is odd under time reversal invariance and parity: Chern-Simons gauge theory. Originally introduced in Quantum Field Theory in 1982 by Stanley Deser, Roman Jackiw and Stephen Templeton [163, 164], Chern-Simons gauge theory can be defined for any compact Lie group, as well as an extension of Einstein's gravity in 2+1 dimensions. In 1989 Edward Witten [165] showed that Chern-Simons gauge theory computes the expectation values of configurations of Wilson loops, regarded as the worldlines of heavy particles, in terms of a set of topological invariants known as the Jones polynomial that classify knots in three dimensions.

We will consider the simplest case, the U(1) Chern-Simons gauge theory. The Chern-Simons action for a U(1) gauge field A_μ in 2+1 dimension is

$$S[A] = \frac{k}{4\pi} \int_{\Omega} d^3x \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda + \int_{\Omega} d^3x J_\mu A^\mu \quad (224)$$

where J_μ is a set of conserved currents (representing the worldlines of a set of heavy particles). On a closed 3-manifold Ω (e.g. a sphere, a torus, etc) the Chern-Simons action is gauge invariant provided the parameter k (known as the level) is an integer. If the manifold Ω has a boundary, the action is not gauge invariant at the boundary. Gauge invariance is restored by additional boundary degrees of freedom. This structure is general, and not just a feature of the U(1) theory.

Since the Chern-Simons action is first order in derivatives it is not invariant under time reversal and under parity (which in 2+1 dimensions is a reflection). These symmetries make this theory relevant to the description of the fractional quantum Hall effect. In the absence of external sources, $J_\mu = 0$, at the classical level the Chern-Simons action is invariant under arbitrary changes of coordinates. This means that the theory is, at least classically, a topological field theory.

For a general non-abelian gauge group G the Chern-Simons action becomes

$$S = \frac{k}{4\pi} \int_M d^3x \text{tr}(AdA + \frac{2}{3}A \wedge A \wedge A) \quad (225)$$

Here, the cubic term is shorthand for

$$\text{tr}(A \wedge A \wedge A) \equiv \text{tr}(\epsilon_{\mu\nu\lambda} A^\mu A^\nu A^\lambda) \quad (226)$$

for a gauge field A^μ that takes values on the algebra of the gauge group G .

9.4. BF gauge theory

A closely related (abelian) gauge theory is the so-called *BF* theory [166] which, in a general spacetime dimension D (even or odd), is a theory of a vector field A^μ (a one-form) and an antisymmetric tensor field B with $D - 2$ Lorentz indices (a $D2$ form), known as a Kalb-Ramond field. In 2+1 dimensions the action of the BF theory is

$$S = \frac{k}{2\pi} \int_{\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} B^\lambda \partial^\mu A^\nu \quad (227)$$

where, once again, k is an integer.

The BF gauge theory has the same content as the topological sector of a discrete \mathbb{Z}_k gauge theory. To see this we will consider the theory of compact electrodynamics which is a U(1) gauge theory minimally coupled to charged scalar field ϕ . We will assume that the gauge theory is defined for a compact U(1) gauge group (meaning that the gauge flux is quantized) and that the complex scalar field has charge integer $k \in \mathbb{Z}$. As usual minimal coupling is implemented by introducing the covariant derivative $D_\mu = \partial_\mu + ikA_\mu$, where A_μ is the U(1) gauge field. Deep in the phase in which the global U(1) symmetry is spontaneously broken, usually called the Higgs regime, the amplitude of the scalar field is frozen at its (real) vacuum expectation value ϕ_0 but its phase ω , representing the Goldstone mode, is unconstrained. In this limit the Lagrangian of this theory becomes

$$\mathcal{L} = |\phi_0|^2 (\partial_\mu \omega - kA_\mu)^2 - \frac{1}{4e^2} F_{\mu\nu}^2 \quad (228)$$

where e is the coupling constant of the gauge field (the electric charge) and $F_{\mu\nu}$ is the field strength of the gauge field A_μ . In general spacetime dimension D the charge e has units of length $^{(D-4)/2}$. In this limit the gauge field becomes massive (this is the Higgs mechanism). This theory has fluxes quantized in units of $2\pi/k$ and has only k distinct fluxes. This is the \mathbb{Z}_k gauge theory.

Here we will consider this theory in 2+1 dimensions. We will use a gaussian (Hubbard-Stratonovich) decoupling of the first term of Eq.(228) in terms of a gauge field C_μ to write the Lagrangian in the equivalent form

$$\mathcal{L} = -\frac{1}{4|\phi_0|^2} C_\mu^2 + C^\mu (\partial_\mu \omega - kA_\mu) - \frac{1}{4e^2} F_{\mu\nu}^2 \quad (229)$$

Up to an integration by parts, we see that the phase field ω plays the role of a Lagrange multiplier field which forces the vector field B_μ to obey the constraint $\partial_\mu C^\mu = 0$. This constraint is solved by writing C_μ as

$$C_\mu = \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\nu B^\lambda \quad (230)$$

of a 1-form gauge field B_μ . Upon solving the constraint the Lagrangian of Eq.(229) becomes

$$\mathcal{L} = \frac{k}{2\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu B^\lambda - \frac{1}{4e^2} F_{\mu\nu}(A)^2 - \frac{1}{32\pi^2 |\phi_0|^2} F_{\mu\nu}(B)^2 \quad (231)$$

For spacetime dimensions $D < 4$ the IR the Maxwell terms for the fields A^μ and B_μ are irrelevant in the IR and, in this limit, this theory reduces to the BF theory at level k of Eq.(227). Therefore, \mathbb{Z}_k gauge theory is equivalent to the BF theory at level k . In fact, this result is essentially valid in all dimensions with the main difference being that the field B_μ is, in general, a rank $D - 2$ Kalb-Ramond antisymmetric field.

9.5. Quantization of Abelian Chern-Simons Gauge Theory

By expanding the action of Eq.(224) the Lagrangian density becomes

$$\mathcal{L} = \frac{k}{4\pi} \epsilon_{ij} A_i \partial_0 A_j + A_0 \left(\frac{k}{2\pi} B - J_0 \right) - J_i A_i \quad (232)$$

where $B = \epsilon_{ij} \partial_i A_j$ is the local flux, J_0 is a local classical density and J_i a local classical current. Then, the first term of Eq.(232) implies that the spatial components of the gauge field obey equal-time canonical commutation relations

$$[A_1(x), A_2(x')] = i \frac{2\pi}{k} \delta(x - x') \quad (233)$$

The second term of the Lagrangian enforces the Gauss Law which for this theory simply implies that the states in the physical Hilbert space obey the constraint

$$B(x) = \frac{2\pi}{k} J_0(x) \quad (234)$$

Thus the Gauss Law requires that a charge density necessarily has a magnetic flux attached to it. In other terms, the physical states are charge-flux composites as postulated in Wilczek's theory [157]. This is the theoretical basis to the concept of flux attachment which, as we will see in section 10.3, is widely used in the theory of the fractional quantum Hall effect.

The third term in Eq.(232) simply states that the Hamiltonian density is just

$$\mathcal{H} = J_i A_i \quad (235)$$

Hence, in the absence of sources the Hamiltonian vanishes, $\mathcal{H} = 0$.

The Chern-Simons action is locally gauge-invariant, up to boundary terms. To see this let us perform a gauge transformation, $A_\mu \rightarrow A_\mu + \partial_\mu \Phi$, where $\Phi(x)$ is a smooth, twice differentiable function. Then,

$$\begin{aligned} S[A^\mu + \partial^\mu \Phi] &= \int_{\mathcal{M}} (A^\mu + \partial^\mu \Phi) \epsilon_{\mu\nu\lambda} \partial^\nu (A^\lambda + \partial^\lambda \Phi) \\ &= \int_{\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda + \int_{\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} \partial^\mu \Phi \partial^\nu A^\lambda \end{aligned} \quad (236)$$

Therefore, the change is

$$S[A^\mu + \partial^\mu \Phi] - S[A^\mu] = \int_{\mathcal{M}} d^3x \partial^\mu \Phi F_\mu^* = \int_{\mathcal{M}} d^3x \partial^\mu (\Phi F_\mu^*) - \int_{\mathcal{M}} d^3x \Phi \partial^\mu F_\mu^* \quad (237)$$

where $F_\mu^* = \epsilon^{\mu\nu\lambda} \partial_\nu A_\lambda$ is the dual field strength. However, in the absence of magnetic monopoles, this field satisfies the Bianchi identity, $\partial^\mu F^*\mu = \partial^\mu (\epsilon_{\mu\nu\lambda} \partial^\nu A^\lambda) = 0$. Therefore, using the Gauss Theorem, we find that the change of the action is a total derivative and integrates to the boundary

$$\delta S = \int_{\mathcal{M}} d^3x \partial^\mu (\Phi F_\mu^*) = \int_{\Sigma} dS^\mu \Phi F_\mu^* \quad (238)$$

where $\Sigma = \partial \mathcal{M}$ is the boundary of \mathcal{M} . In particular, if Φ is a non-zero constant function on \mathcal{M} , then the change of the action under such a gauge transformation is

$$\delta S = \Phi \times \text{flux}(\Sigma) \quad (239)$$

Hence, the action is not invariant if the manifold has a boundary, and the theory must be supplied with additional degrees of freedom at the boundary.

Indeed, the flat connections, i.e. the solution of the equations of motion, $F_{\mu\nu} = 0$, are pure gauge transformations, $A_\mu = \partial_\mu \phi$, and have an action that integrates to the boundary. Let the $\mathcal{M} = D \times \mathbb{R}$ where D is a disk in space and \mathbb{R} is time. The boundary manifold is $\Sigma = S^1 \times \mathbb{R}$, where S^1 is a circle. Thus, in this case, the boundary manifold Σ is isomorphic to a cylinder. The action of the flat configurations reduces to

$$S = \int_{S^1 \times \mathbb{R}} d^2x \frac{k}{4\pi} \partial_0 \phi \partial_1 \phi \quad (240)$$

This implies that the dynamics on the boundary is that of a scalar field on a circle S^1 , and obeys periodic boundary conditions.

Although classically the theory does not depend on the metric, it is invariant under arbitrary transformations of the coordinates. However, any gauge fixing condition will automatically break this large symmetry. For instance, we can specify a gauge condition at the boundary in the form of a boundary term of the form $\mathcal{L}_{\text{gauge fixing}} = A_1^2$. In this case, the boundary action of the field φ becomes

$$S[\varphi] = \int_{S^1 \times \mathbb{R}} d^2x \frac{k}{4\pi} [\partial_0 \phi \partial_1 \phi - (\partial_1 \phi)^2] \quad (241)$$

The solutions of the equations of motion of this compactified scalar field have the form $\phi(x_1 \mp x_0)$, and are right (left) moving chiral fields depending of the sign of k . This boundary theory is not topological but is conformally invariant.

A similar result is found in non-abelian Chern-Simons gauge theory. In the case of the $SU(N)_k$ Chern-Simons theory on a manifold $D \times \mathbb{R}$, where D is a disk whose boundary is Γ , and \mathbb{R} is time, the action is

$$S_{\text{CS}}[A] = \int_{D \times \mathbb{R}} d^3x \left[\frac{k}{8\pi} \text{tr} \left(\epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda + \frac{2}{3} \epsilon^{\mu\nu\lambda} A_\mu A_\nu A_\lambda \right) \right] \quad (242)$$

This theory integrates to the boundary, $\Gamma \times \mathbb{R}$ where it becomes the chiral (right-moving) $SU(N)_k$ Wess-Zumino-Witten model (at level k) at its IR fixed point, $\lambda_c^2 = 4\pi/k$

$$S_{\text{WZW}}[g] = \frac{1}{4\lambda_c^2} \int_{\Gamma \times \mathbb{R}} d^2x \text{tr} \left(\partial_\mu g \partial^\mu g^{-1} \right) + \frac{k}{12\pi} \int_B \epsilon^{\mu\nu\lambda} \text{tr} \left(g^{-1} \partial_\mu g g^{-1} \partial_\nu g g^{-1} \partial_\lambda g \right) \quad (243)$$

Here, $g \in SU(N)$ parametrizes the flat configurations of the Chern-Simons gauge theory. The boundary theory is a non-trivial CFT, the chiral Wess-Zumino-Witten CFT [165].

9.6. Vacuum degeneracy a torus

We will now construct the quantum version of the U(1) Chern-Simons gauge theory on a manifold $\mathcal{M} = T^2 \times \mathbb{R}$, where T^2 is a spatial torus, of linear size L_1 and L_2 . Since this manifold does not have boundaries, the flat connections, $\epsilon_{ij} \partial_i A_j = 0$ do not reduce to local gauge transformations of the form $A_i = \partial_i \Phi$. Indeed, the holonomies of the torus T^2 , i.e. the Wilson loops on the two non-contractible cycles of the torus Γ_1 and Γ_2 are gauge-invariant observables:

$$\int_0^{L_1} dx_1 A_1 \equiv \bar{a}_1, \quad \int_0^{L_2} dx_1 A_2 \equiv \bar{a}_2 \quad (244)$$

where \bar{a}_1 and \bar{a}_2 are time-dependent. Thus, the flat connections now are

$$A_1 = \partial_1 \Phi + \frac{\bar{a}_1}{L_1}, \quad A_2 = \partial_2 \Phi + \frac{\bar{a}_2}{L_2} \quad (245)$$

whose action is

$$S = \frac{k}{4\pi} \int dx_0 \epsilon_{ij} \bar{a}_i \partial_0 \bar{a}_j \quad (246)$$

Therefore, the global degrees of freedom \bar{a}_1 and \bar{a}_2 at the quantum level become operators that satisfy the commutation relations

$$[\bar{a}_1, \bar{a}_2] = i \frac{2\pi}{k} \quad (247)$$

We find that the flat connections are described by the quantum mechanics of \bar{a}_1 and \bar{a}_2 . A representation of this algebra is

$$\bar{a}_2 \equiv -i \frac{2\pi}{k} \frac{\partial}{\partial \bar{a}_1} \quad (248)$$

Furthermore, the Wilson loops on the two cycles become

$$W[\Gamma_1] = \exp \left(i \int_0^{L_1} A_1 \right) \equiv e^{i\bar{a}_1}, \quad W[\Gamma_2] = \exp \left(i \int_0^{L_2} A_2 \right) \equiv e^{i\bar{a}_2} \quad (249)$$

and satisfy the algebra

$$W[\Gamma_1] W[\Gamma_2] = \exp(-i2\pi/k) W[\Gamma_2] W[\Gamma_1] \quad (250)$$

Under large gauge transformations

$$\bar{a}_1 \rightarrow \bar{a}_1 + 2\pi, \quad \bar{a}_2 \rightarrow \bar{a}_2 + 2\pi \quad (251)$$

Therefore, invariance under large gauge transformations on the torus implies that \bar{a}_1 and \bar{a}_2 define a two-torus target space.

Let us define the unitary operators

$$U_1 = \exp(i k \bar{a}_2), \quad U_2 = \exp(-i k \bar{a}_1) \quad (252)$$

which satisfy the algebra

$$U_1 U_2 = \exp(i 2\pi k) U_2 U_1 \quad (253)$$

The unitary transformations U_1 and U_2 act as shift operators on \bar{a}_1 and \bar{a}_2 by 2π , and hence generate the large gauge transformations. Moreover, the unitary operators U_1 and U_2 leave the Wilson loop operators on non-contractible cycles invariant,

$$U_1^{-1}W[\Gamma_1]U_1 = W[\Gamma_1], \quad U_2^{-1}W[\Gamma_2]U_2 = W[\Gamma_2] \quad (254)$$

Let $|0\rangle$ be the eigenstate of $W[\Gamma_1]$ with eigenvalue 1, i.e. $W[\Gamma_1]|0\rangle = |0\rangle$. The state $W[\Gamma_2]|0\rangle$ is also an eigenstate of $W[\Gamma_1]$ with eigenvalue $\exp(-i2\pi/k)$, since

$$W[\Gamma_1]W[\Gamma_2]|0\rangle = e^{i2\pi/k}W[\Gamma_2]W[\Gamma_1]|0\rangle = e^{-i2\pi/k}W[\Gamma_1]|0\rangle \quad (255)$$

More generally, since

$$W[\Gamma_1]W^p[\Gamma_2]|0\rangle = e^{-i2\pi p/k}W^p[\Gamma_2]|0\rangle \quad (256)$$

we find that, provided $k \in \mathbb{Z}$, there are k linearly independent vacuum states $|p\rangle = W^p[\Gamma_2]|0\rangle$, for the $U(1)$ Chern-Simons gauge theory at level k . It is denoted as the $U(1)_k$ Chern-Simons theory. Therefore the finite-dimensional topological space on a two-torus is k -dimensional. It is trivial to show that, on a surface of genus g , the degeneracy is k^g .

We see that in the abelian $U(1)_k$ Chern-Simons theory the Wilson loops must carry k possible values of the unit charge. This property generalizes to the non-abelian theories, which are technically more subtle. We will only state some important results. For example, if the gauge group is $SU(2)$ we expect that the Wilson loops will carry the representation labels of the group $SU(2)$, i.e. they will be labelled by (j, m) , where $j = 0, \frac{1}{2}, 1, \dots$ and the $2j+1$ values of m satisfy $|m| \leq j$. However, it turns out $SU(2)_k$ Chern-Simons theory has fewer states, and that the values of j are restricted to the range $j = 0, \frac{1}{2}, \dots, \frac{k}{2}$.

9.7. Fractional Statistics and Braids

Another aspect of the topological nature of Chern-Simons theory is the behavior of expectation values of products of Wilson loop operators. Let us compute the expectation value of a product of two Wilson loop operators on two positively oriented closed contours γ_1 and γ_2 . We will do this computation in the abelian Chern-Simons theory $U(1)_k$ in 2+1-dimensional Euclidean space. Note that the Euclidean Chern-Simons action is pure imaginary since the action is first-order in derivatives. The expectation value to be computed is

$$W[\gamma_1 \cup \gamma_2] = \left\langle \exp\left(i \oint_{\gamma_1 \cup \gamma_2} dx_\mu A_\mu\right) \right\rangle_{\text{CS}} \quad (257)$$

The result changes depending on whether the loops γ_1 and γ_2 are linked or unlinked. In this section we will compute the contribution to this expression for a pair of contours γ_1 and γ_2 . Here we will not include the contribution to this expectation value for each contours. We will return to this problem in section 11.3.1 where we discuss the problem of fractional spin.

The expectation value of a Wilson loop on the union of two contours, as in the present case, γ can be written as

$$\left\langle \exp\left(i \oint_{\gamma} dx_\mu A_\mu\right) \right\rangle_{\text{CS}} = \left\langle \exp\left(i \int d^3x J_\mu A_\mu\right) \right\rangle_{\text{CS}} \quad (258)$$

where the current J_μ is

$$J_\mu(x) = \delta(x_\mu - z_\mu(t)) \frac{dz_\mu}{dt} \quad (259)$$

Here $z_\mu(t)$ is a parametrization of the contour γ . Therefore, the expectation value of the Wilson loop is [165]

$$\left\langle \exp\left(i \oint_{\gamma} dx_\mu A_\mu\right) \right\rangle_{\text{CS}} \equiv \exp(iI[\gamma]_{\text{CS}}) = \exp\left(-\frac{i}{2} \int d^3x \int d^3y J_\mu(x) G_{\mu\nu}(x-y) J_\nu(y)\right) \quad (260)$$

where $G_{\mu\nu}(x-y) = \langle A_\mu(x)A_\nu(y) \rangle_{\text{CS}}$ is the propagator of the Chern-Simons gauge field. Since the loops are closed, the current J_μ is conserved, i.e. $\partial_\mu J_\mu = 0$, and the effective action $I[\gamma]_{\text{CS}}$ of the loop γ is gauge-invariant.

The Euclidean propagator of Chern-Simons gauge theory (in the Feynman gauge) is

$$G_{\mu\nu}(x-y) = \frac{2\pi}{k} G_0(x-y) \epsilon_{\mu\nu\lambda} \partial_\lambda \delta(x-y) \quad (261)$$

where $G_0(x - y)$ is the propagator of the massless Euclidean scalar field, which satisfies

$$-\partial^2 G_0(x - y) = \delta^3(x - y) \quad (262)$$

Using these results, we find the following expression for the effective action

$$\begin{aligned} I[\gamma]_{\text{CS}} &= \frac{\pi}{k} \int d^3x \int d^3y J_\mu(x) J_\nu(y) G_0(x - y) \epsilon_{\mu\nu\lambda} \partial_\lambda \delta(x - y) \\ &= \frac{\pi}{k} \oint_\gamma dx_\mu \oint_\gamma dy_\nu \epsilon_{\mu\nu\lambda} \partial_\lambda G_0(x - y) \end{aligned} \quad (263)$$

Since the current J_μ is conserved, it can be written as the curl of a vector field, B_μ , as

$$J_\mu = \epsilon_{\mu\nu\lambda} \partial_\nu B_\lambda \quad (264)$$

In the Lorentz gauge, $\partial_\mu B_\mu = 0$, we can write

$$B_\mu = \epsilon_{\mu\nu\lambda} \partial_\nu \phi_\lambda \quad (265)$$

Hence,

$$J_\mu = -\partial^2 \phi_\mu \quad (266)$$

where

$$\phi_\mu(x) = \int d^3y G_0(x - y) J_\mu(y) \quad (267)$$

Upon substituting this result into the expression for B_μ , we find

$$B_\mu = \int d^3y \epsilon_{\mu\nu\lambda} \partial_\nu G_0(x - y) J_\lambda(y) = \oint_\gamma \epsilon_{\mu\nu\lambda} \partial_\nu G_0(x - y) dy_\lambda \quad (268)$$

Therefore, the effective action $I[\gamma]_{\text{CS}}$ becomes

$$I[\gamma]_{\text{CS}} = \frac{\pi}{k} \oint_\gamma dx_\mu \oint_\gamma dy_\nu \epsilon_{\mu\nu\lambda} \partial_\lambda G_0(x - y) = \frac{\pi}{k} \oint_\gamma dx_\mu B_\mu(x) \quad (269)$$

Let Σ be an oriented open surface of the Euclidean three-dimensional space whose boundary is the oriented loop (or union of loops) γ , i.e. $\partial\Sigma = \gamma$. Then, using Stokes Theorem we write in the last line of Eq.(269) as

$$I[\gamma]_{\text{CS}} = \frac{\pi}{k} \int_\Sigma dS_\mu \epsilon_{\mu\nu\lambda} \partial_\nu B_\lambda = \frac{\pi}{k} \int_\Sigma dS_\mu J_\mu \quad (270)$$

The integral in the last line of this equation is the flux of the current J_μ through the surface Σ . Therefore, this integral counts the number of times n_γ the Wilson loop on γ pierces the surface Σ (whose boundary is γ), and therefore it is an integer, $n_\gamma \in \mathbb{Z}$. We will call this integer the *linking number* (or Gauss invariant) of the configuration of loops. In other words, the expectation value of the Wilson loop operator is

$$W[\gamma]_{\text{CS}} = e^{i\pi n_\gamma/k} \quad (271)$$

The linking number is a *topological invariant* since, being an integer, its value cannot be changed by smooth deformations of the loops, provided they are not allowed to cross.

We will now see that this property of Wilson loops in Chern-Simons gauge theory leads to the concept of fractional statistics. Let us consider a scalar matter field that is massive and charged under the Chern-Simons gauge field. The excitations of this matter field are particles that couple minimally to the gauge field. Here we will be interested in the case in which these particles are very heavy. In that limit, we can focus on states that have a few of these particles which will be in their non-relativistic regime.

Consider, for example, a state with two particles which in the remote past, at time $t = -T \rightarrow -\infty$, are located at two points A and B . This initial state will evolve to a final state at time $t = T \rightarrow \infty$, in which the particles either go back to their initial locations (the direct process), or to another one in which they exchange places, $A \leftrightarrow B$. At intermediate times, the particles follow smooth worldlines. These two processes, direct and exchange. There we see that the direct process is equivalent to a history with two unlinked loops (the worldlines of the particles), whereas in the exchange process the two loops form a link. It follows from the preceding discussion that the

two amplitudes differ by the result of the computation of the Wilson loop expectation value for the loops γ_1 and γ_2 . Let us call the first amplitude W_{direct} and the second W_{exchange} . The result is

$$W_{\text{exchange}} = W_{\text{direct}} e^{\pm i\pi/k} \quad (272)$$

where the sign depends on how the two worldlines wind around each other.

An equivalent interpretation of this result is that if $\Psi[A, B]$ is the wave function with the two particles at locations A and B , the wavefunction where their locations are exchanged is

$$\Psi[B, A] = e^{\pm i\pi/k} \Psi[A, B] \quad (273)$$

where the sign depends on whether the exchange is done counterclockwise or clockwise. Clearly, for $k = 1$ the wave function is antisymmetric and the particles are fermions, while for $k \rightarrow \infty$ they are bosons. At other values of k the particles obey *fractional statistics* and are called *anyons* [156, 153]. The phase factor $\phi = \pm\pi/k$ is called the statistical phase.

Notice that, while for fermions and bosons the statistical phase $\varphi = 0, \pi$ is uniquely defined ($\text{mod } 2\pi$), for other values of k the statistical angle is specified up to a sign that specifies how the worldlines wind around each other. Indeed, mathematically the exchange process is known as a *braid*. Processes in which the worldlines wind clock and counterclockwise are braids that are inverse of each other. Braids can also be stack sequentially yielding multiples of the phase φ . In addition to stacking braids, Wilson loops can be fused: seen from some distance, a pair of particles will behave as a new particle with a well defined behavior under braiding. This process of fusion is closely related to the concept of fusion of primary fields in Conformal Field Theory.

Furthermore, up to regularization subtleties [167], the self-linking terms (those with $a = b$) yield a topological spin $1/2k$, consistent with the spin-statistics connection [168]. For $k = 1$ this means that the flux-charge composites have spin $1/2$.

What we have just described is a mathematical structure called the *Braid group*. The example that we worked out using abelian Chern-Simons theory yields one-dimensional representations of the Braid group with the phase φ being the label of the representations. For $U(1)_k$ there are k types of particles (anyons). That these representations are abelian means that, in the general case of $U(1)_k$, acting on a one-dimensional representation p (defined mod k) with a one-dimensional representation q (also defined mod k) yields the representation one-dimensional $p+q$ (mod k). We will denote the operation of *fusing* these representations (particles!) as $[q]_{\text{mod } k} \times [p]_{\text{mod } k} = [q + p]_{\text{mod } k}$. These representations are in one-to-one correspondence with the inequivalent charges of the Wilson loops, and with the vacuum degeneracy of the $U(1)_k$ Chern-Simons theory on a torus.

A richer structure arises in the case of the non-abelian Chern-Simons theory at level k [165], such as $SU(2)_k$. For example, for $SU(2)_1$ the theory has only two representations, both are one-dimensional, and have statistical angles $\varphi = 0, \pi/2$.

However, for $SU(2)_k$, the content is more complex. In the case of $SU(2)_2$ the theory has a) a trivial representation $[0]$ (the identity, $(j, m) = (0, 0)$), b) a (spinor) representation $[1/2]$ ($(j, m) = (1/2, \pm 1/2)$), and c) a the representation $[1]$ ($(j, m) = (1, m)$, with $m = 0, \pm 1$). These states will fuse obeying the following rules: $[0] \times [0] = [0]$, $[0] \times [1/2] = [1/2]$, $[0] \times [1] = [1]$, $[1/2] \times [1/2] = [0] + [1]$, $[1/2] \times [1] = [1/2]$, and $[1] \times [1] = [0]$ (note the truncation of the fusion process!).

Of particular interest is the case $[1/2] \times [1/2] = [0] + [1]$. In this case we have two fusion channels, labeled by $[0]$ and $[1]$. The braiding operations now will act on a two-dimensional Hilbert space and are represented by 2×2 matrices. This is an example of a non-abelian representation of the braid group. These rather abstract concepts have found a physical manifestation in the physics of the fractional quantum Hall fluids, whose excitations are vortices that carry fractional charge and anyon (braid) fractional statistics.

Why this is interesting can be seen by considering a Chern-Simons gauge theory with four quasi-static Wilson loops. For instance in the case of the $SU(2)_2$ Chern-Simons theory the Wilson loops carry the spinor representation, $[1/2]$. If we call the four particles A, B, C and D , we would expect that their quantum state would be completely determined by the coordinates of the particles. This, however, is not the case since, if we fuse A with B , the result is either a state $[0]$ or a state $[1]$. Thus, if the particles were prepared originally in some state, braiding (and fusion) will lead to a linear superposition of the two states. This braiding process defines a unitary matrix, a representation of the Braid Group. The same is true with the other particles. However, it turns out that for four particles there are only two linearly independent states. This two-fold degenerate Hilbert space of topological origin is called a topological qubit.

Moreover, if we consider a system with N (even) number of such particles, the dimension of the topologically protected Hilbert space is $2^{\frac{N}{2}-1}$. Hence, for large N , the entropy per particle grows as $\frac{1}{2} \ln 2 = \ln \sqrt{2}$. Therefore the qubit is not an “internal” degree of freedom of the particles but a collective state of topological origin. Interestingly, there are physical systems, known as non-abelian fractional quantum Hall fluids that embody this physics and are accessible to experiments! For these reasons, the non-abelian case has been proposed as a realization of a topological qubit [169, 170].

10. Topological Phases of Matter

10.1. Topological Insulators

We will now give a brief discussion of the physics of Topological Insulators from a field-theoretic perspective. Topological insulators are systems whose electronic states (band structures) have special topological properties which manifest in the existence of symmetry-protected edge states. For this reason these systems are known as symmetry-protected topological states (or SPTs). The simplest example is found in one space dimension where it is related to the fascinating problem of fractionally charged solitons and electron fractionalization. In several ways many of the concepts involved in these 1+1-dimensional systems can and have been extended to higher dimensions.

10.1.1. Dirac fermions in 2+1 dimensions

We will now see that, in spite of the formal similarities with the 1+1 dimensional case, this theory has different symmetries, particularly concerning parity and time reversal invariance. In addition, in 2+1 dimensions it is not possible to define a γ^5 Dirac matrix which implies that there is no chiral symmetry and no chiral anomaly. We will consider first a theory of a Dirac field which in 2+1 dimensions is a bi-spinor (as is in 1+1 dimensions). The Lagrangian of the Dirac theory coupled to a background gauge field A_μ is

$$\mathcal{L} = \bar{\psi} i\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - eA_\mu \bar{\psi}\gamma_\mu \psi \equiv \bar{\psi} i\gamma^\mu D_\mu(A)\psi - m\bar{\psi}\psi \quad (274)$$

where $\bar{\psi} = \psi^\dagger \gamma_0$, $\bar{\psi}\gamma_\mu \psi \equiv j_\mu$ is the gauge-invariant (and conserved) Dirac current, and $D_\mu(A) = \partial_\mu + ieA_\mu$ is the covariant derivative.

In Eq.(274), γ_μ are the three 2×2 Dirac matrices which obey the algebra, $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}$, where $g_{\mu\nu} = \text{diag}(1, -1, -1)$ is the metric of 2+1 dimensional Minkowski spacetime. The Dirac matrices γ_μ can be written in terms of the three 2×2 Pauli matrices. For instance we can choose $\gamma_0 = \sigma_3$, $\gamma_1 = i\sigma_2$ and $\gamma_2 = -i\sigma_1$ which satisfy the Dirac algebra. Since the three gamma matrices involve all three Pauli matrices, it is not possible to define a γ_5 matrix which would anticommute with the gamma matrices. In this sense, it is not possible to define a chirality for bi-spinors in 2+1 dimensions

We could have also chosen a different set of gamma matrices. For instance, we could have chosen $\gamma_0 = \sigma_3$, $\gamma_1 = i\sigma_2$ and $\gamma_2 = +i\sigma_1$ which also satisfy the Dirac algebra. These two choices are equivalent to the 2D parity transformation $x_1 \rightarrow x_1$ and $x_2 \rightarrow -x_2$. In other words, we can choose the three gamma matrices to be defined in terms of a right handed or a left handed frame (or triad). Thus, the choice of handedness of the frame used to define the gamma matrices can be regarded as a chirality.

The parity transformation is also equivalent to a unitary transformation (a change of basis) $U = \gamma_1 = i\sigma_1$ which flips the sign of both γ_2 and γ_0 . Thus, a parity transformation is equivalent to a change of the sign of γ_0 or, what is the same, as changing the *sign* of the Dirac mass $m \rightarrow -m$. Since the Dirac theory is charge conjugation invariant, the parity transformation is equivalent to time reversal. It is easy to see that in 2+1 dimensions the massive Dirac theory is not invariant under time reversal since this the single particle Dirac Hamiltonian $H(\mathbf{p})$ for momentum \mathbf{p} involves all three Pauli matrices,

$$H(\mathbf{p}) = \alpha \cdot \mathbf{p} + \beta m \quad (275)$$

where, as usual, $\alpha = \gamma_0\gamma$ and $\beta = \gamma_0$. In fact, time reversal is equivalent to the change of the sign of the Dirac mass. These considerations also apply to the massless Dirac theory coupled to a background gauge field.

10.1.2. Dirac Fermions and Topological Insulators

In addition systems in one space dimension, discussed in section 7.1, in which Dirac fermions naturally describe the low energy electronic degrees of freedom, Dirac fermions also play a role in many other system in Condensed Matter Physics. Such systems range from the nodal Bogoliubov fermionic excitations of d-wave superconductors in 2D and p-wave superfluids in 3D, to Chern and \mathbb{Z}_2 topological insulators in 2D and \mathbb{Z}_2 topological insulators and Weyl semimetals in 3D. They also play a role in spin liquid phases of frustrated spin-1/2 quantum antiferromagnets, such as Dirac and chiral spin liquids. Dirac fermions play also a significant role in our understanding of the compressible limit of 2D electron gases (2DEG) in large magnetic fields (see section 10.3). We will not cover here all of these examples. Instead we will focus of the 2D Chern topological insulators (which exhibit the anomalous quantum Hall effect) and on the 3D \mathbb{Z}_2 topological insulators.

In Condensed matter Physics the important low energy electronic degrees of freedom belong to the electronic states close to the Fermi energy. In such systems the lattice periodic potential determines the properties of their band structures. The only significant exception to this general rule is the case of the 2DEGs in GaAs-AlAs heterostructures (the most commonly used platform for the study of quantum hall effects) whose electronic densities are low enough so that the lattice periodic potential can for practical purposes almost always be ignored.

In general, Dirac (and Weyl) fermions arise when the conduction and valence bands cross at isolated points of the Brillouin zone. In such situations, the near the crossing points the low energy electronic states locally (in momentum space) look like cones and, ignoring possible anisotropies, look like the states of a Dirac-Weyl theory. For these reasons, Dirac fermions play a central role in the theory of graphene type materials [171] and, more significantly, in the theory of topological insulators [172].

Dirac fermions play a central role in Quantum Electrodynamics and in the Standard Model of Particle Physics. The problem of quark confinement requires an understanding of these theories in a regime inaccessible to Feynman-diagrammatic perturbation theory and an intrinsically non-perturbative formulation, known as Lattice Gauge Theory [55, 56, 173], needed to be developed. For this reason in the 1970s fermionic Hamiltonians with crossings at points became of great interest in High Energy Physics as a way to describing the short-distance dynamics of quarks in Lattice Gauge Theory.

This program run into difficulties when extended to the theory of Weak Interactions which have an odd number of species of chiral (Weyl) Dirac fermions. All local discretizations of the Dirac equation yield an even number of species. This fact came to be known as the fermion doubling problem. These results are special cases of a general theorem, due to Holger Nielsen and Masao Ninomiya [174, 175, 176] (and extended by Daniel Friedan [177]), which proves that for systems whose kinetic energy is *local* in space (as it must be) there must always be an *even* number of crossings. Therefore, it is impossible to write a local theory with an odd number of chiral fermionic species.

In the context of Condensed Matter Physics, the simplest example of Dirac fermions is found in the electronic structure of graphene (a single layer of graphite) discussed by A. H. Castro Neto and coworkers [171]. Graphene is an allotrope of crystalline carbon in which the carbon atoms are arranged in a 2D hexagonal lattice, which is a lattice with two inequivalent sites in its unit cell, labeled by A and B . Let \mathbf{r}_A and \mathbf{r}_B be the sites of the two sublattices. Each site \mathbf{r}_A has three nearest neighbors on the B sublattice located at $\mathbf{r}'_B = \mathbf{r}_A + \mathbf{d}_i$, where $i = 1, 2, 3$ and

$$\mathbf{d}_1 = \left(\frac{1}{2\sqrt{3}}, \frac{1}{2} \right), \quad \mathbf{d}_2 = \left(\frac{1}{2\sqrt{3}}, -\frac{1}{2} \right), \quad \mathbf{d}_3 = \left(-\frac{1}{\sqrt{3}}, 0 \right) \quad (276)$$

(in units in which the lattice spacing is $a = 1$). The nearest neighbor A sites are related by the vectors

$$\mathbf{a}_1 = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2} \right), \quad \mathbf{a}_2 = (0, 1), \quad \mathbf{a}_3 = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2} \right) \quad (277)$$

where \mathbf{a}_1 and \mathbf{a}_2 are the primitive lattice vector of the hexagonal lattice. The nearest neighbor sites of the B sublattice are also related by these same three vectors.

The low energy electronic degrees of freedom of graphene can be described by a single fermionic state (ignoring spin) at each carbon atom. Let $c^\dagger(\mathbf{r}_A)$ and $c^\dagger(\mathbf{r}_B)$ be the fermion operators that creates an electron at the site \mathbf{r}_A and at the site nearest neighbor sites \mathbf{r}_B , respectively. The Hamiltonian is

$$H = t_1 \sum_{\mathbf{r}_A, i=1,2,3} c^\dagger(\mathbf{r}_A) c(\mathbf{r}_A + \mathbf{d}_i) + \text{h.c.} \quad (278)$$

where t_1 is the hopping matrix element between nearest neighbor A and B sites.

In Fourier space

$$c(\mathbf{r}_A) = \int_{\text{BZ}} \frac{d^2k}{(2\pi)^2} \psi_A(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}_A), \quad c(\mathbf{r}_B) = \int_{\text{BZ}} \frac{d^2k}{(2\pi)^2} \psi_B(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}_B) \quad (279)$$

where BZ is the first Brillouin zone of the hexagonal lattice, a hexagonal region of reciprocal space with vertices at $\pm 2\pi/\sqrt{3}, 1/\sqrt{3}$ (which are denoted by K and K' , respectively) and their two images under $2\pi/3$ rotations. In momentum space the Hamiltonian is [178]

$$H = t_1 \int_{\text{BZ}} \frac{d^2k}{(2\pi)^2} \begin{pmatrix} \psi_A(\mathbf{k}) & \psi_B(\mathbf{k}) \end{pmatrix} \begin{pmatrix} 0 & \sum_{j=1,2,3} \exp(i\mathbf{k} \cdot \mathbf{d}_j) \\ \sum_{j=1,2,3} \exp(-i\mathbf{k} \cdot \mathbf{d}_j) & 0 \end{pmatrix} \begin{pmatrix} \psi_A(\mathbf{k}) \\ \psi_B(\mathbf{k}) \end{pmatrix} \quad (280)$$

The one-particle states of this Hamiltonian have eigenvalues $E(\mathbf{k}) = \pm \left| \sum_{j=1,2,3} \exp(i\mathbf{k} \cdot \mathbf{d}_j) \right|$. Clearly $E(\mathbf{k})$ vanishes at the K and K' points, the corners of the Brillouin zone. Hence, there is a crossing of the two bands at the K and K' points near which the energy eigenvalues are a two cones at K and K' , respectively. Thus, the low energy states of graphene consists of two massless (gapless) Dirac bi-spinors, with opposite chirality/parity.

Other examples with similar fermion content are a theory of fermions on a 2D square lattice with a π magnetic flux (1/2 of the flux quantum) per plaquette which arises, for instance, in the theory of the integer quantum Hall effect on lattices by David Thouless, Mahito Kohmoto, Peter Nightingale and Marcel den Nijs [179] (albeit for more general flux per plaquette), and in the theory of flux phases of 2D spin liquids of Ian Affleck and J. Brad Marston [180]. It also arises in the theory of the chiral spin liquid of Xiao-Gang Wen, Frank Wilczek and Anthony Zee [181]. We will discuss these theories in section 10.

10.1.3. Chern invariants

The single-particle quantum states of free fermions (electrons) in the periodic potential of a crystal are Bloch wave functions labeled by the lattice momentum \mathbf{k} and band indices m . The Bloch states are periodic functions of the lattice momenta whose periods are the Brillouin zones. Topologically each Brillouin zone is a torus: in 1D is a 1-torus, in 2D a 2-torus, etc. The symmetries (and shapes) of the Brillouin zone are dictated by the symmetries of the host crystal.

In this section we will consider the quantum states of systems of fermions on a 2D lattice with M bands with eigenvalues $\{E_m(\mathbf{k})\}$ with $m = 1, \dots, M$. The eigenstates are Bloch states $\{|u_m(\mathbf{k})\rangle\}$ such that the wave functions are $\psi_m(\mathbf{x}) = u_m(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x})$, where \mathbf{k} is a (quasi) momentum in the first Brillouin zone [182]. We will assume that the band spectrum is such that all the bands are separated from each other by a finite gap for all momenta in the first Brillouin zone. Hence, the eigenvalues obey the inequality $|E_m(\mathbf{k}) - E_n(\mathbf{k})| > 0$. We will assume that the system of interest has $N < M$ filled bands and, consequently, that there is a gap separating the top-most occupied band N (the ‘‘valence band’’) and the lowest unoccupied band $N+1$ (the ‘‘conduction band’’) that does not close everywhere in the first Brillouin zone.

Let us consider specifically a 2D system and a Bloch state $|u_m(\mathbf{k})\rangle$ at a momentum \mathbf{k} , and let $\partial_k |u_m(\mathbf{k})\rangle$ be an infinitesimally close Bloch state with momentum $\mathbf{k}' = \mathbf{k} + d\mathbf{k}$. The inner product of these two infinitesimally close Bloch states is the vector field (defined on the first Brillouin zone for each band m)

$$\mathcal{A}_j^{(m)}(\mathbf{k}) = i \langle u_m(\mathbf{k}) | \partial_j | u_m(\mathbf{k}) \rangle \quad (281)$$

where $j = 1, 2$ are the orthogonal direction in momentum space and we have used the notation $\partial_i = \partial/\partial k_i$. This vector field is known as the *Berry Connection* of the electronic states in the m th band.

The theory of the quantum states of non-interacting electrons in periodic potentials as developed by Felix Bloch [182] is the foundation of much of what we know about the physics of many semiconductors and simple metals. This theory is the core subject of many textbooks [183, 184]. This theory is based on the classification of the electronic states as representations of the group of lattice translations and point (and spatial) group symmetry transformations. A key unstated assumption in much of this body of work is that the Bloch states are globally well-defined functions on the Brillouin zone of a given crystal. It is rather remarkable that this assumption was not stated explicitly since the original work by Bloch in 1929 until the work by Thouless, Kohmoto, Nightingale and den Nijs on the quantum states of electrons on 2D lattices in the presence of an uniform magnetic field [179]. The realization that there is a topological obstruction to define the

electronic states globally in the Brillouin zone is the key to the development of the theory of topological insulators and, more generally, of the modern theory of band structures [185, 186].

In quantum mechanics states are defined up to a *phase*. This means that Bloch states that differ by a phase describe the same quantum state. In other words each quantum state $|u_m(\mathbf{k})\rangle$ is a member of a *ray* (or *fiber*) of states each labeled by a phase,

$$|u_m(\mathbf{k})\rangle \mapsto \exp(if_m(\mathbf{k})) |u_m(\mathbf{k})\rangle \quad (282)$$

Changing the states by phase factors defines a $U(1)^M$ unitary transformation of physically equivalent states. Since we have a fibre at each point \mathbf{k} , this amounts at defining a *fibre bundle*. However, under this transformation the Berry connection $\mathcal{A}_j^{(m)}(\mathbf{k})$ changes by a gradient of the phases

$$\mathcal{A}_j^{(m)}(\mathbf{k}) \mapsto \mathcal{A}_j^{(m)}(\mathbf{k}) + \partial_j f_m(\mathbf{k}) \quad (283)$$

where we assumed that the phases $f_m(\mathbf{k})$ are continuous and differentiable functions on the entire first Brillouin zone. Since the physical states cannot change by redefinitions of the phases of the basis states we are led to the condition that only the data that is invariant under the gauge transformations defined by Eqs. (282) and (283) is physically meaningful, which is encoded in the gauge-invariant pseudo-scalar quantity $\mathcal{F}^{(m)}(\mathbf{k})$

$$\mathcal{F}^{(m)}(\mathbf{k}) = \epsilon_{ij} \partial_i \mathcal{A}_j^{(m)}(\mathbf{k}) \quad (284)$$

which is known as the *Berry curvature*.

In what follows we will be interested in the quantity

$$C_1^{(m)} = \frac{1}{2\pi} \oint_{\Gamma} dk_j \mathcal{A}_j^{(m)}(\mathbf{k}) = \frac{1}{2\pi} \int_{BZ} d^2k \mathcal{F}^{(m)}(\mathbf{k}) \quad (285)$$

where Γ is the boundary of the 1st Brillouin zone (which we denote by BZ). We will now show that the quantity $C_1^{(m)}$, which measures the *flux* of the Berry connection $\mathcal{A}_j^{(m)}(\mathbf{k})$ through the first Brillouin zone (in units of 2π), is an integer independent of the particular connection $\mathcal{A}_j^{(m)}(\mathbf{k})$ that we have chosen. This integer-valued quantity is a topological invariant known as the first Chern number.

However, since the first Brillouin zone is a 2-torus, which is a closed manifold, a Berry connection with a net flux cannot obey periodic boundary conditions. Instead, we must allow for generalized (large) gauge transformations that wrap around the 2-torus of the first Brillouin zone. This problem is similar (and closely related to) the problem of the wave functions of a charged particle moving in the presence of a magnetic monopole [151]. We will consider a 2D system whose first Brillouin zone is spanned by two reciprocal lattice vectors \mathbf{b}_1 and \mathbf{b}_2 , related by the two primitive lattice vectors \mathbf{a}_1 and \mathbf{a}_2 by the relation $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$ (with $i, j = 1, 2$). The generalized gauge transformations now are

$$\begin{aligned} |u_m(\mathbf{k} + \mathbf{b}_1)\rangle &= \exp(if_m^{(1)}(\mathbf{k})) |u_m(\mathbf{k})\rangle, & |u_m(\mathbf{k} + \mathbf{b}_2)\rangle &= \exp(if_m^{(2)}(\mathbf{k})) |u_m(\mathbf{k})\rangle \\ \mathcal{A}_j^{(m)}(\mathbf{k} + \mathbf{b}_1) &= \mathcal{A}_j^{(m)}(\mathbf{k}) + \partial_j f_m^{(1)}(\mathbf{k}), & \mathcal{A}_j^{(m)}(\mathbf{k} + \mathbf{b}_2) &= \mathcal{A}_j^{(m)}(\mathbf{k}) + \partial_j f_m^{(2)}(\mathbf{k}) \end{aligned} \quad (286)$$

where $f_m^{(1)}(\mathbf{k})$ and $f_m^{(2)}(\mathbf{k})$ are smooth functions of \mathbf{k} .

We will now show that the Bloch states cannot be defined globally over the Brillouin zone if the flux of the Berry connection does not vanish. More specifically let us assume that at some point $\mathbf{k}_0 \in T$ we know the Bloch state $|u_m(\mathbf{k}_0)\rangle$ which satisfies the generalized periodic boundary conditions of Eq.(286). The question is, can we determine the Bloch state at some other also arbitrary point \mathbf{k}'_0 ? We will now show that there is a topological obstruction that does not allow it if the flux of the Berry phase is not zero. The reason is that the *phase* of the Bloch state cannot be defined since, in general, the Bloch state will vanish at some point of the Brillouin zone where the phase is undefined.

We will prove that this is true by following an elegant construction due to Mahito Kohmoto which goes as follows [187]. The Brillouin zone is a torus, that we will denote by T , can be regarded as the tensor product of two circles, $T \equiv S_1 \times S_1$. Let us split the torus T into two disjoint subsets (or patches) H_I and H_{II} such that $T = H_I \cup H_{II}$. We will assume that in region H_I . Let us assume that the Bloch state vanishes at some point $\mathbf{k}_0 \in H_I$ and that the Bloch state does not vanish for all points $\mathbf{k} \in H_{II}$. This means that we can choose the Bloch state $|u_m(\mathbf{k})\rangle$ to be real for all $\mathbf{k} \in H_{II}$. On the other hand we can always assign some arbitrary phase to the

Bloch state at $\mathbf{k}_0 \in T_I$. Once we have done that we can extend the definition of the phase to a neighborhood of \mathbf{k}_0 wholly included in H_I . If we assume that there is only one zero, then the phase can be defined over all of H_I . The result is that we now have two different definitions of the phase of the Bloch state on H_I and on H_{II} . Let $|u_m(\mathbf{k})\rangle_I$ and $|u_m(\mathbf{k})\rangle_{II}$ be the two resulting definitions of the Bloch state. Let the closed curve γ be the common boundary of regions H_I and H_{II} , $\gamma = \partial H_I = \partial H_{II}$. However on the common boundary γ these the two definitions of the Bloch state must be a gauge transformation,

$$|u_m(\mathbf{k})\rangle_I = \exp(if_m(\mathbf{k})) |u_m(\mathbf{k})\rangle_{II} \quad (287)$$

on all points $\mathbf{k} \in \gamma$. The gauge transformation $f_m(\mathbf{k})$, known as the transition function, is a smooth periodic function of the points \mathbf{k} on the closed curve γ . Likewise, the Berry connection $\mathcal{A}_j^{(m)}(\mathbf{k})$ has two definitions on the regions H_I and H_{II} which differ by a gauge transformation on all the points \mathbf{k} of the common boundary γ ,

$$\mathcal{A}_j^{(m)}(\mathbf{k})\Big|_I - \mathcal{A}_j^{(m)}(\mathbf{k})\Big|_{II} = \partial_j f_m(\mathbf{k}) \quad (288)$$

The transition function defines a mapping of the closed curve γ , which is topologically equivalent to a circle S_1 , to the phase of the Bloch state which is defined mod 2π and hence is also a circle S_1 . Hence the transition functions $f_m(\mathbf{k})$ are homotopies which can be classified by the homotopy group $\Pi_1(S_1) \simeq \mathbb{Z}$. We recognize that the classification of the the transition functions is the same that we used for the vortices of section 5.1.1. Hence, the change of the transition function on a full revolution of the closed curve γ must be an integer multiple of 2π .

We will now compute the quantity $C_1^{(m)}$, given in Eq.(285), which measures the flux of the Berry connection through the 2-torus T that defines the first Brillouin zone. Using the partition of the torus $T = T_I \cup T_{II}$ we can write

$$C_1^{(m)} = \frac{1}{2\pi} \int_T d^2k \mathcal{F}^{(m)}(\mathbf{k}) = \frac{1}{2\pi} \left(\int_{H_I} d^2k \mathcal{F}^{(m)}(\mathbf{k}) + \int_{H_{II}} d^2k \mathcal{F}^{(m)}(\mathbf{k}) \right) \quad (289)$$

Using Stokes Theorem on the two regions H_I and H_{II} we get

$$C_1^{(m)} = \frac{1}{2\pi} \oint_\gamma dk_j \left(\mathcal{A}_j(\mathbf{k})\Big|_I - \mathcal{A}_j(\mathbf{k})\Big|_{II} \right) \quad (290)$$

Since the two definitions of the Berry connection differ by a gauge transformation, Eq.(288), we can express $C_1^{(m)}$ in terms of the transition function $f_m(\mathbf{k})$ on the curve γ

$$C_1^{(m)} = \frac{1}{2\pi} \oint_\gamma d\mathbf{k} \cdot \partial f_m(\mathbf{k}) = \frac{1}{2\pi} (\Delta f_m(\mathbf{k}))_\gamma = n \quad (291)$$

where we used that the transition functions are classified by the integer-valued quantity $n \in \mathbb{Z}$.

We conclude that the flux of the Berry curvature $C_1^{(m)}$ takes only integer values. It is known of the first Chern number $C_1^{(m)}$ which is a topological invariant since it cannot change by a smooth redefinition of the curve γ . Since $C_1^{(m)} \neq 0$ implies that the Bloch states must vanish at least at some point (or points) $\mathbf{k} \in H_I$, then the integer-valued Chern number can only change if a redefinition of the curve γ crosses at least one point \mathbf{k} at which the Bloch state vanishes.

The conclusion of this analysis is that whenever the flux of the Berry curvature through the Brillouin zone does not vanish the Bloch states cannot be defined globally on the Brillouin zone. A direct consequence is that in this case the band is characterized by a topological invariant called the first Chern number. This number is a property of a given band and, in general, it is different for each band. This is a particular case of a more general topological classification of the states of free fermions on a lattice which depends on on the dimensionality of the system [188, 189, 190].

10.1.4. The quantum Hall effect on a lattice

We will now show that 2D free fermion systems with Chern bands, i.e. bands characterized by a non-vanishing Chern number, are insulators that have a quantized Hall conductivity. We will do this for the theory of the integer quantum Hall effect on a 2D lattice of Thouless, Kohmoto, Nightingale and den Nijs (TKNN) [179, 187] (see, also, Ref. [191]). Although this problem played a key role in the development of the theory of topological phases of matter, for many

years it was viewed as an academic problem since it would require gigantic magnetic fields to be in the regime of interest for typical solids. The situation changed with the development of twisted bilayer graphene and similar systems which have unit cells large enough for this physics to be observable. These new materials has allowed to study this problem experimentally.

TKNN considered a system of free fermions on a planar (actually square) lattice in a uniform magnetic field B with flux $2\pi p/q$ per plaquette (in units in which the flux quantum $\phi_0 = e\hbar/c = 1$) with p and q two co-prime integers. The tight-binding Hamiltonian for this simple problem is

$$H = \sum_{\mathbf{r}, j=1,2} t_j c^\dagger(\mathbf{r}) e^{iA_j(\mathbf{r})} c(\mathbf{r} + \mathbf{e}_j) + \text{h.c.} \quad (292)$$

where $c(\mathbf{r})$ and $c^\dagger(\mathbf{r})$ are fermion creation and annihilation operators on the lattices sites $\{\mathbf{r} = (m, n)\}$, \mathbf{e}_j are lattice unit vectors, and t_j , with $j = x, y$, is a hopping matrix element between nearest neighbor sites of the square lattice along the x and y directions. On each link of the lattice we defined a vector potential $A_j(\mathbf{r})$, where the oriented sum of the lattice vector potentials on each plaquette $2\pi\phi$, where $\phi = \frac{p}{q}$, is the flux on each plaquette of the square lattice. In the axial (Landau) gauge $A_1(m, n) = 0$, $A_2(m, n) = 2\pi m\phi$ is a periodic function of m with period q . In this gauge the Hamiltonian becomes

$$\begin{aligned} H = & t \sum_{m,n} \left(c^\dagger(m+1, n)c(m, n) + c^\dagger(m, n)c(m+1, n) \right) \\ & + t \sum_{m,n} \left(c^\dagger(m, n+1) e^{i2\pi\phi m} c(m, n) + c^\dagger(m, n) e^{-i2\pi\phi m} c(m, n+1) \right) \end{aligned} \quad (293)$$

In this gauge the problem reduces to a system with a $q \times 1$ unit cell with q inequivalent sites. In momentum space, the (*magnetic*) Brillouin zone of this system is $-\frac{\pi}{q} \leq k_1 \leq \frac{\pi}{q}$ and $-\pi \leq k_2 \leq \pi$. The Fourier transform of the operators $c^\dagger(\mathbf{r})$ and $c(\mathbf{r})$ are, respectively, $c^\dagger(\mathbf{k})$ and $c(\mathbf{k})$ which obey the standard anticommutation relations, $\{c(\mathbf{k}), c(\mathbf{q})\} = \{c^\dagger(\mathbf{k}), c^\dagger(\mathbf{q})\} = 0$, and $\{c^\dagger(\mathbf{k}), c(\mathbf{q})\} = \delta(\mathbf{k} - \mathbf{q})$. In momentum space the Hamiltonian becomes

$$H = q \int_{-\pi/q}^{\pi/q} \frac{dk_x}{2\pi} \int_{-\pi}^{\pi} \frac{dk_y}{2\pi} \tilde{H}(k_x, k_y) \quad (294)$$

where

$$\begin{aligned} \tilde{H}(k_x, k_y) = & \frac{1}{q} \sum_{n=0}^{q-1} \left\{ 2t_x \cos(k_x + 2\pi\phi n) c^\dagger(k_x + 2\pi\phi n, k_y) c(k_x + 2\pi\phi n, k_y) \right. \\ & \left. + t_y [e^{-ik_y} c^\dagger(k_x + 2\pi(n+1)\phi, k_y) c(k_x + 2\pi n\phi, k_y) + e^{ik_y} c^\dagger(k_x + 2\pi(n-1)\phi, k_y) c(k_x + 2\pi n\phi, k_y)] \right\} \end{aligned} \quad (295)$$

The spectrum of this system was first investigated by Hofstadter [192] consists of q bands which for q an odd integer are separated by finite energy gaps.

For fixed values of k_x and k_y (in the magnetic Brillouin zone), the Hamiltonian $\tilde{H}(k_x, k_y)$ is the tight-binding model of a one-dimensional chain on the 1D lattice of q sites located at $k_x + 2\pi n\phi$ with nearest neighbor hopping. The single-particle (Bloch) states $u_n(\mathbf{k})$ of this 1D model obeys the Schrödinger Equation

$$2t_x \cos(k_x + 2\pi n\phi) u_n(\mathbf{k}) + t_y \left(e^{-ik_y} u_{n-1}(\mathbf{k}) + e^{ik_y} u_{n+1}(\mathbf{k}) \right) = E_n u_n \quad (296)$$

which is known as Harper's equation. In general this equation does not admit an analytic solution. However, the nature qualitative features of the spectrum can be obtained by an expansion either on t_x/t_y or on t_y/t_x . TKNN used degenerate perturbation theory to show that the spectrum has q bands and that the r -th band is characterized by two integers s_r and ℓ_r which are the solution of the Diophantine equation

$$r = q s_r + p \ell_r \quad (297)$$

with $\ell_0 = s_0 = 0$. It is also obvious that for $r = q s_q = 1$ and $\ell_q = 0$ (for all p).

Furthermore, TKNN showed that there was an, until then unsuspected, relation between the Hall conductivity σ_{xy} of a gapped system of electrons in periodic potentials in a uniform magnetic field and the Berry connection of the filled bands. This relation implies the quantization of the

Hall conductivity and its computation in terms of a topological invariant, the Chern number of the occupied bands. They showed that in this context each band has a non-trivial Berry connection $\mathcal{A}_j^{(r)} = i\langle u_r(\mathbf{k}) | \partial_j | u_r(\mathbf{k}) \rangle$ (with $\partial_j = \partial_{k_j}$) of the form of Eq.(281) with a non-vanishing (first) Chern number $C_1^{(r)}$, i.e. the flux through the magnetic Brillouin zone.

The conductivity tensor characterizes the electrical properties of a physical system. Linear Response Theory provides a framework for computing the conductivity tensor by perturbing the system with a weak external electromagnetic field A_μ and computing the currents that they induce [193, 10]. The expectation value of the gauge-invariant current operator $J_\mu(x)$ is

$$\langle J_\mu(x) \rangle = -\frac{i}{\hbar} \frac{\delta}{\delta A_\mu(x)} \ln Z[A_\mu] \quad (298)$$

where $Z[A_\mu]$ is the partition function in the presence of a background (i.e. classical) electromagnetic field $A_\mu(x)$. In general spacetime dimension $D = d + 1$, the lowest order in the vector potential A_μ , the induced current is given in terms of the polarization tensor $\Pi_{\mu\nu}(x, y)$,

$$\ln Z[A_\mu] = \frac{i}{2} \int d^D x \int d^D y A_\mu(x) \Pi^{\mu\nu}(x, y) A_\nu(y) + O(A^3) \quad (299)$$

Therefore, the induced current is related to the external field by

$$\langle J_\mu(x) \rangle \equiv J_\mu(x) = \int d^D y \Pi_{\mu\nu}(x, y) A^\nu(y) \quad (300)$$

Expressions of this type are known as a *Kubo formula*. The polarization tensor can be regarded as a *generalized susceptibility*.

Although we are using a continuum relativistic notation, these expressions are generally valid, even in lattice systems. Gauge invariance requires that the polarization tensor be conserved,

$$\partial_\mu \Pi^{\mu\nu}(x, y) = 0 \quad (301)$$

In general, the (retarded) polarization tensor $\Pi^{\mu\nu}(x, y)$ is related to the (retarded) current-current correlation function $\mathcal{D}_{\mu\nu}^R(x, y)$,

$$\mathcal{D}^{\mu\nu}(x, y) = -\frac{i}{\hbar} \Theta(x_0 - y_0) \langle [J_\mu(x), J_\nu(y)] \rangle \quad (302)$$

by the identity

$$\Pi_{\mu\nu}^R(x, y) = \mathcal{D}_{\mu\nu}^R(x, y) - i\hbar \left\langle \frac{\delta J_\mu(x)}{\delta A_\nu(y)} \right\rangle \quad (303)$$

The last term in Eq.(303), usually called a “contact term”. This term vanishes only for a theory of relativistic fermions (in the continuum). In all other cases, lattice or continuum, relativistic or not, the contact term does not vanish and its form depends on the specific theory. In non-relativistic systems, e.g. in a Fermi liquid, this term is the origin of the *f*-sum rule [193, 194].

When the external field A_μ represents a (locally) uniform electric field E , the induced current is $J_i = \sigma_{ij} E_j$, where σ_{ij} is the conductivity tensor, which can be obtained from the polarization tensor as the limit

$$\sigma_{ij} = \lim_{\omega \rightarrow 0} \frac{1}{i\omega} \lim_{\mathbf{q} \rightarrow 0} \Pi_{ij}(\omega, \mathbf{q}) \quad (304)$$

In a metal, which has a Fermi surface, the order of limits in which ω and \mathbf{q} vanish matters and only the order of limits specified above is the correct one to take. In a Dirac system, that we will discuss below, the order does not matter due to relativistic invariance. The other case in which the order does not matter is the Chern insulator that we are interested in. In general, in an isotropic system, the conductivity tensor has a symmetric part and an antisymmetric part. The symmetric part of the conductivity tensor yields the longitudinal conductivity which has all the effects of dissipation. The antisymmetric part does not vanish in the presence of a magnetic field or, more generally, if time reversal symmetry is broken, and yields the Hall conductivity.

A Chern insulator is an insulator and as such is a state with an energy gap. In such a state the longitudinal conductivity vanishes since there is no dissipation in a gapped state. But in a Chern insulator time reversal invariance is broken. In the system that we are discussing is broken by

the magnetic field. The Hall conductivity can be calculated from the antisymmetric part of the polarization tensor as the limit

$$\sigma_{xy} = \lim_{\omega \rightarrow 0} \frac{i}{\omega} \Pi_{xy}(\omega, 0) \quad (305)$$

In addition, the contact term does not contribute to the Hall conductivity. As a result the hall conductivity can be computed in terms of the antisymmetric component of the current-correlation function.

let us now return to the theory of free fermions on a lattice in a (commensurate) magnetic field takes. As we saw the electronic states are split into q bands with single particle states $|\psi_n(\mathbf{k})\rangle$ where \mathbf{k} takes values on the first magnetic Brillouin zone. We will compute the Hall conductivity for this system assuming that the Fermi energy E_F lies in the gap between the n -th and the $n - 1$ -th bands. Let us label by α the occupied bands and by β the unoccupied bands. Hence $E_\alpha(\mathbf{k}) < E_F < E_\beta(\mathbf{k})$. In this case, the Kubo formula for the Hall conductivity becomes

$$\sigma_{xy} = i \frac{e^2}{\hbar} \sum_{E_\alpha < E_F < E_\beta} \int_{BZ} \frac{d^2 k}{(2\pi)^2} \frac{\langle u_\alpha(\mathbf{k}) | J_y | u_\beta(\mathbf{k}) \rangle \langle u_\beta(\mathbf{k}) | J_x | u_\alpha(\mathbf{k}) \rangle - \langle u_\alpha(\mathbf{k}) | J_x | u_\beta(\mathbf{k}) \rangle \langle u_\beta(\mathbf{k}) | J_y | u_\alpha(\mathbf{k}) \rangle}{(E_\alpha(\mathbf{k}) - E_\beta(\mathbf{k}))^2} \quad (306)$$

In the lattice model the current operator is simply

$$\mathbf{J} = \frac{e}{\hbar} \partial_k \tilde{H}(\mathbf{k}) \quad (307)$$

Using this form of the current and the Schrödinger equation of Eq.(296) we can rewrite the Kubo formula for the Hall conductivity, Eq.(306), as a sum over only the occupied states labeled by α . The expression for the Hall conductivity now takes the more compact form

$$\sigma_{xy} = \frac{e^2}{\hbar} \sum_{\alpha} \int_{BZ} \frac{d^2 k}{(2\pi)^2} \epsilon_{jl} \partial_j \langle u_\alpha(\mathbf{k}) | \partial_k | u_\alpha(\mathbf{k}) \rangle = \frac{e^2}{\hbar} \sum_{\alpha} \frac{1}{2\pi} \int_{BZ} \mathcal{F}^{(\alpha)}(\mathbf{k}) = \frac{e^2}{\hbar} \sum_{\alpha} C_1^{(\alpha)} \quad (308)$$

Therefore, we conclude that each occupied band contributes to the Hall conductivity an amount equal to its first Chern number (multiplied by e^2/h). This is the main result of TKNN [179].

In the specific case of the model that we are discussing here the Chern number is given explicitly in terms of the solution of the Diophantine equation, Eq.(297)

$$C_1^{(r)} = \frac{1}{2\pi} \oint_{\Gamma} dk_j \mathcal{A}_j^{(r)}(\mathbf{k}) = \ell_r - \ell_{r-1} \quad (309)$$

where the integers ℓ_r and ℓ_{r-1} are the solutions of the Diophantine equation for the bands r and $r - 1$ (A detailed derivation can be found in Ref.[9], chapter 12). The TKNN integers ℓ_r are topological invariants of the bands.

We conclude, with TKNN, that for a system with r filled bands the Hall conductivity $\sigma_{xy}^{(r)}$ is given by the expression

$$\sigma_{xy}^{(r)} = \frac{e^2}{h} \sum_{n=1}^r (\ell_r - \ell_{r-1}) = \frac{e^2}{h} \ell_r \quad (310)$$

thus showing that this system exhibits an integer quantum Hall effect and that the Hall conductivity is (up to the ratio of universal constants e^2/h) given by a topological invariant, the first Chern number of the band.

We conclude with two comments. The first is that when *all* the bands are occupied the Hall conductivity vanishes. This obvious physical constraint is automatically satisfied by the Diophantine equation. The other comment concerns the case of what happens when q is an even integer. In this case it is possible to show that there are band crossings. This is what happens for $q = 2$ which is the π flux lattice. in this case one has a theory of two species of Dirac bi-spinors. As we will see in section 10.1.5 what happens depends on how the gapless state is reached.

10.1.5. The Anomalous Quantum Hall Effect

We will now discuss a simple and insightful model of a Chern insulator that displays the quantum anomalous Hall effect, that is a Hall effect in a system with broken time reversal symmetry in the absence of a uniform magnetic field. In order to do that we will reconsider the graphene model with additional terms in the Hamiltonian of Eq.(280) that will open up energy gaps in its spectrum. There are several ways to do this. One way requires breaking the symmetry between the two sublattices A and B , by assigning them a site energy of $\pm M$, which breaks

the mirror reflection symmetry of the hexagon. This is what happens with graphene on a boron nitride substrate [178] (there are other ways that involve breaking the point group symmetry [195]). Another way to open a gap is to break time reversal symmetry which is accomplished by complex hopping amplitudes $t_2 \exp(i\phi)$ between nearest neighbor A sites (and between nearest neighbor B sites as well). The phase ϕ is oriented from \mathbf{r}_A to $\mathbf{r}_A + \mathbf{a}_i$, where $\mathbf{a}_1 = \mathbf{d}_2 - \mathbf{d}_3$, etc. (and similarly for the B sites). Physically, the oriented sum of the phases ϕ on the triangles generated by the hopping terms t_1 and t_2 represent the magnetic fluxes through each triangle such that the total flux on the elementary hexagon is zero. This model was considered by Haldane [196]. With these changes the one-particle Hamiltonian becomes 2×2 Hermitian matrix

$$H(\mathbf{k}) = h_0(\mathbf{k}) \mathbf{I} + \mathbf{h}(\mathbf{k}) \cdot \boldsymbol{\sigma} \quad (311)$$

where \mathbf{I} is the 2×2 identity matrix and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is a vector made of the three Pauli matrices. The functions $h_0(\mathbf{k})$ and $\mathbf{h}(\mathbf{k}) = (h_1(\mathbf{k}), h_2(\mathbf{k}), h_3(\mathbf{k}))$ are, respectively, given by

$$\begin{aligned} h_0(\mathbf{k}) &= 2t_2 \cos \phi \sum_{i=1,2,3} \cos(\mathbf{k} \cdot \mathbf{a}_i), & h_1(\mathbf{k}) &= t_1 \sum_{i=1,2,3} \cos(\mathbf{k} \cdot \mathbf{d}_i), \\ h_2(\mathbf{k}) &= t_1 \sum_{i=1,2,3} \sin(\mathbf{k} \cdot \mathbf{d}_i), & h_3(\mathbf{k}) &= M - 2t_2 \sin \phi \sum_{i=1,2,3} \sin(\mathbf{k} \cdot \mathbf{a}_i) \end{aligned} \quad (312)$$

Provided $|t_2/t_1| < 1/3$ the spectrum of the Hamiltonian of Eq.(280) has two bands that do not overlap and have a finite energy gap unless they cross at the corners K and K' the Brillouin zone where the gap is smallest. The two bands cross at K and/or K' if $M = \pm 3\sqrt{3}t_2 \sin \phi$ (for K and K' , respectively). When either one of these conditions are not met the spectrum is gapped at either K or K' or, more generally, at both. For values of the parameters for which the band gaps are small, the low energy states of the Hamiltonian of Eq.(280) is that of a theory of two Dirac fields, that we will label as $\psi_1(x)$ and $\psi_2(x)$, whose one-particle Hamiltonians have the form of Eq.(275) but with generally different masses $m_1 \propto M - 3\sqrt{3}t_2 \sin \phi$ and $m_2 \propto M + 3\sqrt{3}t_2 \sin \phi$. So, in general, both fields will have non-vanishing masses which can have the same sign or opposite sign.

We conclude that the effective low-energy theory consist of two species of massive Dirac bi-spinors, albeit generally with different masses. To examine the electromagnetic response we will couple the lattice model and the effective low energy theory to a slowly varying background electromagnetic field A_μ . The Lagrangian of the low energy theory is

$$\mathcal{L} = \bar{\psi}_1(i\partial^\mu - m_1)\psi_1 + \bar{\psi}_2(i\partial^\mu - m_2)\psi_2 - eA_\mu(\bar{\psi}_1\gamma^\mu\psi_1 + \bar{\psi}_2\gamma^\mu\psi_2) \quad (313)$$

In general the two masses will be different in magnitude and sign. Equivalently the Dirac fermions may have the same or opposite chirality. We will examine the electromagnetic response for both cases in section 10.1.6 where we will see that in the phase in which the masses of both Dirac fermions have the same sign this model describes a topological insulator that exhibits the anomalous quantum Hall effect, whereas in the phase where the signs are opposite it is a conventional insulator.

However, by varying the parameters we can tune to special values of these parameters where one of the two masses vanishes. In section 10.1.6 we will see that these special values of the parameters correspond to quantum phase transitions at which the topological properties of the ground state change. We will see that this is a quantum critical point with special properties. In particular, time reversal invariance is explicitly broken at this quantum critical point. In this regime the heavy Dirac fermion plays the role of the heavy regulating field as in the Pauli-Villars regularization of the theory of a single Dirac fermion. In the lattice model this is a manifestation of the Nielsen-Ninomiya theorem which requires that there should be an even number of Dirac fermions [174, 175, 177]. In this context, the heavy fermion is the “doubler”.

10.1.6. The Parity Anomaly

We saw in section 7.3 that Dirac fermions in 1+1 dimensions have a chiral anomaly. That statement mean that although the theory of free massless Dirac fermions has a global chiral symmetry, the associated current, which we called j_μ^5 , is not conserved when the system is coupled to an external gauge field. This effect happened because the right and left moving (Weyl) components of the massless Dirac fermion cannot be separately conserved in the presence of a gauge field.

Analogs of the 1+1-dimensional chiral anomaly exist on 3+1 dimensions (and, more generally, in all even spacetime dimensions). The cancelation of these anomalies is a key condition

for the consistency of gauge theories [96]. We will see in section 10 that these anomalies play a role on the theory of 3D topological insulators. More generally, anomalies play a central role in the theory of symmetry-protected topological (SPT) phases of matter [197, 198, 199].

Dirac fermions in 2+1 dimensions do not have an analog of the chiral anomaly for the simple reason that the chiral symmetry does not exist. There is, however, an anomaly involving parity or, equivalently, time reversal invariance, known as the parity anomaly. The parity anomaly was discovered in the computation of the effective action of the electromagnetic field in a theory of massive Dirac fermions in 2+1 dimensions by Redlich [200, 201].

Consider theory on a single massive Dirac fermion (a bi-spinor) $\psi(x)$ in 2+1 dimensions coupled to a background U(1) gauge field A_μ . The Lagrangian is

$$\mathcal{L}[\psi, \bar{\psi}, A_\mu] = \bar{\psi} i \not{D} \psi - m \bar{\psi} \psi \quad (314)$$

where the covariant derivative is $D_\mu = \partial_\mu + iA_\mu$. The effective action $S_{\text{eff}}[A_\mu]$ of the gauge field A_μ is formally obtained by computing the path integral

$$Z[A_\mu] = \exp(iS_{\text{eff}}[A_\mu]) = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp(i \int d^3x \mathcal{L}[\psi, \bar{\psi}, A_\mu]) = \text{Det}(i\not{D}[A_\mu] - m) \quad (315)$$

where we used that the theory is free to write the partition function as a functional determinant. In 1+1 dimensions, in the case of a massless theory, this determinant can be computed exactly using the heat kernel (or zeta-function) technique [202, 203]. Aside from important technicalities, the reason for why this works is that as a consequence of the chiral anomaly, the effective action computed at one loop order in the fermions is exact in 1+1 dimensions. In higher dimensions this is no longer true.

We will sketch the computation of the effective action to one loop order, which means quadratic in the gauge field A_μ , which has the form

$$S_{\text{eff}}[A_\mu] = \frac{1}{2} \int d^3x \int d^3y A_\mu(x) \Pi^{\mu\nu}(x-y) A_\nu(y) + \dots \quad (316)$$

where the ellipsis contains the higher order contributions. The polarization operator $\Pi^{\mu\nu}(x-y)$ in momentum space is (here p_μ is the momentum transfer) by the one-loop result

$$\Pi^{\mu\nu}(p) = \int \frac{d^3q}{(2\pi)^3} \text{tr}[S(q+p/2)\gamma^\mu S(q-p/2)\gamma^\nu] \quad (317)$$

The symbol tr signifies the trace over the spinor labels and $S(p)$ is the (massive) Feynman propagator

$$S(p) = \frac{1}{\not{p} - m + i\epsilon} \quad (318)$$

The polarization tensor $\Pi^{\mu\nu}(p)$ has the explicit decomposition into a parity even and a parity odd contributions

$$\Pi^{\mu\nu}(p) = (p^2 g^{\mu\nu} - p^\mu p^\nu) \Pi_0(p^2) - i\epsilon^{\mu\nu\lambda} p_\lambda \Pi_A(p^2) \quad (319)$$

where $p^2 = p_0^2 - \mathbf{p}^2$. The polarization tensor is manifestly transverse, $p_\mu \Pi^{\mu\nu}(p) = 0$ and, hence, the one-loop effective action of Eq.(316) is gauge-invariant (as it should be). The parity even $\Pi_0(p^2)$ and parity odd $\Pi_A(p^2)$ kernels are given by

$$\Pi_0(p^2) = -\frac{|m|}{4\pi p^2} + \frac{1}{8\pi \sqrt{p^2}} \left(\frac{4m^2}{p^2} + 1 \right) \sinh^{-1} \left(\frac{\sqrt{p^2}}{\sqrt{4m^2 - p^2}} \right) \quad (320)$$

$$\Pi_A(p^2) = -\frac{m}{2\pi \sqrt{p^2}} \sinh^{-1} \left(\frac{\sqrt{p^2}}{\sqrt{4m^2 - p^2}} \right) \quad (321)$$

These two results are obtained in the limit in which the regulator scales are sent to infinity leaving behind these finite results, which hold provided the Dirac mass m does not vanish.

In the low energy regime, $|p^2| \ll m^2$, the kernels take the asymptotic values

$$\Pi_0(p^2) \simeq \frac{1}{16\pi|m|} \quad \Pi_A(p^2) \simeq -\frac{1}{4\pi} \text{sgn}(m) \quad (322)$$

This result is the leading contribution to the parity even term (the first term) up to corrections of the order of p^2/m^2 . However, the second term, which is parity odd, is the exact contribution

in the limit $p \rightarrow 0$ for all non-vanishing values of the Dirac mass. In this regime, the effective Lagrangian of the gauge field A_μ takes the local form of the sum of a Maxwell term and a Chern-Simons term

$$\mathcal{L}_{\text{eff}}[A_\mu] = -\frac{1}{4\pi|m|} F_{\mu\nu} F^{\mu\nu} \pm \frac{1}{2} \frac{1}{4\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (323)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength of the gauge field. Clearly, since the Maxwell term has two derivatives while the Chern-Simons term has one derivative, in the long distance regime the Maxwell term is irrelevant (subdominant). Also, while the prefactor of the Chern-Simons term is a pure number, the prefactor of the Maxwell term is proportional to $\frac{1}{|m|}$ (as required by dimensional analysis).

The second term in the Lagrangian of Eq.(323) is a Chern-Simons term for the background gauge field A_μ [163, 164]. This term is gauge-invariant (up to boundary terms) but breaks parity and time reversal invariance. This phenomenon is called the *parity anomaly* [200], although Witten [199] has argued that it is better to call this phenomenon a *time-reversal anomaly*. The *sign* of the prefactor of the Chern-Simons term depends of the *sign* of the Dirac mass m : positive for $m > 0$ and negative for $m < 0$ or, alternatively for positive and negative chirality of the Dirac (bi-spinor) field which fixes the sign of the breaking of time reversal invariance.

In addition to the sign, the prefactor of the Chern-Simons term is equal to $\frac{1}{2}$, and it is not an integer. In section 9.6 we showed that for a *dynamical* Chern-Simons gauge theory to be defined consistently on a closed surface this coefficient must be quantized and should take *integer* values. The fact that the coefficient is half-quantized means that the classical global symmetry of a theory of a single Dirac bi-spinor cannot be gauged, which means that the gauge theory cannot be quantized. This is an example of an obstruction to the quantization of a gauge theory due to an anomaly [96].

We will now use these results to compute the effective low-energy action for the theories of lattice Dirac fermions of section 10.1.2. In those systems the low energy theory is that of *two* Dirac (bi-spinors) with different masses m_1 and m_2 . Since both Dirac fields are coupled (minimally) to the same background gauge field A_μ , the effective Lagrangian is just the sum of the contribution for each Dirac fermion

$$\mathcal{L}_{\text{eff}}[A_\mu] = -\frac{1}{4\pi|m|_{\text{eff}}} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} (\text{sgn}(m_1) + \text{sgn}(m_2)) \frac{1}{4\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (324)$$

where $|m|_{\text{eff}}$ is

$$\frac{1}{|m|_{\text{eff}}} = \frac{1}{|m_1|} + \frac{1}{|m_2|} \quad (325)$$

This result implies that, as anticipated in section 10.1.2, this theory has two phases: a parity even phase and a parity-odd phase. In the regime in which the signs of the two mass terms are equal and opposite, $\text{sgn}(m_1) = -\text{sgn}(m_2)$, the coefficient of the Chern-Simons term cancels and the low-energy effective Lagrangian is a Maxwell term (generally with an effective speed of light much smaller than that in vacuum): this phase is a conventional insulator.

Conversely, in the phase in which the two masses have the same sign, $\text{sgn}(m_1) = \text{sgn}(m_2)$, the coefficient of the Chern-Simons terms does not cancel and it is given by $\pm \frac{1}{4\pi}$, where the sign is the sign of both mass terms. This phase is also an insulator but one in which time-reversal invariance is broken. Moreover, in this phase the induced current j_μ by the background field A_μ in the long-distance limit is controlled by the Chern-Simons term and it is given by

$$j_\mu = \frac{\delta \mathcal{L}_{\text{eff}}}{\delta A_\mu(x)} = \pm \frac{e^2}{2\pi\hbar} \epsilon_{\mu\nu\lambda} \partial^\nu A^\lambda \quad (326)$$

From this result we see that in the parity-broken anomalous quantum Hall phase the system has a correctly quantized and non-vanishing Hall conductivity

$$\sigma_{xy} = \pm \frac{e^2}{h} \quad (327)$$

Therefore, the phase with $\text{sgn}(m_1) = \text{sgn}(m_2)$ displays the *quantum anomalous Hall effect* with a Hall conductivity whose sign equals the signs of both masses. Notice that this is true regardless the magnitudes of the masses m_1 and m_2 , and that only their signs matter. In section 10.1 we will identify this phase with a topological phase of matter known as the *Chern Insulator*.

We showed that the Hall conductivity of the anomalous quantum Hall phase is, as expected, equal to e^2/h using the effective low energy Dirac theory. One may wonder if this approximation

may be missing some contributions to the Hall conductivity. Fortunately there is an alternative quite elegant way of computing the Hall conductivity as a property of the entire occupied band. This approach, originally introduced by Gregory Volovik [204] in the context of superfluid ^3He -A, and adapted to the theory of the quantum anomalous Hall effect by Viktor Yakovenko [205], by Maarten Golterman, Karl Jansen and David Kaplan for a theory of Wilson fermions in odd dimensional hypercubic lattices [206], and by Xiao-Liang Qi, Yong-Shi Wu and Shou-Cheng Zhang [207], involves the derivation of a topological invariant for a two-band model.

As we saw above the Hall conductivity is obtained from the xy component of the polarization operator as the limit

$$\sigma_{xy} = \lim_{\omega \rightarrow 0} \frac{i}{\omega} \lim_{\mathbf{q} \rightarrow 0} \Pi_{xy}(\omega, \mathbf{q}) \quad (328)$$

For a free fermion system $\Pi_{xy}(\omega, 0)$ is given by the current correlator

$$\Pi_{xy}(\omega, 0) = \int_{\text{BZ}} \frac{d^2 k}{(2\pi)^2} \int_{-\infty}^{\infty} \frac{d\Omega}{2\pi} \text{tr} [J_x(\mathbf{k}) G(\mathbf{k}, \omega + \Omega) J_y(\mathbf{k}) G(\mathbf{k}, \Omega)] \quad (329)$$

where $G(\mathbf{k}, \Omega)$ is the propagator for a two-band free fermion system. A generic two-band free fermion system has a one-particle Hamiltonian of the form given in Eq.(311). The (one-body) current operator for such a system is obtained as

$$J_l(\mathbf{k}) = \frac{\partial h_0(\mathbf{k})}{\partial k_l} \mathbf{1} + \frac{\partial h_a(\mathbf{k})}{\partial k_l} \sigma_a \quad (330)$$

The propagator $G(\mathbf{k}, \omega)$ is the 2×2 matrix (in band indices)

$$G(\mathbf{k}, \omega) = \frac{1}{\omega \mathbf{1} - \mathbf{h}(\mathbf{k}) \cdot \boldsymbol{\sigma} + i\epsilon} = \frac{P_+(\mathbf{k})}{\omega - E_+(\mathbf{k}) + i\epsilon} + \frac{P_-(\mathbf{k})}{\omega - E_-(\mathbf{k}) + i\epsilon} \quad (331)$$

where $P_{\pm}(\mathbf{k})$ are the operators that project onto the (empty) conduction band and the (filled) valence band whose energies are, respectively $E_{\pm}(\mathbf{k})$,

$$P_{\pm}(\mathbf{k}) = \frac{1}{2} (\mathbf{1} \pm \hat{\mathbf{h}}(\mathbf{k}) \cdot \boldsymbol{\sigma}) \quad (332)$$

where $\hat{\mathbf{h}}(\mathbf{k})$ is the unit vector defined for every momentum \mathbf{k} of the first Brillouin zone

$$\hat{\mathbf{h}}(\mathbf{k}) = \frac{\mathbf{h}(\mathbf{k})}{\|\mathbf{h}(\mathbf{k})\|} \quad (333)$$

Upon performing the frequency integration and the band traces in the expression for $\Pi_{xy}(\omega, 0)$ of Eq.(329), we find that the Hall conductivity takes the form

$$\sigma_{xy} = \frac{e^2}{2\hbar} \int_{\text{BZ}} \frac{d^2 k}{(2\pi)^2} \epsilon_{abc} \frac{\partial \hat{h}_a(\mathbf{k})}{\partial k_x} \frac{\partial \hat{h}_b(\mathbf{k})}{\partial k_y} \hat{h}_c(\mathbf{k}) (n_+(\mathbf{k})) - n_-(\mathbf{k})) \quad (334)$$

where $n_{\pm}(\mathbf{k})$ are the Fermi functions (at zero temperature) for the two bands. Since

$$E_+(\mathbf{k}) - E_-(\mathbf{k}) = 2\|\mathbf{h}(\mathbf{k})\| = 2\sqrt{\mathbf{h}^2(\mathbf{k})} > 0 \quad (335)$$

there is a non-vanishing gap on the entire Brillouin zone between the occupied valence band, with $n_-(\mathbf{k}) = 1$, and the unoccupied conduction band, with $n_+(\mathbf{k}) = 0$. For the insulating state the Fermi energy lies inside this gap and the expression for the Hall conductivity reduces to the following

$$\sigma_{xy} = -\frac{e^2}{2\hbar} \int_{\text{BZ}} \frac{d^2 k}{(2\pi)^2} \epsilon_{abc} \frac{\partial \hat{h}_a(\mathbf{k})}{\partial k_x} \frac{\partial \hat{h}_b(\mathbf{k})}{\partial k_y} \hat{h}_c(\mathbf{k}) \quad (336)$$

In our discussion of quantum antiferromagnets in one space dimensions in section 5.2 we found that their effective low-energy action contained a crucial topological term proportional to the integer-valued topological invariant Q (the winding number) of Eq.(103) which classifies the smooth maps (homotopies) of the 2D surface (say a sphere S^2) onto the target space of a three-component unit vector field which also a sphere S^2 . These equivalence classes are represented by the notation $\Pi_2(S^2) \simeq \mathbb{Z}$. In the case at hand the unit vector $\hat{\mathbf{h}}(\mathbf{k})$ are points on a 2-sphere. Hence, $\hat{\mathbf{h}}(\mathbf{k})$ is a map of the first Brillouin zone (which is a 2-torus) to the sphere S^2 . Such maps

are also classified by the same integer-valued topological invariant defined in Eq.(103). These results imply that the Hall conductivity of the two-band system is

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar} Q[\hat{h}] \quad (337)$$

In other words we have shown that the Hall conductivity is given in terms of a topological invariant of the occupied band (in units of e^2/h), the winding number Q . In the two-band model the topological invariant Q plays the same role as the Chern number does in the work of TKNN [179, 187] When $Q \neq 0$, the two-band system exhibits the quantized anomalous quantum Hall effect. This is a property of the entire band of occupied states, and not just a consequence of the low energy approximation. This result implies that the low energy approximation captures all of the topology of the band. It also implies that in these lattice models the Berry curvature is highly concentrated near the points in momentum space where the two bands are close in energy.

We will now consider the problem of the quantum phase transition between the trivial and the Chern insulator. To address this problem we will tune the parameters of the lattice model discussed in section 10.1.2, the phase ϕ and the ratio of hopping amplitudes t_2/t_1 , to the point at which the mass of one of the two species of Dirac fermions, say the fermion ψ_1 , is zero, $m_1 = 0$, while keeping the fermion ψ_2 massive, $m_2 \neq 0$. In the low energy regime we have a massless fermion and a massive fermion. This point in parameter space is a quantum phase transition between a trivial insulator and a Chern insulator. In the low energy regime the fermion ψ_1 is massless. A massless fermion is a scale-invariant system in the sense that the correlators of all its observables exhibit power law behavior (free field in this case).

We will now discuss briefly the electromagnetic response of this system at the quantum phase transition where one fermion becomes massless. In particular, it is natural to ask if the coefficient of the parity-odd (Chern-Simons) term non-vanishing *at the quantum phase transition*. To answer this question we will look at the behavior of the parity-even kernel $\Pi_0(p^2)$ and the parity-odd kernel $\Pi_A(p^2)$ in the massless limit for the light fermion, $m_1 \rightarrow 0$. We find that the *total* contribution of both the light fermion and of the heavy fermion to the polarization kernels is (assuming $m_2/m_1 \rightarrow \infty$)

$$\lim_{m_1 \rightarrow 0} \Pi_0(p^2) = \frac{i}{16\sqrt{p^2}}, \quad \lim_{|m_2| \rightarrow \infty} \Pi_A(p^2) = -\frac{1}{4\pi} \text{sgn}(m_2) \quad (338)$$

The important conclusion is that at this quantum critical point the parity-even kernel $\Pi_0(p^2)$ is *non-local* and that the heavy regulator fermion (the “doubler”) yields the leading finite non-vanishing and local contribution to the parity-odd kernel $\Pi_A(p^2)$.

We can now use these results to compute the conductivity *tensor* σ_{ij} at the quantum critical point where $m_1 \rightarrow 0$. Since the system is spatially isotropic the conductivity tensor has the form

$$\sigma_{ij} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix} \quad (339)$$

where we used that $\sigma_{yy} = \sigma_{xx}$ since the system is isotropic. The *longitudinal* conductivity σ_{xx} and the Hall conductivity σ_{xy} are

$$\sigma_{xx} = \frac{\pi e^2}{8h}, \quad \sigma_{xy} = \pm \frac{1}{2} \frac{e^2}{h} \quad (340)$$

in other words, *at the quantum critical point* the system has a finite (and universal) longitudinal conductivity. This result may seem surprising as there is no disorder in this model. A finite universal longitudinal conductivity is a standard occurrence in 2D systems at a quantum critical point. For example, at the superconductor-insulator transition the conductivity is (conjectured) to be $\sigma_{xx} = e^2/2h$ (with $\sigma_{xy} = 0$). In addition, it also has finite and also universal Hall conductivity. The Hall conductivity at the quantum critical point is due to the heavy fermionic “doubler” and is equal to $1/2$ (in units of e^2/h). So, the quantum critical point is not time-reversal invariant since this symmetry is broken at the UV (lattice).

We can examine the massless theory using a more formal approach [199, 208]. In a gauge-invariant regularization of the massless theory the partition function of a Dirac fermion coupled to a background (unquantized) U(1) gauge field, the partition function is not time-reversal invariant.

It is given by

$$Z[A_\mu] = \int \mathcal{D}\bar{\psi} \mathcal{D}\psi \exp\left(i \int d^3x \bar{\psi} i\slashed{D}[A_\mu]\psi\right) = \det(i\slashed{D}[A_\mu]) = |\det(i\slashed{D}[A_\mu])| \exp\left(\pm i\frac{\pi}{2}\eta[A_\mu]\right) \quad (341)$$

where the sign depends on how time-reversal invariance is broken by the choice of regularization. The quantity $\eta[A_\mu]$ that appears in the phase factor is the Atiyah-Patodi-Singer η -invariant [147] which we already encountered in the discussion of the fractionally charged solitons in section 8.3. With some caveats [199, 208], the phase factor of the partition function of Eq.(341) is commonly written in the form of a 1/2-quantized Chern-Simons term

$$\frac{\pi}{2}\eta[A_\mu] \equiv \pm \frac{1}{2} \int d^3x \frac{1}{4\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (342)$$

often denoted as a $U(1)_{1/2}$ Chern-Simons term. This term plays the same role as the contribution of the heavy fermion doubler in the lattice theory.

However, we can also wonder if there is a way to have a time-reversal invariant theory of a *single* massless Dirac fermion, $m \rightarrow 0$. This case cannot be realized in a 2D lattice model but, as will see, it can be realized on the surface states of a 3D time-reversal invariant \mathbb{Z}_2 topological insulator, which we will discuss in section 10.1.

10.2. Three-dimensional \mathbb{Z}_2 topological insulators

The classification of topological insulators in terms of a Chern number is only possible in even space dimensions in systems with broken time reversal symmetry: in $d = 2$ the Berry connection is abelian and the topological invariant is the first Chern number while in $d = 4$ the Berry connection in a non-abelian $SU(2)$ gauge field and the topological invariant is the second Chern number [135], etc. We will now discuss the time-reversal invariant topological insulators [209, 190, 135] and, in particular, those that are invariant under inversion symmetry. Such states exist in both two and space dimensional insulator with strong spin-orbit coupling.

10.2.1. \mathbb{Z}_2 Topological Invariants

Let $\{|u_n(\mathbf{k})\rangle\}$ be the Bloch states. We will represent time reversal by the anti-unitary operator Θ that acts on the single particle (Bloch) states by complex conjugating the state and reverses the spin. For spin-1/2 fermions $\Theta = \exp(i\pi\sigma_2)\mathcal{K}$, where \mathcal{K} is the complex conjugation operator. In this case $\Theta^2 = -1$. Let us assume that we have two occupied Bloch bands for each point \mathbf{k} f the Brillouin zone. In this case the states form a rank-2 vector bundle over the torus of the Brillouin zone. In time reversal invariant systems the anti-unitary time reversal transformation \mathcal{T} induces an involution in the Brillouin zone that identifies the points \mathbf{k} and $-\mathbf{k}$. Time reversal the acts on the one-particle (Bloch) Hamiltonian as $\Theta H(\mathbf{k})\Theta^{-1} = H(-\mathbf{k})$. The states $|u_n(\pm\mathbf{k})\rangle$ are related by time reversal as $|u_n(-\mathbf{k})\rangle = \Theta|u_n(\mathbf{k})\rangle$ which implies that the bundle is real. The condition $\Theta^2 = -1$ implies that the bundle is real. In algebraic topology these bundles are classified by an integer (here the number of occupied bands) and a \mathbb{Z}_2 index that will allow us to classify these states.

In a periodic lattice there exists a set of points \mathbf{Q}_i of the Brillouin zone with the property that they differ by their images under the action of time reversal by a reciprocal lattice vector, $-\mathbf{Q}_i = \mathbf{Q}_i + \mathbf{G}$. In $d = 2$ there four such points and $d = 3$ there are eight points, and are given by $\mathbf{Q}_i = \frac{1}{2} \sum_j n_j \mathbf{b}_j$, where $n_j = 0, 1$, $j = 1, 2$ in $d = 2$ and $j = 1, 2, 3$ in $d = 3$. Here \mathbf{b}_j are the primitive lattice vectors. Kane and Mele [210] defined the $2N \times 2N$ antisymmetric matrix $w_{m,n}(\mathbf{k}) = \langle u_m(-\mathbf{k}) | \Theta | u_n(\mathbf{k}) \rangle$. They showed that at each time reversal invariant point \mathbf{Q}_i one can define an index δ_i

$$\delta_i = \frac{\sqrt{\det[w(\mathbf{Q}_i)]}}{\text{Pf}[w(\mathbf{Q}_i)]} = \pm 1 \quad (343)$$

where $\det[w]$ and $\text{Pf}[w]$ are, respectively, the determinant and the Pfaffian of the matrix w , and $\det[w] = \text{Pf}[w]^2$. The sign of the quantities δ_i can be made unambiguous by requiring that the Bloch states be continuous. In addition, the quantities δ_i are gauge-dependent. However the products

$$(-1)^v = \prod_{i=1}^4 \delta_i \quad (344)$$

in $d = 2$, and

$$(-1)^{\nu_0} = \prod_{i=1}^8 \delta_i, \quad (-1)^{\nu_k} = \prod_{n_k=1, n_{j\neq k}=0,1} \delta_i(n_1, n_2, n_3) \quad (345)$$

are gauge and are also topological invariant. The \mathbb{Z}_2 -valued indices ν and ν_0 are robust to disorder and are called strong topological indices. Furthermore, Fu and Kane showed that if $\xi_n(\mathbf{Q}_i) = \pm 1$ are the parity eigenvalues of the occupied parity eigenstates, the quantities δ_i are given by $\delta_i = \prod_{m=1}^N \xi_{2m}(\mathbf{Q}_i)$. In $d = 3$ the index ν_0 does not rely on the existence of inversion symmetry.

In the case of a two-band model in $d = 2$ and in $d = 3$, the states are four-component spinors reflecting the two bands and the two spin components. In the context of these systems with strong spin-orbit coupling spin is actually the z -component of the atomic total angular momentum \mathbf{J} of the electrons with energies close to the Fermi energy. In these systems the one-particle Hamiltonian $H(\mathbf{k})$ is a 4×4 Hermitian matrix which can be expanded as a linear combination of Dirac matrices. A simple and very useful model of systems of this type is the Wilson fermion model (with continuous time) [55, 211, 135] of a square (cubic) lattice with 4 states per site (parity and spin) whose Hamiltonian three dimensions is

$$H(\mathbf{k}) = \sin \mathbf{k} \cdot \boldsymbol{\alpha} + M(\mathbf{k}) \beta \quad (346)$$

where $\boldsymbol{\alpha} = \gamma_0 \boldsymbol{\gamma}$ and $\beta = \gamma_0$ are the conventional 4×4 Dirac matrices. The γ matrices satisfy the Clifford algebra $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}\mathbf{1}$, where $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ is the metric tensor of four-dimensional Minkowski space time. An additional γ matrix of interest is $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$.

The (Wilson) mass term $M(\mathbf{k})$ in two dimensions is $M(\mathbf{k}) = M + \cos k_1 + \cos p_2 - 2$, and in three dimensions is $M(\mathbf{k}) = M + \cos k_1 + \cos k_2 + \cos k_3 - 3$. In Eq.(346) we see that, consistent with the requirements of the Nielsen-Ninomiya theorem [174, 175], in addition to a possible low energy Dirac fermion (if M is small) there are three more massive Dirac fermions (in $d = 2$) and seven other Dirac fermions in $d = 3$, so that the total number of Dirac fermions is always even, 8 in this case. With Wilson's mass term the additional Dirac fermions (the "doublers") are always heavy even if the Dirac fermion near the Γ point $\mathbf{Q} = 0$ is light.

In the Dirac basis time reversal is the operation $\Theta = (i\sigma_2 \otimes I)\mathcal{K}$, where \mathcal{K} is complex conjugation, and parity is $P = \beta$. The matrices $\boldsymbol{\alpha}$ and β commute with $P\Theta$. At the time-reversal and parity invariant points of the Brillouin zone $\{\mathbf{Q}_i\}$ the Hamiltonian only depends of the the matrix β , $H(\mathbf{Q}_i) = M(\mathbf{Q}_i)\beta$. Since the parities of the spinors are the eigenvalues of the matrix β , we conclude that in the two-band models the quantities δ_i are simply equal to the sign of the mass of the fermions defined at the time-reversal-invariant points \mathbf{Q}_i of the BZ [209, 135], $\delta_i = \delta(\mathbf{Q}_i) = -\text{sgn } M(\mathbf{Q}_i)$. Using this result it follows that in two dimensions the system is a \mathbb{Z}_2 topological insulator with index $\nu = 1 \pmod{2}$ if $0 < M < 4$ while it is trivial for other values of M , i.e. $\nu = 0 \pmod{2}$. Similarly, in three dimensions a \mathbb{Z}_2 (strong) topological insulator exists only if $0 < M < 2$ (in the other regimes this system is in a weak topological insulator state or in a trivial one).

We conclude that both in two and three dimensions the low energy theory of the \mathbb{Z}_2 topological insulators consists of a single Dirac fermion whose mass is small compared to that of the fermion doublers and has the opposite sign. In both cases there is a quantum phase transition between a trivial insulator and the time-reversal invariant topological insulator at the point where he mass of the light fermion vanishes.

10.2.2. The Axial Anomaly and the Effective Action

We will now look at the electromagnetic response of a \mathbb{Z}_2 topological insulator in 3+1 dimensions. This problem can be addressed more easily using the continuum field theory description which is valid in the regime where the mass M is weak. In this regime one species of Dirac fermions is light (i.e. its mass is small) while the Dirac doublers remain heavy. Much as we did in our discussion of the Chern insulators and the *parity anomaly* in section 10.1.6 we will keep in mind that the fermion doublers play the role of heavy regulators, such as in the Pauli-Villars scheme, in Quantum Field Theory. Here too, we will see that a field-theoretic anomaly, known as the *axial anomaly* plays a central role in the physics. The analysis is very similar to what we did with the *chiral anomaly* in 1+1 dimensions in section 7.3.

Let us begin with a theory of a single massive Dirac fermion in 3+1 dimensions. The Lagrangian of the free massive Dirac theory is

$$\mathcal{L} = \bar{\psi} (i\partial^\mu - m) \psi \quad (347)$$

The equation of motion of the spinor field operator $\psi(x)$ (I omit the spinor indices here) is the Dirac equation

$$(i\partial - m)\psi = 0 \quad (348)$$

The Dirac Lagrangian has a global gauge symmetry $\psi \rightarrow e^{i\theta}\psi$ which requires that the Dirac current $j_\mu = \bar{\psi}\gamma_\mu\psi$ is locally conserved, $\partial_\mu j^\mu = 0$. The *massless* Dirac theory can be decomposed into a theory of two Weyl bi-spinor fields which obey separate Dirac Lagrangians. Furthermore, the massless theory has the additional global symmetry under the transformation $\psi \rightarrow e^{i\theta\gamma_5}\psi$ and the additional formally locally conserved *axial* current $j_\mu^5 = i\bar{\psi}\gamma_\mu\gamma^5\psi$. However the conservation law of the axial current is violated in the *massive* Dirac theory

$$\partial^\mu j_\mu^5 = -2mi\bar{\psi}\gamma^5\psi \quad (349)$$

since the two Weyl bi-spinors transmute into each other in the presence of a mass term. This is the origin of the phenomenon of neutrino oscillations. In the condensed matter physics context there is a similar phenomenon in systems of Weyl semimetals which have crossings between the valence and conducting bands at two locations $\pm\mathbf{Q}$ of the BZ, with each crossing associated with each Weyl bi-spinor. A charge density wave with ordering wavevector $2\mathbf{Q}$ mixes the two Weyl fermions which become gapped, becoming effectively a single massive Dirac fermion [212]. This state is often called an “axionic”-charge-density-wave.

We will now show that the axial symmetry has an anomaly and cannot be gauged. Thus, we will consider the problem of a Dirac theory coupled to a background U(1) gauge field A_μ and reexamine the putative conservation of the axial current j_μ^5 . This question can be addressed in different ways. Quite early on Steven Adler [133], and John Bell and Roman Jackiw [134] examined this problem by computing a Dirac fermion triangle diagram for the process of a neutral pion decaying into two photons, $\pi^0 \rightarrow 2\gamma$. In particle physics the pion is the Goldstone boson of the spontaneously broken chiral symmetry. The analog of this problem in condensed matter physics is the phase mode of an incommensurate charge density wave. The relativistic Lagrangian for this problem is a theory of Dirac fermions coupled to a complex scalar field $\phi = \phi_1 + i\phi_2$ through two Yukawa couplings

$$\mathcal{L} = \bar{\psi}iD\psi + g\phi_1\bar{\psi}\psi + ig\phi_2\bar{\psi}\gamma^5\psi - V(\phi_1, \phi_2) = \bar{\psi}iD\psi + g|\phi|\bar{\psi}e^{i\gamma_5\theta}\psi - V(|\phi|^2) \quad (350)$$

where $|\phi|^2 = \phi_1^2 + \phi_2^2$, $\tan\theta = \phi_2/\phi_1$, and $D_\mu = \partial_\mu + ieA_\mu$ is the covariant derivative. Here we are regarding the gauge field A_μ as a background probe field.

The triangle Feynman diagram computes the polarization tensor for the electromagnetic field A_μ with an insertion of the coupling to the complex scalar field in an otherwise massless theory. Assuming a gauge-invariant regularization of the diagram, this computation finds that the axial current j_μ^5 is anomalous and is not conserved even in the massless theory [133, 134]

$$\partial^\mu j_\mu^5 = -\frac{e^2}{16\pi^2}F^{\mu\nu}F_{\mu\nu}^* \quad (351)$$

where $F_{\mu\nu}^* = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}F^{\lambda\rho}$ is the dual of the electromagnetic field tensor. This is the *axial anomaly*.

To see how the axial anomaly arises we will follow the physically transparent approach of Nielsen and Ninomiya [176], that we also employed in section 7.3 in 1+1 dimensions. We will consider a theory of free massless dirac fermions coupled to a background electromagnetic field. Since the theory is massless, the Dirac equation decouples into an equation to the right handed Weyl fermion ψ_R (with positive chirality $\gamma_5\psi_R = +\psi_R$) and a left handed Weyl fermion ψ_L (with negative chirality, $\gamma_5\psi_L = -\psi_L$). In the gauge $A_0 = 0$, the Dirac equations become

$$[i\partial_0 - (-i\partial - e\mathbf{A}) \cdot \boldsymbol{\sigma}]\psi_R = 0, \quad [i\partial_0 - (i\partial - e\mathbf{A}) \cdot \boldsymbol{\sigma}]\psi_L = 0 \quad (352)$$

Let us now consider look at the solutions of the Weyl equation for right-handed fermions ψ_R , Eq.(352). The left-handed fermions ψ_L are analyzed similarly. We will consider a gauge field $A_1 = 0$ and $A_2 = Bx_1$ representing a uniform static magnetic field of strength B pointing along the x_3 direction. In this this gauge the eigenstates are plane waves along the directions x_2 and x_3 and harmonic oscillator states along the direction x_1 . The eigenvalue spectrum consists of Landau levels with energies

$$E(n, p_3, \sigma_3) = \pm \left[2eB \left(n + \frac{1}{2} \right) + p_3^2 + eB\sigma_3 \right]^{1/2} \quad (353)$$

for $n = 0, 1, 2, \dots$, except for the *zero mode* with $n = 0$ and $\sigma_3 = -1$, for which

$$E(n = 0, p_3, \sigma_3 = -1) = \pm p_3 \quad (354)$$

where the + sign holds for ψ_R and the – sign for ψ_L . Just as in the case of non-relativistic fermions the relativistic Landau levels are degenerate. We will consider a system of Dirac fermions at charge neutrality and, hence, $E_F = 0$. The positive and negative energy states are charge conjugate of each other and the negative energy states are filled. For right-handed fermions, the zero mode states with $p_3 < 0$ are filled while for right handed states the zero modes with $p_3 > 0$ are filled. The density of states of the zero modes is $LeB/(4\pi^2)$ (where L is the linear size of the system).

Let us consider now turning on an external electric field E parallel to the magnetic field B . Just as we saw in 1+1 dimensions in section 7.3, the electric field leads to pair creation by shifting the Fermi momentum to p_F for the zero modes. There is no particle creation for the states with $n \neq 0$ and they do not contribute to the anomaly. The rate of creation of right-handed fermions N_R ,

$$\frac{dN_R}{dx_0} = \frac{1}{L} \frac{LeB}{4\pi^2} \frac{p_F}{dx_0} = \frac{e^2}{4\pi^2} EB \quad (355)$$

The annihilation rate of left handed particles is

$$\frac{dN_L}{dx_0} = -\frac{1}{L} \frac{LeB}{4\pi^2} \frac{p_F}{dx_0} = -\frac{e^2}{4\pi^2} EB \quad (356)$$

and the creation rate of left handed anti-particles is

$$\frac{d\bar{N}_L}{dx_0} = \frac{1}{L} \frac{LeB}{4\pi^2} \frac{p_F}{dx_0} = \frac{e^2}{4\pi^2} EB \quad (357)$$

the axial anomaly is the total rate of creation of right handed particles and of left-handed antiparticles:

$$\frac{dQ_5}{dx_0} = \frac{dN_R}{dx_0} + \frac{d\bar{N}_L}{dx_0} = \frac{e^2}{2\pi^2} EB \quad (358)$$

which agrees with the expression of Eq.(351).

We now turn to the \mathbb{Z}_2 topological insulator. As we saw this is a system with two species of (4 component) Dirac spinors. In the topological phase the sign of the Dirac mass term one of the Dirac fermions (the one near the Γ point in the lattice model) is opposite (negative) to the sign of the mass term of the other Dirac fermion (the fermion doubler). Explicit calculations [135, 213, 214] on the lattice model obtain the result that the effective low-energy action for the electromagnetic gauge field A_μ in the topological phase is

$$S_{\text{eff}}[A_\mu] = \int d^4x \left[-\frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu} + \frac{\theta}{32\pi^2} e^2 \epsilon_{\mu\nu\lambda\rho} F^{\mu\nu} F^{\lambda\rho} \right] + \dots \quad (359)$$

In a time-reversal invariant system the allowed values of the θ angle of Eq.(359) are restricted to be $\theta = n\pi$, with $n \in \mathbb{Z}$. The case $\theta = 0 \pmod{2\pi}$ represents a trivial insulator whereas $\theta = \pi \pmod{2\pi}$ holds for a \mathbb{Z}_2 time-reversal invariant topological insulator.

The second term in the effective action of the electromagnetic gauge field of Eq.(359), known as the θ term, has been extensively discussed in the high-energy physics literature [202, 215, 216]. The derivation of this term is subtle. As it stands, unless θ is varying in space-time (in which case this is known as the axion field) this term is a total derivative. In non-abelian Yang-Mills gauge theories this term is proportional to a topological invariant known as the Pontryagin index which counts the instanton number of the gauge field configurations [99, 93]. In the context of the Lagrangian of Eq.(350) this term is induced by the coupling of the Dirac fermion to the Yukawa coupling of the complex scalar field ϕ to the Dirac and γ_5 mass terms. In this case, the lowest order contribution is given by the triangle diagram. The fact that this term is exact at lowest order reflects the fact that the axial anomaly is in fact a non-perturbative effect which is the same in both the weak coupling and the strong coupling regimes [96].

In the phase where the potential $V(|\phi|^2)$ in the Lagrangian of Eq.(171) has a minimum at $|\phi_0| \exp(i\theta)$ the chiral symmetry is spontaneously broken and the phase field $\theta(x)$ is the Goldstone boson of the spontaneously broken chiral symmetry. In this phase the Dirac fermions become massive through the Yukawa couplings to the complex scalar field, and the phase of the field ϕ enters in the effective action as an axion field which couples to the gauge field A_μ through the θ term of the effective action. We should note that in a recently studied axionic CDW state of a Weyl semimetal [212] the phase of the CDW plays the role of the axion field.

10.2.3. Theta terms, and Domain walls: Anomaly and the Callan-Harvey Effect

We will now discuss some remarkable behaviors of three-dimensional \mathbb{Z}_2 topological insulators. We will begin with the electromagnetic response encoded in the effective action of Eq.(359). We will assume that the θ angle is effectively a slowly varying Goldstone mode of the spontaneously broken $U(1)$ chiral symmetry (i.e. the axion field) present in the Lagrangian of Eq.(350). If the field θ is constant, the θ term is a total derivative and it does not contribute to the local equations of motion. We will see below that this term [lays a key role in the physics of a domain wall, which we will regard as the interface of a \mathbb{Z}_2 topological insulator and a trivial insulator. In the general case in which θ varies slowly (as Goldstone modes do) its presence leads to interesting modification of Maxwell's equations, known as axion electrodynamics [215]:

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \tilde{\rho} - e^2 \nabla \theta \cdot \mathbf{B}, & \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{B} &= \partial_t \mathbf{E} + \tilde{\mathbf{j}} + e^2 (\partial_t \theta + \nabla \theta \times \mathbf{E}), & \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (360)$$

where $\tilde{\rho}$ and $\tilde{\mathbf{j}}$ are external probe electric charge and current densities. The equations of axion electrodynamics have many remarkable properties. Here we will focus on effects: the topological magnetoelectric effect [135] and the Witten effect [217].

Let us consider a \mathbb{Z}_2 topological insulator with a flat open boundary (the $x_1 - x_2$ plane) perpendicular to the direction x_3 . We will assume that the topological insulator lies at $x_3 < 0$. This means that for $x_3 < 0$, and far from the surface, $\theta(x_3) \rightarrow \pi$, while in the trivial vacuum, and also far from the surface, $\theta(x_3) \rightarrow 0$. We will assume that the change of θ from 0 to π occurs on a short distance ξ . We will call this configuration an axion domain wall. In the region where $\theta(x_3)$ is changing, $|x_3| \lesssim \xi$, an applied uniform magnetic field \mathbf{B} induces a uniform electric field \mathbf{E} parallel to \mathbf{B} whose magnitude is proportional to the change in θ . This is the topological magnetoelectric effect [135].

A similar striking effect is obtained by considering the case of a magnetic monopole of magnetic charge $2\pi/e$ (as required by Dirac quantization) inside a sphere of the trivial region of radius R , with $\theta = 0$, surrounded by a region with $\theta \neq 0$. The two regions are separated by a thin (axion) wall in which θ changes for 0 to π in a narrow shell of thickness $\xi \ll R$. The equations of axion electrodynamics imply that the magnetic monopole induces an electric charge Q_e on the surface of the sphere

$$Q_e = e \frac{\Delta\theta}{2\pi} \quad (361)$$

Thus, a magnetic monopole acquires an electric charge and becomes a *dyon*. This is the Witten effect [217]. In the particular case of a \mathbb{Z}_2 topological insulator, time reversal invariance requires that $\Delta\theta = \pi$, which implies that a monopole with unit magnetic charge has an electric charge $e/2$. A similar argument implies that an external magnetic field perpendicular to the open surface of the \mathbb{Z}_2 topological insulator (which is essentially an axion domain wall) induces an *electric charge polarization* on the surface proportional to the total magnetic flux, which is another manifestation of the topological magnetoelectric effect.

The reader should readily recognize that the result of Eq.(361) for the electric charge induced by the magnetic monopole is the same as the Goldstone-Wilczek equation for the fractional charge for a one dimensional soliton of Eq.(215). We will now see that this is not just an analogy. We will follow here the general approach of Curtis Callan and Jeffrey Harvey [218] who extended the earlier work of Goldstone and Wilczek [141]. Let us consider the bulk of a \mathbb{Z}_2 topological insulator and assume that $\theta(x)$ is slowly varying. We can use the triangle diagram calculation to find that an electromagnetic gauge field A_μ induces a current $\langle J_\mu(x) \rangle$ given by

$$\langle J_\mu(x) \rangle = -i \frac{e}{16\pi^2} \epsilon_{\mu\nu\lambda\rho} \frac{\phi^*(x) \partial^\nu \phi(x) - \phi(x) \partial^\nu \phi^*(x)}{|\phi(x)|^2} F^{\lambda\rho}(x) = \frac{e}{8\pi^2} \epsilon_{\mu\nu\lambda\rho} \partial^\nu \theta(x) F^{\lambda\rho}(x) \quad (362)$$

This result implies that, in the case of a domain wall in the $x_1 - x_2$ plane, a magnetic field perpendicular to the wall induces a current towards the wall and, hence, a charge accumulation on the wall. Where does this charge come from? To understand this problem we will consider a \mathbb{Z}_2 topological insulator occupying a slab of macroscopic size L between a wall at $x_3 = 0$ and a far way “anti-wall” at $x_3 = L$. In this configuration a magnetic field normal to the wall(s) induces a transfer of charge from one wall to the other. Similarly, an electric field *parallel* to the wall induces a current also parallel to the wall and *perpendicular* to the electric field, i.e. a Hall current.

In 1+1 dimensions we saw that soliton configurations acquire a fractional charge associated to states bound to the soliton (which are zero modes when $\Delta\theta = \pi$). We will see that the surfaces

of the 3D \mathbb{Z}_2 topological insulators also have zero modes which are Weyl fermions propagating on the wall. To see how this works we will consider a 3+1 dimensional Dirac fermion with a Dirac mass that changes sign at $x_3 = 0$. The Lagrangian now will be

$$\mathcal{L} = \bar{\psi} i\partial\psi + g\phi(x)\bar{\psi}\psi \quad (363)$$

where $\phi(x)$ is now a *real* scalar field that has the asymptotic behaviors $\phi(x_3) = \phi_0 f(x_3)$ such that $\lim_{x_3 \rightarrow \pm\infty} f(x_3) = \pm 1$. We will assume that $f(x_3)$ is a monotonous function of x_3 , but its actual dependence on x_3 is immaterial aside from the requirement that it should change sign at some point which we will take to be $x_3 = 0$. This is a special case of Eq.(350) with $\phi_1 = \phi$ and $\phi_2 = 0$. We will now recognize that this is just a Dirac fermion with a position-dependent Dirac mass $m(x_3) = g\phi(x_3)$. The one-particle Dirac Hamiltonian for this system is

$$H = -i\alpha \cdot \nabla + m(x_3)\beta \quad (364)$$

where α and β are the four 4×4 Dirac matrices. By symmetry, this Hamiltonian can be split into two Hamiltonians, $H = H_{\text{wall}} + H_{\perp}$,

$$H_{\text{wall}} = -i\alpha_1\partial_1 - i\alpha_2\partial_2, \quad H_{\perp} = -i\alpha_3\partial_3 + m(x_3)\beta \quad (365)$$

Let the spinor ψ_{\pm} be an eigenstate of the anti-hermitian Dirac matrix $\gamma_3 = \beta\alpha_3$ with eigenvalues $\pm i$, respectively

$$\gamma_3\psi_{\pm} = \pm i\psi_{\pm} \quad (366)$$

We seek a spinor solution ψ_{\pm} of the Dirac equation Eq.(364) such that $H_{\perp}\psi_{\pm} = 0$

$$\pm \partial_3\psi_{\pm} + m(x_3)\psi_{\pm} = 0 \quad (367)$$

which is also a solution of

$$(i\gamma_0 - i\gamma_1\partial_1 - i\gamma_2\partial_2)\psi_{\pm} = 0 \quad (368)$$

In other words, it is a solution of the massless Dirac equation in 2+1 dimensions. The requirement that ψ_{\pm} be an eigenstate of γ_3 reduces the number of spinor components from four to two. The full solution has the form

$$\psi_{\pm} = \eta_{\pm}(x_0, x_1, x_2)F_{\pm}(x_3), \quad \pm \partial_3 F_{\pm}(x_3) = -m(x_3)F_{\pm}(x_3) \quad (369)$$

For a domain wall with $\lim_{x_3 \rightarrow \infty} m(x_3) = +m$, the normalizable solution is $F_+(x_3)$ is

$$F_+(x_3) = F(0) \exp\left(-\int_0^{\infty} dx'_3 m(x'_3)\right) \quad (370)$$

where $F(0)$ is a constant. For the anti-domain wall, for which $\lim_{x_3 \rightarrow \infty} m(x_3) = -m$, the normalizable solution is $F_-(x_3)$.

We conclude that there is a 2+1-dimensional massless Dirac theory that describes the quantum states bound to the wall which propagate along the wall. These states are the generalization of the fractionally charged mid-gap states of solitons in one dimension discussed in section 8.3. The energy of these states is $E(\mathbf{p}) = \pm|\mathbf{p}|$, where $\mathbf{p} = (p_1, p_2)$ (where I set the Fermi velocity to unity). Experimental evidence for 2+1 dimensional massless Dirac fermions on the surface of the 3D \mathbb{Z}_2 topological insulator Bi_2Te_3 has been found in spin-polarized angle-resolved photoemission studies of the surface states which showed that they have a linear energy-momentum relation (expected of massless Dirac fermions) as well as the spin-momentum locking characteristic of these spinor states [219].

Having succeeded in showing that the surface of the \mathbb{Z}_2 time-reversal invariant 3D topological insulator has a two-component massless Dirac spinor we now want to determine its electromagnetic response. In fact we have already discussed this problem in our discussion of the parity anomaly in section 10.1.6 where we showed that the effective action for the electromagnetic field A_{μ} of Eq.(342) of a single massless Dirac bi-spinor is a Chern-Simons term with a prefactor which is 1/2 of the allowed value. In that purely 2+1-dimensional context we saw that time reversal invariance is actually broken. However, the 3D problem is time-reversal invariant so there must be a contribution that cancels this time-reversal anomaly. The answer is that the requisite cancellation is supplied by the bulk.

To see how this can happen we return to the bulk effective action of the 3D topological insulator of Eq.(359) where we observed that the θ -term is a total derivative. Suppose that the

system has a boundary at $x_3 = 0$ and that the topological insulator exists for $x_3 > 0$ (where the fermion mass is negative, $m < 0$) and trivial for $x_3 < 0$ (where $m > 0$). Only the region with $m < 0$ contributes to the θ -term. The region of four-dimensional space time occupied by the topological insulator is \mathcal{M} and in this region the θ -term becomes

$$S_\theta[A] = \frac{\theta}{32\pi^2} \int_{\mathcal{M}} d^4x \epsilon_{\mu\nu\rho\lambda} F^{\mu\nu} F^{\rho\lambda} = \frac{\theta}{8\pi^2} \int_{\mathcal{M}} d^4x \epsilon_{\mu\nu\rho\lambda} \partial^\mu A^\nu \partial^\lambda A^\lambda = \frac{\theta}{8\pi^2} \int_{\partial\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (371)$$

where $\partial\mathcal{M} \equiv \Sigma \times \mathbb{R}$ is the boundary of the region \mathcal{M} , Σ is the surface of the 3D topological insulator and \mathbb{R} is time.

Thus we see that the θ -term of the effective action $S_\theta[A]$ integrates to the boundary where it has the form of a 2+1-dimensional Chern-Simons term. Since for a time-reversal-invariant 3D topological insulator $\theta = \pi$, we see that in this case the boundary Chern-Simons term becomes

$$S_{\theta=\pi}[A] = \frac{1}{8\pi} \int_{\partial\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (372)$$

with a coefficient which is 1/2 of the allowed value. This bulk contribution either cancels the parity anomaly of the boundary state, rendering the full system time-reversal invariant. In other words, in this system time-reversal symmetry is realized by cancellation of the anomaly between the bulk of the topological insulator and the boundary or, equivalently, by an *inflow* of the parity anomaly between the boundary and the bulk of the system. Another way to phrase this result is the statement that a single Dirac fermion cannot exist on its own in 2+1 dimensions but it can as the boundary state of a 3+1-dimensional system whose anomaly cancels the anomaly of the boundary.

10.3. Chern-Simons Gauge Theory and The Fractional Quantum Hall Effect

We will now turn to the problem of the quantum Hall effects. This is a problem that revealed the existence of profound and far reaching connections between condensed matter physics, quantum field theory, conformal field theory and topology. In particular, the *fractional* quantum hall effect is the best studied and best understood topological phase of matter. As such, it has become the conceptual springboard for its manifold generalizations.

The integer (IQHE) and fractional (FQHE) quantum Hall effects are fascinating phenomena observed in fluids of electrons in two dimensions in strong perpendicular magnetic fields. The integer quantum Hall effect was discovered by Klaus von Klitzing in 1980 [220] in transport measurements of the longitudinal and Hall resistivity of the surface states of metal oxide field-effect transistors (MOSFET) in magnetic field of up to 15 Tesla. The effect that von Klitzing discovered (for which he was awarded the 1985 Nobel Prize in Physics) was that in the high field regime the measured Hall conductivity showed a series of sharply defined plateaus at which took the values $\sigma_{xy} = ne^2/h$, where n is an integer, and the longitudinal conductivity appeared to vanish $\sigma_{xx} \rightarrow 0$ as the temperature was lowered down to $T \simeq 1.5$ K. Remarkably the measured value of the Hall conductivity was obtained with a precision of $\sim 10^{-9}$. To this date the measurement of the Hall conductivity in the IQHE yields the most precise definition of the fine structure constant.

Subsequent transport experiments in ultra-high-purity GaAs-AlAs heterostructures by Dan Tsui, Horst Störmer and Art Gossard found that, in addition to the IQHE, two-dimensional electron fluids in high magnetic fields exhibit the *fractional* quantum Hall effect meaning that there are equally sharply-defined plateaus of the Hall conductivity at the values $\sigma_{xy} = \frac{p}{q} \frac{e^2}{h}$, where p and q are co-prime integers [221]. Much as in the IQHE case, in the FQHE the longitudinal conductivity vanishes at low temperatures. It is important to note that both in the IQHE and in the FQHE the observed temperature dependence in the highest purity samples of the longitudinal conductivity is activated, $\sigma_{xx} \sim \exp(-W/T)$. The observed value of the energy scale W is the experimental estimate of an energy gap in the electron fluid in the quantum Hall states.

In addition to MOSFETS and GaAs-AlAs heterostructures both the IQHE and the FQHE have been seen in several other experimental platforms, particularly in graphene and other 2D materials [222, 223]. The explanation of these effects and of a panoply of startling consequences that were uncovered in the course of understanding this phenomenon is the focus of this section.

10.3.1. Landau levels and the Integer Hall effect

At some level the integer quantum Hall effect can be explained by the Landau quantization of the energy levels of a free charged moving in two dimensions in a perpendicular magnetic field

[224]. The Hamiltonian for a non-relativistic particle of charge $-e$ and mass M in a perpendicular magnetic field B is

$$H = \frac{1}{2M} \left(-i\hbar \nabla + i\frac{e}{c} \mathbf{A}(\mathbf{x}) \right)^2 \quad (373)$$

for a uniform perpendicular magnetic field $\mathbf{B} = B \hat{\mathbf{e}}_z = \nabla \times \mathbf{A}(\mathbf{x})$.

In the circular gauge the vector potential is $A_i = -\frac{1}{2}B\epsilon_{ij}x_j$. We will assume that the 2D plane has linear size L . The total magnetic flux is $\Phi = BL^2$ and we will assume that there is an integer number N_ϕ of magnetic flux quanta piercing the plane, $\phi = N_\phi\phi_0$, where $\phi_0 = hc/e$ is the flux quantum. In units such that $\hbar = e = c = 1$ the flux quantum is $\phi_0 = 2\pi$ and $\Phi = 2\pi N_\phi$.

In the presence of a magnetic field the components of the canonical momentum operator $\mathbf{p} = -i\hbar \nabla - \frac{e}{c} \mathbf{A}$ do not commute with each other,

$$[p_i, p_j] = i\frac{e\hbar}{c} B \epsilon_{ij} \quad (374)$$

This means that translations in two directions do not commute with each other. However, the components of the operator $\mathbf{k} = \mathbf{p}(-B)$ commute with \mathbf{p} (and hence with the one-particle Hamiltonian) but do not commute with each other: the commutator is the same as in Eq. (374). Since $[\mathbf{k}, H] = 0$ they act as symmetry generators of the group of *magnetic translations*. For arbitrary displacements \mathbf{a} and \mathbf{b} the translation operators $t(\mathbf{a}) = \exp(i\mathbf{a} \cdot \mathbf{k}/\hbar)$ (and similarly with \mathbf{b}) satisfy

$$t(\mathbf{a})t(\mathbf{b}) = \exp(i\mathbf{a} \times \mathbf{b} \cdot \mathbf{e}_z/\ell_0^2) t(\mathbf{b})t(\mathbf{a}) \quad (375)$$

Magnetic translations only commute with each other if the area subtended by \mathbf{a} and \mathbf{b} contains an integer number of magnetic flux quanta.

Given the rotational symmetry of the circular gauge it is natural to work in complex coordinates $z = x_1 + ix_2$. We will also use the notation $\partial_z = (\partial_1 - i\partial_2)/2$ and $\partial_{\bar{z}} = (\partial_1 + i\partial_2)/2$. In this gauge, up to a normalization, the eigenstate wave functions have the form

$$\psi(z, \bar{z}) = f(z, \bar{z}) \exp\left(-\frac{|z|^2}{4\ell_0^2}\right) \quad (376)$$

where $\ell_0 = \sqrt{\frac{hc}{e|B|}}$ is the magnetic length. In this gauge (and in complex coordinates) the angular momentum operator $L_z = -i\hbar(x_1\partial_2 - x_2\partial_1) = \hbar(z\partial_z - \bar{z}\partial_{\bar{z}})$.

Any analytic function $f(z)$ is an eigenstate with energy $E_0 = \frac{1}{2}\hbar\omega_c$. A complete basis of analytic functions are the monomials $f_n(z) = z^n$ and have energy E_0 and angular momentum $L_z = n\hbar$. This is the lowest Landau level whose wave functions are $\psi_n(z) = z^n \exp(-|z|^2/4\ell_0^2)$. On the other hand, an anti-analytic function $f_N = \bar{z}^N$ is an eigenstate of energy $E_N = \hbar\omega_c(N + \frac{1}{2})$, where $\omega_c = \frac{e|B|}{Mc}$ is the cyclotron frequency, and angular momentum $L_z = -N\hbar$. States with angular momentum $n\hbar$ have the same energy and the degeneracy is equal to the number of flux quanta N_ϕ . For the most part we will be interested in the states in the lowest Landau level.

In the absence of disorder the Landau levels have an extensive degeneracy equal to the number of flux quantum N_ϕ . If we consider a system of N electrons in a Landau level the natural measure of density is not the areal density $\rho = \frac{N_e}{L^2}$ but the fraction $v = \frac{N_e}{N_\phi}$ the states in the Landau level which are occupied by electrons. The many-body state in which all N_ϕ states of the lowest Landau level has filling fraction $v = 1$. The wave function for this state is the Slater determinant of the Landau states in the $m = 0$ level. After some simple algebra the wave function of this state is found to be

$$\Psi_{v=1}(z_1, \dots, z_{N_e}) = \prod_{i < j} (z_i - z_j) \exp\left(-\sum_{i=1}^{N_e} \frac{|z_i|^2}{4\ell_0^2}\right) \quad (377)$$

This is the ground state of the non-interacting system and it is non-degenerate. It is easy to see that this state has a finite energy gap. Indeed, in the Hilbert space with a fixed number of electrons, the lowest energy excitation is a particle-hole pair in which the particle is in an unoccupied state of the first excited Landau level, with $m = 1$ and energy $E_1 = \frac{3}{2}\hbar\omega_c$ and a hole a single-particle state in the lowest Landau level, with energy $E_0 = \hbar\omega_c$. The excitation energy of the electron-hole pair is just $\hbar\omega_c$ which is finite. However, in the free particle system this excited states has a huge degeneracy as the particle and the hole can be in any single particle state of the Landau level.

At a very naive level one can estimate the Hall conductivity for a translationally invariant system filling up n Landau level by the following simple argument. If n Landau levels are filled, the number of electrons $N = nN_\phi$ and the total charge is then $Q = eN = e n N_\phi = e n BL^2/\phi_0 = ne^2BL^2/hc$. If a weak and uniform in-plane electric field \mathbf{E} is applied in the presence of a perpendicular magnetic field $\mathbf{B} = B \mathbf{e}_z$, then the entire system (its center of mass) moves at the drift velocity \mathbf{v} such that $\mathbf{v} \times \mathbf{B} = -Ec$. Thus, there is a Hall current $\mathbf{J} = Q\mathbf{v}$. The current density is $\mathbf{j} = \mathbf{J}/L^2 = Q\mathbf{v}/L^2$. Hence, $j_i = \frac{Qc}{BL^2} \epsilon_{ij} E_j$, which implies that the Hall conductivity is $\sigma_{xy} = Qc/BL^2 = ne^2/h$.

While the above argument is formally correct and yields the correct value of the Hall conductivity in this highly idealized setting, at a very basic level it is faulty. To begin with, it assumes exact translational invariance and, hence, Galilean symmetry, which are not obeyed in any realistic system. The second objection is that, even in this idealized system, as the number of electrons is varied, the chemical potential (and hence the Fermi energy) jumps discontinuously from one Landau level to the next resulting in a linearly increasing Hall conductivity (as predicted classically) without any of the observed plateaus. Furthermore, this argument does not explain which the Hall conductivity is so precisely quantized whereas, a priori, one would expect that being a transport coefficient the Hall conductivity would depend on lots of complicated materials details. But, it turns out that it does not!

Let us first discuss the observed *universality* of the Hall conductivity, namely its robustness and independence of microscopic details. In a remarkable paper Qian Niu, David Thouless and Yong-Shi Wu showed that, provided the ground state is separated by a finite energy gap from the excited states, the Hall conductivity is actually a topological invariant [225]. The argument has a similarity with what we discussed in the case of the Hall effect on a 2D lattice (in section 10.1.4) but in this more general system one considers the full many-body wave function, rather than the one-particle states. To show that this is true they considered a system of N electrons with periodic boundary conditions, i.e. on a two-dimensional torus, $T^2 = S_1 \times S_1$, with N_ϕ flux quanta going through the torus. The torus is a rectangle of dimensions L_1 and L_2 with opposite ends identified. This is necessary to allow for a Hall current to be globally allowed. The electromagnetic gauge field for a system on a torus with uniform non-vanishing flux cannot obey periodic boundary conditions but rather across the torus the gauge fields must differ by (large) gauge transformations

$$A_1(x_1, x_2 + L_2) = A_1(x_1, x_2) + \partial_1 \beta_2(x_1, x_2), \quad A_2(x_1 + L_1, x_2) = A_2(x_1, x_2) + \partial_2 \beta_1(x_1, x_2) \quad (378)$$

while the wave functions themselves obey twisted boundary conditions

$$\begin{aligned} \Psi([\mathbf{x}^{(j)} + L_1 \mathbf{e}_1]) &= \exp\left(-i \frac{e}{\hbar c} \sum_{j=1}^{N_e} \beta_1([\mathbf{x}^{(j)}]) + i\theta_1\right) \times \Psi([\mathbf{x}^{(j)}]) \\ \Psi([\mathbf{x}^{(j)} + L_2 \mathbf{e}_2]) &= \exp\left(-i \frac{e}{\hbar c} \sum_{j=1}^{N_e} \beta_2([\mathbf{x}^{(j)}]) + i\theta_2\right) \times \Psi([\mathbf{x}^{(j)}]) \end{aligned} \quad (379)$$

where \mathbf{e}_j (with $j = 1, 2$) are two unit vectors on the torus, and θ_1 and θ_2 are two angles that *twist* the boundary conditions of the wave functions.

In order to have a current on the torus they assumed that, in addition to the uniform flux going through the torus, there a weak uniform electric field \mathbf{E} on the torus represented by the constant in space vector potential $\delta \mathbf{A} = \mathbf{E} t = \nabla[U(\mathbf{x})t]$ whose circulation on the two non-contractible circles Γ_1 and Γ_2 , which wrap around the directions x_1 and x_2 of the torus T^2 , are

$$I_j = \oint_{\Gamma_j} \delta \mathbf{A} \cdot d\mathbf{x} = t \oint_{\Gamma_j} \mathbf{E} \cdot d\mathbf{x} = it E_j L_j \quad (380)$$

Line integrals of a gauge field on non-contractible loops in space (or space-time) are called *holonomies*. By inspection we see that the angles θ_1 and θ_2 are given by

$$\theta_j = \frac{e}{\hbar c} I_j \quad (381)$$

Alternatively, the angles $\boldsymbol{\theta} = (\theta_1, \theta_2)$ can be interpreted as magnetic fluxes (in units of the flux quantum ϕ_0) through the two non-contractible circles of the torus T^2 .

The angles $\boldsymbol{\theta}$ are defined mod 2π since they are phase factors that twist the phase of the wave functions, and define a 2-torus of boundary conditions. By comparing with what we did in

section 10.1.4 in the case of the single-particle wave functions in the lattice model, we see that the *many-body wave function* Ψ_θ and the twist angles θ is the same as the relation between the phase of the Hofstadter wave functions with the momentum \mathbf{k} of the magnetic BZ (which is also a 2-torus). Indeed, as it is always the case, while the phase of (in this case) the many-body wave function Ψ_θ is arbitrary, the changes of this phase as the the twist angle θ is changed are not. To quantify this dependence, for a given many-body state $\Psi_\theta^{(\alpha)}$, where α labels the state, we define a Berry connection $\mathcal{A}^{(\alpha)}(\theta)$ on the torus of boundary conditions θ

$$\mathcal{A}^{(\alpha)}(\theta) = i \langle \Psi_\theta^{(\alpha)} | \partial_\theta | \Psi_\theta^{(\alpha)} \rangle \quad (382)$$

Under a redefinition of the phase of the state

$$\Psi_\theta^{(\alpha)} \rightarrow \exp(if(\theta)) \Psi_\theta^{(\alpha)} \quad (383)$$

the Berry connection $\mathcal{A}^{(\alpha)}(\theta)$ changes by a gauge transformation

$$\mathcal{A}^{(\alpha)}(\theta) \rightarrow \mathcal{A}^{(\alpha)}(\theta) - \partial_\theta f(\theta) \quad (384)$$

Niu, Thouless and Wu [225] showed that the expression of the Kubo formula for the Hall conductivity of the state $\Psi_\theta^{(\alpha)}$, averaged over the boundary conditions θ , is given in terms of the flux of the Berry connection $\mathcal{A}^{(\alpha)}(\theta)$ through the 2-torus of boundary conditions, by the gauge-invariant quantity

$$\langle (\sigma_{xy})_\theta \rangle_\theta = \frac{e^2}{\hbar} \int_0^{2\pi} \frac{d\theta_1}{2\pi} \int_0^{2\pi} \frac{d\theta_2}{2\pi} (\partial_1 \mathcal{A}_2^{(\alpha)} - \partial_2 \mathcal{A}_1^{(\alpha)}) = \frac{e^2}{h} \frac{1}{2\pi} \oint_\gamma \mathcal{A}^{(\alpha)}(\theta) \cdot d\theta \quad (385)$$

where γ is the square contour with corners at $(0, 0)$, $(2\pi, 0)$, $(0, 2\pi)$ and $(2\pi, 2\pi)$, that defines the 2-torus of boundary conditions θ . This result is known as the Niu-Thouless-Wu formula.

Eq.(385) implies that for the Hall conductivity to be non-vanishing the Berry connection $\mathcal{A}^{(\alpha)}(\theta)$ must have a non-vanishing flux through the torus of boundary conditions. For this to be true, just as we did in section 10.1.3, we must consider large gauge transformations of the form

$$\begin{aligned} \mathcal{A}_1(\theta + 2\pi e_2) &= \mathcal{A}_1(\theta) + \partial_1 f_2(\theta) \\ \mathcal{A}_2(\theta + 2\pi e_1) &= \mathcal{A}_2(\theta) + \partial_2 f_1(\theta) \\ \Psi^{(\alpha)}(\{\mathbf{x}^{(j)}\}; \theta + 2\pi e_1) &= \exp(if_1(\theta)) \Psi^{(\alpha)}(\{\mathbf{x}^{(j)}\}; \theta) \\ \Psi^{(\alpha)}(\{\mathbf{x}^{(j)}\}; \theta + 2\pi e_2) &= \exp(if_2(\theta)) \Psi^{(\alpha)}(\{\mathbf{x}^{(j)}\}; \theta) \end{aligned} \quad (386)$$

where e_1 and e_2 are two orthogonal unit vectors on the torus of boundary conditions. We can now repeat the analysis we did in section 10.1.3 which, in this context, implies that the wave function $\Psi_\theta^{(\alpha)}$ cannot be globally well defined on the torus of boundary conditions, where it must have zeros, and where it should be defined in patches. The end result is that the wave functions labeled by α are classified by a *topological invariant*, the first-valued Chern number $C_1^{(\alpha)}$, which here it is given by

$$C_1^{(\alpha)} = \frac{1}{2\pi} \oint_\gamma \mathcal{A}^{(\alpha)}(\theta) \cdot d\theta \quad (387)$$

and, hence, that the Hall conductivity in the state α (averaged over the twisted boundary conditions) exhibits the integer quantum Hall effect,

$$\langle (\sigma_{xy})^{(\alpha)} \rangle_\theta = C_1 \frac{e^2}{h} \quad (388)$$

Expressing the Hall conductivity in terms of the first Chern number, which is a topological invariant, proves that the value of the conductivity cannot be changed by local physics effects, such as disorder, etc. The topological nature of the Hall conductivity is the reason for the robustness and high precision of the measured value of the Hall conductivity. In subsequent work Niu and Thouless showed that, provided the state has a finite energy gap to all excitations, in the thermodynamic limit the Hall conductivity averaged over boundary conditions and for a given boundary condition are equal.

The result of Eq.(388) seemingly implies that there can only be an *integer* quantum Hall effect which, as we see shortly, it is not the case: there is a fractional quantum Hall effect in

which the Hall conductivity is a *fraction*, $\sigma_{xy} = \frac{p}{q} \frac{e^2}{h}$, where p and q are co-prime integers. This value of the Hall conductivity is also found experimentally to be obeyed with the same precision as in the integer quantum Hall effect. In other words, the Hall conductivity in the fractional case must also have a topological character. To understand why this can be the case we need an implicit assumption that we made in our derivation. In fact, in deriving the result of Eq.(388) we assumed (implicitly) that for each value of θ there is only one many body state $\Psi_\theta^{(\alpha)}$ (up to gauge transformations). As we will see below, in the *fractional* quantum Hall effect on a torus (and, in fact on any closed surface except the sphere) the ground state is degenerate in the thermodynamic limit. We will also see that traversing the torus of boundary conditions once maps one degenerate state to another one. In general, we will find that on the 2-torus in the thermodynamic limit there are $m \in \mathbb{Z}$ exactly degenerate states. In this case, one returns to the original state after sweeping m times the torus of boundary conditions. We will also see below that the degenerate states are labeled by quantum numbers related to the *anyons* that they support.

While the topological argument proves the robustness (and universality) of the value of the Hall conductivity, it does not provide an insight for why there are plateaus in its magnetic field dependence. If for some value of the filling fraction $\nu = N/N_\phi$ the electron fluid is exactly at a ground state $\Psi^{(\alpha)}$ upon changing either the number of particles N or the magnetic field B (but not both) the fluid will now be at the state $\Psi^{(\alpha)}$ plus or minus some number of electrons. The existence of the plateau in σ_{xy} can be understood if the extra electrons (or holes) do not contribute to the Hall conductivity. This happens in the presence of disorder as these extra particles will be localized by the disorder and localized states, whose wave functions decay exponentially are insensitive to boundary conditions and, hence, these states do not contribute to the conductivity. This argument was first put forth by Laughlin [226] who build on the result that in a disordered system in two dimensions all states are localized [63] but made the key assumption that the state at the center of the Landau level (broadened by disorder) is extended and contributes to the conductivity if this state is filled. This picture implies that this is actually a quantum phase transition from an insulator (dubbed a Hall insulator) to a state with a quantized Hall conductivity.

This intuitive and appealing proposal has been the focus of research for many years and it is largely an unsolved problem. There is in fact a proposed field theory for this quantum phase transition based on a non-linear sigma model (in replica space) with a topological θ -term [227, 228], analogous to the theory we discussed in section 5.2 for quantum antiferromagnets in 1+1 dimensions. in this proposal the (Boltzmann) longitudinal conductivity σ_{xx} plays the role of the inverse of the coupling constant of the non-linear sigma model while σ_{xy} plays the role of the coefficient of the θ -term. Although the RG flows suggested by this approach qualitatively agree with currently existing experimental data, and actual derivation is still wanting.

However, aside from the fact that is is still an unsolved problem, there is the problem that this explanation ignores the fact that in a Landau level interactions are dominant and are the key to the understanding of the fractional case even in the same samples! A symptom of this problem is that, at least in the cleaner samples, as we noted above the longitudinal conductivity vanishes exponentially with temperature. This dependence means that there is a clean energy gap (and not just a mobility gap) which suggests that the state with additional excitations may have some sort of at least local order which would provide for a local energy gap. An actual theory based on this picture has yet to be developed.

10.3.2. The Laughlin Wave Function

We will turn now to the *fractional* quantum Hall effect (FQHE). The FQHE is observed in fractionally filled Landau levels. These states have fractional filling fractions, namely that $\nu = \frac{N_e}{N_\phi}$ is a fraction. Most of the observed states are in the lowest Landau level, with $N = 0$, but a few states with very intriguing properties are seen in the first excited Landau Level with $N = 1$.

Let us consider for now the states in the lowest Landau level, $N = 0$. We will assume that there is no disorder and, for now, we will ignore the effects of boundaries of the sample. Physically, the important interaction here is the Coulomb interaction although, in some samples with screening layers, short-range interactions may be important as well. Since there is a large magnetic field, the Zeeman interaction is typically a large energy scale and the electrons are expected to be fully polarized. However, there are situations under which the gyromagnetic factor can be made small (and even zero) and spin unpolarized states have to be considered. In some platforms, such as graphene, additional quantum numbers are present, such as “flavor” associated with the different Dirac states. In our discussion we will not cover these richer possibilities.

Under these circumstances we have a problem in which all N_ϕ available one-particle states are

exactly degenerate (the Landau level is “flat”) but only a fraction of them are filled. In this limit, this system as as strongly interaction as it gets. Not surprisingly, in a system of this type time-honored standard approximations fail, such as Hartree-Fock. Given the macroscopic degeneracy and the nature of the states in a magnetic field, a system of this type may harbor many different types of actual ground states depending of the types of interactions that are considered.

Tsui, Störmer and Gossard [221] did transport experiments in the two-dimensional electron gas (2DEG) in high-purity GaAs-AlAs heterostructures and reported the discovery of a state with a highly precise value of the Hall conductivity of $\sigma_{xy} = \frac{1}{3} \frac{e^2}{h}$. Such a state cannot be understood in any obvious way in terms of the integer quantum Hall effect. The observation of a clean gap in the low temperature longitudinal resistivity $\sigma_{xx} \rightarrow 0$ as $T \rightarrow 0$ implied that the 2DEG is incompressible and quite likely uniform.

The first breakthrough was Laughlin’s unique insight which led to his explanation of the experiments in terms of a novel quantum liquid of fully spin polarized electrons at filling fractions $\nu = \frac{1}{m}$ (with m and odd integer) whose wave function he proposed to be

$$\Psi_m(z_1, \dots, z_N) = \prod_{i < j} (z_i - z_j)^m \times \exp\left(-\sum_{i=1}^{N_e} \frac{|z_i|^2}{4\ell_0^2}\right) \quad (389)$$

where $N = N_e$ is the number of electrons, $\{z_i\}$ are their (complex) coordinates and m is an odd integer (which makes the wave function antisymmetric, as required by Fermi statistics).

Laughlin extracted the physics encoded in this wave function by considering the probability $|\Psi_m(z_1, \dots, z_N)|^2$ of finding the N electrons in the locations $\{z_i\}$ (with $i = 1, \dots, N$). The norm of the Laughlin wave function is

$$\|\Psi_m\|^2 = \int d^2 z_1 \dots d^2 z_N |\Psi_m(z_1, \dots, z_N)|^2 \quad (390)$$

The Laughlin wave function $\Psi_m(z_1, \dots, z_N)$ is an eigenstate of angular momentum with eigenvalue $L_m = \frac{1}{2}mN(N-1)$. This remarkable wave function has a large overlap ($\sim 98\%$) with the exact wave function with $V(R) = \frac{\epsilon^2}{\epsilon R}$ (Coulomb) interactions in systems of up to eight particles (ϵ is the dielectric constant).

The norm of the state can ten be interpreted as the classical partition function of a system of a system of N particles at the locations $\{z_i\}$ at inverse temperature $\beta = m$ with the effective interaction

$$U(z_1, \dots, z_N) = -2 \sum_{1 \leq i < j \leq N} \ln |z_i - z_j| + \frac{1}{2m\ell_0^2} \sum_{j=1}^N |z_j|^2 \quad (391)$$

This classical partition function is a one-component plasma, of a gas of particles with unit (negative) charge interacting with each other with the repulsive 2D classical (logarithmic!) Coulomb interaction, $V_{CG} = -\ln |z_i - z_j|$ (not $1/R!$). The last term represents the contribution of a neutralizing background positive charge (which here it originates in the gaussian factor of the wave function). In other words, this is a one-component plasma. We can check that this is correct since $\nabla^2 \left(\frac{|z|^2}{2m\ell_0^2} \right) = \frac{2}{m\ell_0^2}$ which corresponds to a uniform (areal) charge density $\rho_0 = \frac{1}{2\pi m\ell_0^2}$. Classical Monte Carlo simulations of this probability distribution showed that this wave function describes an incompressible fluid state if $m \lesssim 7$, while for larger values it represents a crystalline state. This approach is known as the plasma analogy.

10.3.3. Quasiholes have fractional charge

Laughlin considered a vortex-like excitation in the fluid created by the adiabatic insertion of an infinitesimally thin solenoid at some coordinate z_0 carrying a magnetic flux Φ , a *fluxoid*. In the presence of this solenoid the single particle state $z^n \exp(|z|^2/4\ell_0^2)$ acquires a branch cut and becomes the state $z^{n+\alpha} \exp(|z|^2/4\ell_0^2)$ where $\alpha = \Phi/\phi_0$, with $\phi_0 = hc/e$ being the flux quantum. This analytic structure is the Aharonov-Bohm effect [155].

However, if the solenoid carries just one flux quantum $\alpha = 1$ and the net effect is that the state becomes $z^{n+1} \exp(-|z|^2/4\ell_0^2)$ which has one more unit of angular momentum. For these reasons, in the presence of a solenoid with one flux quantum inserted at z_0 , the Laughlin wave function for the 2DEG in a large magnetic field now becomes

$$\Psi_m^{qh}(z_0; z_1, \dots, z_N) = \prod_{i=1}^N (z_i - z_0) \times \Psi_m(z_1, \dots, z_N) \quad (392)$$

A calculation using the plasma analogy reveals that the net effect of the solenoid is to expel a charge equal to $-e/m$ from the vicinity of the location of the solenoid or, equivalently, to create a *fractionally charged quasihole* at z_0 with charge $Q_{\text{qh}} = +e/m$. The charge distribution is uniform at long distances but is depleted (hence a quasihole) on a length scale ξ . One can check that this state costs a finite energy ε_0 which depends on the particular interaction.

On the other hand, if we consider a system in which the solenoid is inserted adiabatically so that after a long time t there is a quasihole at z_0 , an amount of charge equal to $-e/m$ will flow radially outwards to the outer edge of the 2DEG. The adiabatic insertion of the solenoid generates a azimuthal electric field (an e.m.f.) and a radial Hall current, and a Hall conductivity $\sigma_{xy} = \frac{1}{m} \frac{e^2}{h}$. Hence, the Laughlin wave function exhibits the fractional quantum Hall effect and its excitations are vortices with fractional charge e/m .

It is important to note the non-local nature of the Laughlin quasiholes. The non-locality is inherited from the fluxoid (the solenoid) inserted at z_0 : even though the magnetic field of the fluxoid vanishes away from it, $B(\mathbf{x}) = \Phi \delta^2(\mathbf{x} - \mathbf{x}_0)$, its vector potential $\mathbf{A}(\mathbf{x})$ does not since its circulation on any non-contractible loop γ that contains the location of the fluxoid inside must be equal to the flux Φ carried by the fluxoid. Although in principle one can choose a singular gauge in which $\mathbf{A}(\mathbf{x})$ vanishes locally, this cannot be true everywhere as it must leave behind a Dirac-like string on a curve Γ ranging from the location z_0 of the fluxoid to the boundary of the system with \mathbf{A} taking a singular value on Γ . This feature has the same form as the 2D vortex that we discussed in section 5.1.1. Kivelson and Roček [229] showed that the existence of this Dirac string causes the *phase* of a quantum state that carries charge e^* to have a branch cut on the curve Γ and to change by $\exp(i2\pi(e^*/e)\Phi/\phi_0)$, as required by the Aharonov-Bohm effect [155]. This effect is unobservable for integer-charged states, i.e. electrons.

10.3.4. The Jain States

The construction of the Laughlin quasiholes motivated Jainendra Jain [230, 231] to rewrite the Laughlin wave function in the form

$$\Psi_m(z_1, \dots, z_N) = \prod_{i < j} (z_i - z_j)^{m-1} \times \Psi_{v=1}(z_1, \dots, z_N) \quad (393)$$

Here the prefactor represents particles each attached with an even number, $m - 1$, of flux quanta. The second factor, $\Psi_{v=1}(z_1, \dots, z_N)$, is the wave function of fermions filling up a lowest Landau level, i.e. the Vandermonde determinant of Eq.(377). In this form, the inverse of the filling factor, $1/v$, which is the number of flux quanta per particle, is written as $\frac{1}{v} = (m - 1) + 1 = m$. The interpretation now is that the $(m - 1)\phi_0$ magnetic flux carried by each particle, which partially screens the uniform field down to an effective field $B_{\text{eff}} = B - (m - 1)\phi_0$ corresponding to one effective flux quantum per particle, i.e. a filled lowest Landau level of the effective field B_{eff} . In other words, the wave function for the $v = 1/m$ FQH state is reinterpreted as the $v_{\text{eff}} = 1$ *integer* QH state of N *composite fermions*, each being an electron carrying attached an even integer $m - 1$ flux quanta.

This procedure is known as flux attachment. Physically it means that the strong interactions between the electrons for other electrons to be further away which, in virtue of being in a strong magnetic field, is equivalent to an increase of their relative angular momentum, as if there was a local change in the magnetic field. There is no reason to believe that the composite fermions are weakly coupled, even though this is frequently stated in the literature without any real supporting evidence.

Jain then generalized this reinterpretation of the Laughlin state to a whole sequence of FQH states obtained by filling p levels of these partially screened magnetic fluxes

$$\Psi_{m,p}(z_1, \dots, z_N) = P_{LLL} \prod_{i < j} (z_i - z_j)^{m-1} \Psi_{v=p}(z_1, \dots, z_N) \quad (394)$$

where $\Psi_{v=p}(z_1, \dots, z_N)$ is the wave function of p filled Landau levels (of composite fermions), and P_{LL} is an operator that projects onto the lowest Landau level. By counting fluxes we see that the filling fractions for the Jain states satisfy

$$\frac{1}{v(m, p)} = m - 1 \pm \frac{1}{p} \quad (395)$$

where we allowed for the possibility that the external fluxes may be over-screened by the composite fermions. Alternatively we can write

$$\nu(m, p) = \frac{p}{p(m-1) + \pm 1} \quad (396)$$

The Jain states with $p = 1$ are the Laughlin states and each is called the primary state of a Jain sequence. Almost all the observed FQH states belong to one of these sequences (and its generalizations). In particular, the most prominent states (i.e. those with larger plateaus) belong to the sequence $\frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{4}{9}, \frac{5}{11}, \dots$ and to the reversed sequence $1, \frac{2}{3}, \frac{3}{5}, \frac{4}{7}, \dots$. For $p \rightarrow \infty$, the Jain sequences converge to the values of the filling fraction $\lim_{p \rightarrow \infty} \nu(m, p) = \frac{1}{m-1}$. In this limit, the $m-1$ fluxes attached to each particle exactly cancels the external flux leading us to a (possibly) Fermi liquid of composite fermions [232].

10.3.5. Quasiholes have fractional statistics

The non-locality and topological nature of the Laughlin quasihole implies that this state must be regarded as a soliton (or vortex) of the charged quantum fluid. The vortex-like state with wave function $\Psi_m^{qh}(z_0; z_1, \dots, z_N)$ is called the Laughlin quasihole. This state represents a composite object of a flux quantum and a fractional charge. This structure of this quantum state is strongly reminiscent of the flux-charge composite objects considered by Frank Wilczek who showed that such objects should be *anyons*, states that exhibit *fractional statistics* [156]. Although at a conceptual level anyons were proposed (almost) prior to the discovery of the FQHE, it is in this setting that fractional statistics entered actual physics. The actual experimental observation took work by many people, and was only confirmed in 2020 [233].

Fractional statistics dictates the analytic form of the wave functions of two or more quasiholes [234]. Let us consider a Laughlin-type wave function for two quasiholes located at (complex) coordinates u and w . Naively we expect the wave function for two quasiholes in the $\nu = 1/m$ Laughlin state to have the form

$$\Psi(u, w; z_1, \dots, z_N) = N(u, w) \prod_{j=1}^N (z_j - u)(z_j - w) \Psi_m(z_1, \dots, z_N) \quad (397)$$

The prefactor $N(u, w)$ must be chosen to account for the fact that there is an additional change of angular momentum due to the additional fluxoid at w . We also expect that as the quasihole with charge e/m is adiabatically carried around the other quasihole (also with charge e/m) there will be an accrued phase in the wave function due to the Aharonov-Bohm effect of the charge of one quasihole circling around the flux of the other (or, equivalently, crossing the branch cut) [229]. The requirement of translation invariance and analyticity are met by the choice [234] (ignoring a multiplicative normalization constant)

$$N(u, w) = (u - w)^{1/m} \exp\left(-\frac{1}{4\ell_0^2 m}(|u|^2 + |w|^2)\right) \quad (398)$$

which has a branch cut stretching from u to w . A calculation using the plasma analogy represents this state as a set of N charge -1 particles interacting with two additional particles of charge $-1/m$ at u and w (and a neutralizing background). The branch cut implies that dragging a quasihole around the other during a π rotation followed by a translation (i.e. an exchange), induces a monodromy in the wave with a jump in its phase of $\exp(\pm i\pi/m)$ (with the sign determined by the orientation of the monodromy) in such a way that the wave function changes by a phase factor

$$\Psi(u, w; z_1, \dots, z_N) \mapsto e^{\pm i\frac{\pi}{m}} \Psi(u, w; z_1, \dots, z_N) \quad (399)$$

In other words, the quasihole is an *anyon* with fractional statistics π/m . Daniel Arovas, J. Robert Schrieffer and Frank Wilczek [235] further refined this argument by showing that under such a slow adiabatic change the wave function of two quasiholes acquires a Berry phase which accounts for the fractional statistics. An explicit path-integral derivation of this effect, which uses the fact that the quasihole states are coherent states can be found in Ref. [9], chapter 13.

10.3.6. Hydrodynamic Effective Field Theory

We will now turn to the approaches to the FQHE using the methods of quantum field theory. As we will see the concept of flux attachment plays a key role. In section 9.5 we showed that

Chern-Simons gauge theory is in fact a theory of flux attachment. Thus we expect that Chern-Simons gauge theory should play a key role as well.

It is useful to ask first why should Chern-Simons gauge theory play a central role at least in the description of the low energy physics. There is a simple, yet powerful, phenomenological argument due to Jörg Fröhlich and Anthony Zee that shows why this should be the case [236]. This in essence a hydrodynamic argument. The fractional quantum Hall effect occurs in a 2DEG in the presence of a large magnetic field which breaks explicitly time reversal invariance. In the regime of the FQHE the 2DEG is a uniform charged incompressible fluid in which charge is conserved. This means that the charge 3-current $j_\mu = (j_0, \mathbf{j})$ (where j_0 is the charge density and \mathbf{j} is the charge current density) must be locally conserved,

$$\partial_\mu j^\mu = 0 \quad (400)$$

which is the continuity equation. The solution of this equation is that the conserved current can be written in terms of a *dual* (in the Hodge sense) vector field \mathcal{B}_μ such that

$$j_\mu = \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{B}^\lambda \quad (401)$$

The factor of $\frac{1}{2\pi}$ is introduced for later convenience. The current field j_μ is not changed by a smooth redefinition of the vector field $\mathcal{B}_\mu \rightarrow \mathcal{B}_\mu + \partial_\mu \Phi$, where $\Phi(x)$ is a non-singular field. This means that the vector field \mathcal{B}_μ is a gauge field.

The effective action of this theory is a functional of the current distribution j_μ and, hence, of the gauge field \mathcal{B}_μ . It should be gauge-invariant, and odd under parity and time-reversal. Since the fluid is incompressible and uniform, at long distances and low energies the action must be a local functional of the gauge field which should be at least Galilean invariant. In addition, the coupling of the fluid to an external electromagnetic probe field A_μ must be of the usual form $-e j^\mu A_\mu$. To lowest orders in derivatives, there the unique local and gauge-invariant Lagrangian density which is odd under time reversal (and parity) is the Chern-Simons theory

$$\mathcal{L}_{\text{eff}}[\mathcal{B}_\mu] = \frac{m}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu \partial^\nu \mathcal{B}^\lambda - \frac{1}{4g^2} \mathcal{F}_{\mu\nu}^2 - \frac{e}{2\pi} A^\mu \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{B}^\lambda + \dots \quad (402)$$

The first term is the Chern-Simons term of the Lagrangian. The coefficient m is a dimensionless integer which, as we saw in section 9.3 is required for the theory to be invariant on a closed manifold. The second is a Maxwell term. Here $\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{B}_\nu - \partial_\nu \mathcal{B}_\mu$ is the field strength tensor of the gauge field \mathcal{B}_μ . By power counting we see that the coupling constant g^2 has units of length $^{-1}$ and, hence this term is irrelevant at long distances and will be neglected. The third term is the coupling of the current j_μ to the electromagnetic field A_μ . Upon an integration by parts this term can be written as $\mathcal{B}_\mu \mathcal{J}^\mu$ where

$$\mathcal{J}_\mu = -\frac{e}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\nu A^\lambda \quad (403)$$

is a current minimally coupled to \mathcal{B}_μ .

The equation of motion of the (dynamical) gauge field \mathcal{B}_μ is

$$\frac{\delta \mathcal{L}}{\delta \mathcal{B}} = 0 \quad (404)$$

which yields the relation

$$\frac{m}{2\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu \partial^\nu \mathcal{B}^\lambda = \mathcal{J}_\mu \quad (405)$$

From the definition of the current \mathcal{J}_μ , Eq.(403), we see that the solution of the equation of motion of Eq.(405), up to a gauge transformation, is

$$m \mathcal{B}_\mu = -e A_\mu \quad (406)$$

By plugging this relation back into the Lagrangian for the gauge field \mathcal{B}_μ , Eq (402) (which is equivalent to integrate out the gauge field \mathcal{B}_μ) we find that the effective Lagrangian of the electromagnetic field A_μ is just a Chern-Simons term

$$\mathcal{L}_{\text{eff}}[A_\mu] = \frac{e^2}{4\pi m} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda + \dots \quad (407)$$

This result allows us to read-off the Hall conductivity of the fluid

$$\sigma_{xy} = \frac{1}{m} \frac{e^2}{2\pi\hbar} \quad (408)$$

where we restored units such that \hbar is not unity. In other words, this fluid exhibits the fractional quantum Hall effect for a fluid at filling fraction $\nu = 1/m$.

We should note that this heuristic argument applies equally to systems of fermions, for which m is odd, as well as to systems of bosons, for which m is even.

10.3.7. Composite Boson Field Theory

We will now discuss two field-theoretic approaches to the FQHE. These approaches use the concept of flux attachment as a mapping of fermions to bosons and as a mapping of fermions to fermions. These approaches are a form of duality transformation which engineers a form of statistical transmutation. We will first show that Chern-Simons theory can be used to effect such a mapping. That this is possible should not be surprising since, as we saw in section 9.7, Chern-Simons theory is a theory of fractional statistics and, hence, of anyons.

We should make some important comments on both approaches before we get into a detailed description. While the mapping of fermions to composite bosons and fermions to composite bosons is correct, their approximate mean field descriptions violate symmetries of the 2DEG in a magnetic field. At the root level is the fact that the flux attachment is local in space time and approximate descriptions of these theories bring about a large amount of Landau level mixing, even in the large field limit. For instance, at the mean field theory level the composite boson approach is equivalent to a Bose-Einstein condensate which is invariant under translations and under time reversal. In this picture the breaking of time reversal is present in the coupling to a Chern-Simons gauge field. Likewise, the mean field theory of the composite fermion theory is a problem of composite fermions in a partially screened. While a partially screened magnetic field still breaks time reversal, it does not behave properly under magnetic translations. As we will see, these problems will be solved, at least in the low energy regime, by quantum fluctuations which restore the symmetries. In both cases, in the fractional states one recovers an effective topological field theory which captures the universal physics of these states. Needless to say it, both approaches do poorly in the computation of non-universal dimensionful quantities such as energy gaps, etc. These difficulties become very severe in the compressible states whose low energy behavior is not topological.

Unlike the wave function approaches which project these states into a specific Landau level (usually the lowest), the field theoretic approach does not effect such projection. Several papers have been written attempting to do a field theory projected into a Landau level. These approaches transformed the problem into quantum field theory on a non-commutative plane, which is inherent the nature of the states in a magnetic field. Although some significant progress has been made in this direction, these theories remain poorly understood [237, 238, 239, 240, 241, 242, 243]

In this section we will focus on the composite boson theory. Let us consider a theory in 2+1 dimensions with two dynamical abelian gauge fields \mathcal{A}_μ and \mathcal{B}_μ , whose Lagrangian is a sum of a Chern-Simons term at level k and a BF term:

$$\mathcal{L} = \frac{k}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{A}^\lambda + \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{B}^\lambda \quad (409)$$

The equation of motion for the field \mathcal{A}_μ is

$$\frac{k}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{A}^\lambda + \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{B}^\lambda = 0 \quad (410)$$

whose solution (up to a gauge transformation) is $k\mathcal{A}_\mu = -\mathcal{B}_\mu$. Plugging this relation into the Lagrangian of Eq.(409) we find that the effective Lagrangian for the field \mathcal{B}_μ is

$$\mathcal{L}[\mathcal{B}_\mu] = -\frac{1}{4\pi k} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu \partial^\nu \mathcal{B}^\lambda \quad (411)$$

which states that exchanging $\mathcal{A}_\mu \leftrightarrow \mathcal{B}_\mu$ is equivalent to the “duality” $k \leftrightarrow -1/k$.

In section 9.7 we showed that particles coupled to a Chern simons gauge field at level k have fractional statistics $\exp(\pm i\pi/k)$. Hence, we see that particles coupled to the field \mathcal{B}_μ have fractional statistics $\exp(\pm i\pi k)$. In other words, for k odd this coupling amounts to exchanging a theory of fermions to a theory of bosons coupled to the gauge field \mathcal{B}_μ . This is a form of

bosonization. Alternatively, for k even it maps a theory of fermions to another theory of fermions coupled to a gauge field \mathcal{B}_μ . These observations have led to distinct, but ultimately equivalent description of the quantum Hall effect.

We will begin the approach of mapping fermions to bosons and use it in the problem of the FQHE. This is the Landau-Ginzburg theory (or *composite boson theory*) of Shoucheng Zhang, Hans Hansson and Steven Kivelson [244] (ZHK). In this approach the problem of fermions coupled to a magnetic field interacting with each other through a two-body potential (that we will assume is ultra with coupling constant λ , for simplicity) becomes the same problem but for a theory of bosons which are also coupled to a Chern-Simons gauge field, which we will denote by \mathcal{A}_μ . The Lagrangian density for this equivalent system of *composite bosons* is

$$\mathcal{L}_{CB} = \phi^*(x)[iD_0 + \mu]\phi(x) + \frac{1}{2M}|\mathbf{D}\phi(x)|^2 - \lambda(|\phi(x)|^4 + \frac{1}{4\pi m}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu\partial^\nu\mathcal{A}^\lambda) \quad (412)$$

where $x = (x_0, \mathbf{x})$ are the space-time coordinates, μ is the chemical potential, M is the mass of the fermions, and m is an arbitrary odd integer. The covariant derivative

$$D_\mu = \partial_\mu + i\frac{e}{\hbar c}A_\mu + i\mathcal{A}_\mu \quad (413)$$

which effects the minimal coupling of the complex scalar field $\phi(x)$ to the background electromagnetic field A_μ and to the Chern-Simons gauge field \mathcal{A}_μ , known in this context os known as the statistical gauge field.

ZHK showed that the FQH state can be thought as being closely related to to a phase in which the complex scalar field condenses, much as in the case of a superfluid even though the FQH *is not* a superfluid. To see how this works we will write

$$\phi(x) = \sqrt{\rho(x)} \exp(i\omega(x)) \quad (414)$$

The classical equations of motion of the theory of composite bosons, Eq.(412), are

$$\frac{\delta\mathcal{L}_{CB}}{\delta\phi^*(x)} = 0 \Rightarrow (iD_0 + \mu)\phi(x) - \frac{1}{2M}\mathbf{D}^2\phi(x) - 2\lambda|\phi(x)|^2\phi(x) = 0 \quad (415)$$

$$\frac{\delta\mathcal{L}_{CB}}{\delta\mathcal{A}_0(x)} = 0 \Rightarrow \frac{1}{2\pi m}\epsilon_{ij}\partial_i\mathcal{A}_j + |\phi(x)|^2 = 0 \quad (416)$$

$$\frac{\delta\mathcal{L}_{CB}}{\delta\mathcal{A}_i(x)} = 0 \Rightarrow \frac{1}{2\pi m}\epsilon_{i\alpha\beta}\partial^\alpha\mathcal{A}^\beta + \frac{i}{2M}[\phi^*(x)D_i\phi(x) - (D_i\phi(x))^*\phi(x)] = 0 \quad (417)$$

$$\frac{\delta\mathcal{L}_{CB}}{\delta\mu} = \rho_0 L^2 T \Rightarrow \int d^3x |\phi(x)|^2 = \rho_0 L^2 T \quad (418)$$

where ρ_0 is the areal density of electrons.

A uniform solution of Eqs. (415)-(418) with constant amplitude of the composite bosons, i.e. a uniform and static condensate $\bar{\phi}$, with uniform statistical gauge field strength $\bar{\mathcal{B}}$, requires that the there should be no current in the ground state and that each term of Eq.(417) should be zero. This condition implies that the external field is canceled by the average statistical field, $\frac{eB}{\hbar c} + \bar{\mathcal{B}} = 0$. This condition together with Eqs.(416) and (418) imply that

$$\rho_0 = \frac{1}{m}\frac{eB}{\hbar c} = \frac{1/m}{2\pi\ell_0^2} \quad (419)$$

and we conclude that the filling fraction is $\nu = \frac{1}{m}$, which are the Laughlin states. In addition we get that the chemical potential is $\mu = 2\lambda\rho_0$ and that $|\bar{\phi}|^2 = \rho_0$, with ρ_0 given in Eq.(419).

However, these mean field results seeming imply that this system is a Bose-Einstein condensate. To understand why this is incorrect, and to determine what it actually is, we need to go beyond the mean field theor that we just described and evaluate the effects of quantum fluctuations and write

$$\phi(x) = (\rho_0 + \delta\rho(x))^{1/2} \exp(i\omega(x)), \quad \frac{eA_\mu}{\hbar c} + \mathcal{A}_\mu = \delta\mathcal{A}_\mu \quad (420)$$

The partition function of this theory is a path integral over the fluctuating density fields $\delta\rho$, the Goldstone field ω and the fluctuation of the Chern-Simons gauge field $\delta\mathcal{A}_\mu$. Upon integrating out

the massive density fluctuations to lowest (quadratic) order we find that the effective Lagrangian for ω and $\delta\mathcal{A}_\mu$ is

$$\mathcal{L}_{\text{eff}}[\omega, \delta\mathcal{A}_\mu] = \frac{\kappa}{2} (\partial_0 \omega - \delta\mathcal{A}_0)^2 - \frac{\rho_s}{2} (\partial_i \omega - \delta\mathcal{A}_i)^2 + \frac{1}{4\pi m} \epsilon_{\mu\nu\lambda} \delta\mathcal{A}^\mu \partial^\nu \delta\mathcal{A}^\lambda + \dots \quad (421)$$

The first two terms of Eq.(421) is the effective Lagrangian of the Goldstone mode ω in the Bogoliubov theory of superfluidity a compressibility κ and a superfluid stiffness ρ_s given by

$$\kappa = \frac{1}{2\lambda}, \quad \rho_s = \frac{\rho_0}{M} = \frac{\nu}{2\pi} \hbar \omega_c \quad (422)$$

However, as we see, this is not a superfluid since the phase field is “eaten” by the Chern-Simons gauge field $\delta\mathcal{A}_\mu$ which now acquires a mass term. In this sort of Higgs mechanism the fluctuating gauge field has a massive longitudinal mode and a massive transverse mode. Thus, this state doe not have any massless modes.

To see that this theory describes the fractional quantum Hall effect we need to compute the linear response to a weak electromagnetic perturbation $\delta A_\mu(x)$. This can be accomplished by considering the new Lagrangian

$$\mathcal{L}_{\text{eff}}[\omega, \delta\mathcal{A}_\mu, \delta A_\mu] = \frac{\kappa}{2} \left(\partial_0 \omega - \delta\mathcal{A}_0 - \frac{e}{\hbar c} \delta A_0 \right)^2 - \frac{\rho_s}{2} \left(\partial_i \omega - \delta\mathcal{A}_i - \frac{e}{\hbar c} \delta A_i \right)^2 + \frac{1}{4\pi m} \epsilon_{\mu\nu\lambda} \delta\mathcal{A}^\mu \partial^\nu \delta\mathcal{A}^\lambda + \dots \quad (423)$$

and integrate out the phase field ω (which can be set to zero in the London/unitary gauge) and the fluctuation of the statistical field $\delta\mathcal{A}_\mu$ to find that effective electromagnetic action is

$$S_{\text{eff}}[\delta A_\mu] = \frac{1}{4\pi m} \frac{e^2}{\hbar} \int d^3x \epsilon_{\mu\nu\lambda} \delta A^\mu \partial^\nu \delta A^\lambda + \dots \quad (424)$$

from which we conclude that the Hall conductivity predicted by the composite boson theory is

$$\sigma_{xy} = \frac{1}{m} \frac{e^2}{\hbar} \quad (425)$$

as it should be for a FQHE at filling fraction $\nu = 1/m$.

We conclude by looking at vortex states in the composite boson theory. A vortex state is a time-independent solution of the equations of motion Eqs.(415)-(418) with the asymptotic behavior (in the temporal gauge $\delta\mathcal{A}_0 = 0$)

$$\lim_{|\mathbf{x}| \rightarrow \infty} \phi(\mathbf{x}) = \sqrt{\rho_0} e^{i\varphi(\mathbf{x})} \quad (426)$$

$$\lim_{|\mathbf{x}| \rightarrow \infty} \delta\mathcal{A}_i(\mathbf{x}) = \pm \partial_i \varphi(\mathbf{x}) = \pm \epsilon_{ij} \frac{x_j}{|\mathbf{x}|^2} \quad (427)$$

where $\varphi(\mathbf{x})$ is the azimuthal angle on the plane

$$\varphi(\mathbf{x}) = \tan^{-1} \left(\frac{x_2}{x_1} \right) \quad (428)$$

The energy of a neutral vortex (i.e. not coupled to a gauge field) is logarithmically divergent, $E_{\text{vortex}} \simeq \frac{\rho_s}{2} \ln(R/a_0)$ where R is the linear size of the system and a_0 is a short-distance cutoff (see section 5.1.1). However, the situation here is different since the complex scalar field $\phi(\mathbf{x})$ is coupled to a dynamical Chern-Simons gauge field \mathcal{A}_μ . Except for the Chern-Simons nature of this gauge field, the problem we are dealing with is similar to that of a superconductor coupled to a dynamical gauge field or to the Abelian-Higgs model in quantum field theory. In the case of interest here we have finite energy vortex solutions which satisfy the asymptotic condition

$$\lim_{|\mathbf{x}| \rightarrow \infty} \left| (i\partial_j - \delta\mathcal{A}_j) \phi(\mathbf{x}) \right|^2 = 0 \quad (429)$$

which is obeyed by configurations that obey Eq.(427). Thus, at long distances, the circulation of the Chern-Simons gauge field on a large closed contour Γ that contains the vortex satisfies

$$\oint_{\Gamma} \delta\mathcal{A}_j dx_j = \pm 2\pi \quad (430)$$

This vortex carries charge. To see this we compute the local charge density $j_0(\mathbf{x})$

$$j_0(\mathbf{x}) = -\frac{\delta S_{\text{eff}}}{\delta A_0(x)} = -e \frac{\delta S_{\text{eff}}}{\delta \delta \mathcal{A}_0(x)} = +e \frac{\delta S_{\text{CS}}}{\delta \delta \mathcal{A}_0(x)} = \frac{e}{2\pi m} \epsilon_{ij} \partial_i \delta \mathcal{A}_j(\mathbf{x}) \quad (431)$$

and compute the total charge of the vortex on a larger region Σ whose boundary is Γ and obtain

$$Q = e \int d^2x j_0(\mathbf{x}) = \frac{e}{2\pi m} \int_{\Sigma} d^2x \epsilon_{ij} \partial_i \delta \mathcal{A}_j(\mathbf{x}) = \frac{e}{m} \frac{1}{2\pi} \oint_{\Gamma} dx_j \delta \mathcal{A}_j(\mathbf{x}) = \pm \frac{e}{m} \quad (432)$$

Therefore, the vortex of the composite boson theory has the same charge as the Laughlin quasi-hole.

To determine the statistics of the vortex we go back to the effective Lagrangian written in the form of Eq.(409), (from now on we set $\delta \mathcal{A}^\mu \equiv \mathcal{A}^\mu$)

$$\mathcal{L}_{\text{eff}} = \frac{\kappa}{2} (\partial_0 \omega - \mathcal{A}_0)^2 - \frac{\rho_s}{2} (\partial_j \omega - \mathcal{A}_j)^2 + \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{B}^\lambda - \frac{e}{2\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu \mathcal{B}^\lambda - \frac{m}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu \partial^\nu \mathcal{B}^\lambda \quad (433)$$

where A_μ is the external electromagnetic probe field, and where we used units with $\hbar = c = 1$. The first two terms make the statistical gauge field \mathcal{A}_μ massive. In the low energy limit the statistical gauge field is frozen to the vortex configurations of the complex scalar field ϕ or, equivalently, to the vortex singularities of its phase field ω . The vorticity current Ω_μ is

$$\Omega_\mu = \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{A}^\lambda \quad (434)$$

The effective Lagrangian for the field \mathcal{B}_μ is

$$\mathcal{L}_{\text{eff}}[\mathcal{B}_\mu] = -\frac{m}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu \partial^\nu \mathcal{B}^\lambda - \frac{e}{2\pi} A^\mu \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{B}^\lambda + \Omega_\mu \mathcal{B}^\mu \quad (435)$$

which, except for the coupling to the electromagnetic field, has the same form as in Eq.(224), and of the effective action introduced in section 10.3.6 on phenomenological grounds.

The effective topological field theory of Eq.(435) encodes all the *universal* data of fractional quantum Hall fluids. Being topological this effective field theory has no energy scales and describes the physics at energies low compared to any excitation energy gap. In addition to yielding the correct Hall conductivity $\sigma_{xy} = ve^2/h$ (with $v = 1/m$) and the fractional vortex charge $Q = e/m$, this theory will allow us to draw important additional results about the low energy physics.

By using the results of section 9.7, we conclude that the vortices of the composite boson are anyons with fractional statistics $\exp(\pm i\pi/m)$, consistent with the conclusions of section 10.3.5. In addition, on a 2-torus the ground state of this system is m -fold degenerate. This feature, characteristic of systems with *topological order*, is that the ground state degeneracy depends on the topology of the surface on which the 2DEG resides. The ground state on a torus degeneracy was shown by Duncan Haldane and Edward Rezayi [245] by an explicit construction of a model wave function for the Laughlin states on a torus. On a surface with g handles (i.e. genus g) the degeneracy is m^g . Xiao-Gang Wen and Qian Niu [246] showed that these results hold for any FQH state which are, quite generally, topological quantum fluids.

Finally, we may ask how many *distinct* types of vortices does this theory have. The fundamental (Laughlin) vortex has charge e/m and statistics π/m . In principle we could have a vortex with any integer topological charge (winding number) $n \in \mathbb{Z}$. Such a vortex has charge ne/m and statistics $\pi n^2/m$. However, a vortex with topological charge $n = m$ has charge e and statistics $m\pi$. This vortex is indistinguishable from a hole (a missing electron) since it has the same charge and statistics. We conclude that a Laughlin state has m distinct vortices with $n = 1, 2, \dots, m-1$. We notice that this number is the as the number of degenerate states on a torus. In section 10.3.10 we will see that this fact is closely related to the number of primary fields of a conformal field theory associated with the FQH states.

10.3.8. Composite Fermion Field Theory

At the beginning of section 10.3.7 we noted that flux attachment can be used to define equivalent theories: a) by attaching an *odd* number, m , fluxes to each electron thereby becoming a composite boson, and b) by attaching an *even* number, $m-1$, fluxes by which the electrons turn into *composite fermions*, as in Jain's construction [230]. In this section we discuss the most salient features of the field theory approach to composite fermions, introduced by Ana López and

me in 1991[247, 248] as a theory of all the Jain states. This approach was extended in 1993 by Bertrand Halperin, Patrick Lee and Nicholas Read [232] to the case of the compressible states.

By following the same approach that we used in section 10.3.7, but adapted to the case in which we map to a theory of a composite Fermi field $\psi(x)$, we find their dynamics is described by the effective action [247]

$$\begin{aligned} \mathcal{S}_{\text{CF}} = & \int d^3x \left\{ \psi^*(x)[iD_0 + \mu]\psi(x) + \frac{1}{2M}|\mathbf{D}\psi(x)|^2 \right\} + \frac{1}{4\pi n} \int d^3x \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{A}^\lambda \\ & - \frac{1}{2} \int d^3x \int d^3x' (|\psi(x)|^2 - \rho_0)V(x - x')(|\psi(x')|^2 - \rho_0) \end{aligned} \quad (436)$$

where $n = m - 1$ is an even integer. Here $V(x - x') = \delta(x_0 - x'_0)V(|x - x'|)$ is the instantaneous electron-electron repulsive interaction, and ρ_0 is the average areal neutralizing charge density. As before, the covariant derivative is $D_\mu = \partial_\mu + ieA_\mu + \mathcal{A}_\mu$, where A_μ is the electromagnetic field (including the uniform magnetic field B), and \mathcal{A}_μ is the statistical gauge field. We are using units such that $\hbar = c = 1$.

We can simplify somewhat the form of the actions for the composite fermions by using the fact that the Gauss law of Chern-Simons gauge theory states that the particle density and the gauge flux are rigidly tied together, $|\psi(x)|^2 = \frac{1}{2\pi n}\mathcal{B} \equiv \epsilon_{ij}\partial_i \mathcal{A}_j$ (this is the flux attachment) as a operator identity. Thus, we can write the action as

$$\begin{aligned} \mathcal{S}_{\text{CF}} = & \int d^3x \left\{ \psi^*(x)[iD_0 + \mu]\psi(x) + \frac{1}{2M}|\mathbf{D}\psi(x)|^2 \right\} + \frac{1}{4\pi n} \int d^3x \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{A}^\lambda \\ & - \frac{1}{2} \int d^3x \int d^3x' \left(\frac{\mathcal{B}(x)}{2\pi n} - \rho_0 \right) V(x - x') \left(\frac{\mathcal{B}(x')}{2\pi n} - \rho_0 \right) \end{aligned} \quad (437)$$

Since this action is a quadratic form in the Fermi (Grassmann) fields ψ , we can integrate them out to obtain the following effective action for the statistical gauge field \mathcal{A}_μ

$$S_{\text{eff}}[\mathcal{A}_\mu] = -i\text{tr} \left(iD_0 + \mu + \frac{\mathbf{D}^2}{2M} \right) + \frac{1}{4\pi n} \int d^3x \epsilon_{\mu\nu\lambda} (\mathcal{A}^\mu - eA^\mu) \partial^\nu (\mathcal{A}^\lambda - eA^\lambda) + S_{\text{int}}[\mathcal{A}_\mu - eA_\mu] \quad (438)$$

where A_μ is a probe external electromagnetic field (with vanishing average) and does not include the uniform magnetic field. The interaction term is

$$S_{\text{int}}[\mathcal{A}_\mu - eA_\mu] = -\frac{1}{2} \int d^3x \int d^3x' \left(\frac{(\mathcal{B}(x) - eB(x))}{2\pi n} - \rho_0 \right) V(x - x') \left(\frac{(\mathcal{B}(x') - eB(x'))}{2\pi n} - \rho_0 \right) \quad (439)$$

Here too $B(x)$ is an external probe with vanishing average.

We will investigate the properties of the path integral for the statistical gauge field

$$Z[A_\mu] = \int \mathcal{D}\mathcal{A}_\mu \exp(iS_{\text{eff}}[\mathcal{A}_\mu, A_\mu]) \quad (440)$$

using a saddle point expansion. The saddle-point condition of the effective action of Eq.(438)

$$\frac{\delta S_{\text{eff}}}{\delta \mathcal{A}_\mu(x)} = 0 \quad (441)$$

leads to the equation of motion for the field \mathcal{A}_μ

$$\langle j_\mu^F(x) \rangle + \frac{1}{4\pi n} \epsilon_{\mu\nu\lambda} [\mathcal{F}^{\nu\lambda}(x) - eF^{\nu\lambda}] = 0 \quad (442)$$

where $\langle j_\mu^F \rangle$ is the expectation value of the fermionic current. In addition we need to impose the condition for the particle density to be uniform and equal to the neutralizing background charge $\langle j_0^F(x) \rangle = \rho_0$.

We will assume that the electromagnetic field A_μ describes just a static uniform magnetic field of strength B . Under these conditions, the ground state should also be static and uniform. This means that the field strength of the statistical gauge field should be constant and uniform value $\langle \mathcal{B} \rangle$, and that the current $\langle j^F \rangle$ should vanish in the ground state. On the other hand, the Gauss Law of the Chern-Simons gauge field implies that $\langle \mathcal{B} \rangle = -2\pi n \rho_0$, while the zero current condition requires the the (statistical) electric field vanishes, $\langle \mathcal{E} \rangle = 0$. Since the composite

fermion couples in the same way to the electromagnetic field A_μ and to the statistical field \mathcal{A}_μ we conclude that they experience an effective magnetic field

$$B_{\text{eff}} = B + \frac{1}{e} \langle \mathcal{B} \rangle = B - 2\pi n \frac{\rho_0}{e} \quad (443)$$

If the total number of electrons is N , then $\rho_0/e = N/L^2$ where L is the linear size of the system. Let us denote by N_ϕ^{eff} the total number of flux quanta of the effective magnetic field (in units of $\hbar = c = 1$ in which the flux quantum is $\phi_0 = 2\pi$), and $2\pi N_\phi = BL^2$ is the total flux. Then, the total effective flux is

$$2\pi N_\phi^{\text{eff}} = 2\pi N_\phi - 2\pi n N \quad (444)$$

Let $\nu = N/N_\phi$ be the filling fraction and $\nu^{\text{eff}} = N/N_\phi^{\text{eff}}$ the effective filling fraction. Eq.(444) implies that these filling fractions are related by

$$\frac{1}{\nu^{\text{eff}}} = \frac{1}{\nu} - n \quad (445)$$

Recall that n is an *even* integer. However, the system will be incompressible only if $\nu^{\text{eff}} = p \in \mathbb{Z}$. In other words, the composite fermions fill an integer number p of the partially screened Landau levels, which is Jain's condition. For this condition to be satisfied the filling fraction $\nu(p, n)$ is

$$\nu_\pm(n, p) = \frac{p}{np \pm 1} \quad (446)$$

which are the Jain fractions. In the special case $p = 1$ the Jain fractions are the Laughlin fractions, with $n = m - 1$. Similarly we find that B_{eff} is

$$B_{\text{eff}} = \pm \frac{B}{np \pm 1} \quad (447)$$

So we see that the external field B is partially screened to the smaller value B_{eff} which can be parallel (screened) to B or anti-parallel (overscreened) to B . For the same reason, at this mean field level the cyclotron frequency ω_c is reduced by the same amount to $\omega_c^{\text{eff}} = \omega_c/(np \pm 1)$.

We will discuss now the effects of quantum fluctuations for the action of Eq.(438) about the saddle-point solutions. We will work to the lowest order (quadratic) in the fluctuations, which can be regarded as a semi-classical (or "RPA") treatment of this theory. The quadratic effective action for the fluctuations of statistical gauge field, which we will still denote by \mathcal{A}_μ is

$$S_{\text{eff}}^{(2)}[\mathcal{A}_\mu] = \frac{1}{2} \int d^3x \int d^3y \mathcal{A}^\mu(x) \Pi_{\mu\nu}^{\text{CF}}(x, y) \mathcal{A}^\nu(y) + \frac{1}{4\pi n} \int d^3x \epsilon_{\mu\nu\lambda} (\mathcal{A}_\mu(x) - eA_\mu(x)) \partial^\nu (\mathcal{A}_\nu(x) - eA_\nu(x)) + S_{\text{int}}(\mathcal{A}_\mu - eA_\mu) \quad (448)$$

Here, $\Pi_F^{\mu\nu}(x, y)$ is the polarization tensor of the composite fermions for p filled effective Landau levels. The interaction term of the action is given in Eq.(439).

As required by gauge invariance, the composite fermion polarization tensor is transverse, $\partial^\mu \Pi_{\mu\nu}^{\text{CF}} = 0$, which fixes the tensorial structure. In momentum and frequency space the composite fermion polarization tensor $\Pi_{\mu\nu}^{\text{CF}}(\mathbf{q}, \omega)$ depends on three kernels $\Pi_0^{\text{CF}}(\mathbf{q}, \omega)$, $\Pi_1^{\text{CF}}(\mathbf{q}, \omega)$ and $\Pi_2^{\text{CF}}(\mathbf{q}, \omega)$, where Π_0^{CF} and Π_2^{CF} contain the parity-even response of the composite fermions and Π_1^{CF} is their parity-odd response. Each kernel is given as a series terms representing particle-hole processes with simple poles at the excitations energies $\omega_{rs} = (r - s)\omega_c^{\text{eff}}$, with $r > p$ (particles) and $s \leq p$ (holes). Each term has a residue which is an integer power of \mathbf{q}^2 times a Laguerre polynomial in \mathbf{q}^2 . Details of these kernels can be found in Ref.[247] and in chapter 13 of Ref.[9]. However, what will be important here is that, in the gapped states these kernels have a low energy and low momentum limit which is local. This will enable us to find a local topological effective action only for the gapped states.

After integrating out the statistical gauge field \mathcal{A}_μ we find the effective action for the external electromagnetic probe field A_μ , which has the standard form

$$S_{\text{eff}}[A_\mu] = \frac{1}{2} \int d^3x \int d^3y A_\mu(x) \Pi^{\mu\nu}(x - y) A_\nu(y) \quad (449)$$

which encodes the full electromagnetic response of the system, not just of the composite fermions. Once again, gauge invariance requires that the full polarization tensor be transverse, $\partial^\mu \Pi_{\mu\nu} = 0$.

In section 10.1.4 we showed that there is a relation between the polarization tensor $\Pi_{\mu\nu}$ and the current-current correlation function $\mathcal{D}_{\mu\nu}$. There we mentioned the these correlators are required (by gauge invariance) to satisfy a Ward identity known as the f-sum rule:

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} i\omega \mathcal{D}_{00}^R(\omega, \mathbf{q}) = \frac{\rho_0}{M} \mathbf{q}^2 \quad (450)$$

where the (retarded) density-density correlation function is related to the (retarded) polarization tensor, $\mathcal{D}_{00}^R(\omega, \mathbf{q}) = \Pi_{00}^R(\omega, \mathbf{q})$. In the limit of low momentum \mathbf{q} , at fixed frequency ω , The expression for $\Pi_{00}^R(\omega, \mathbf{q})$ is

$$\Pi_{00}^R(\omega, \mathbf{q}) \simeq \mathbf{q}^2 \Pi_0^R(\omega, \mathbf{q}) = -\frac{\rho_0}{M} \frac{\mathbf{q}^2}{(\omega + i\epsilon)^2 - \omega_c^2} \quad (451)$$

where $\omega_c = eB/(Mc)$ is the cyclotron frequency of the *electrons*. This expression saturates the sum rule. This means that whichever corrections Eq.(451) may have must vanish in the low \mathbf{q} limit. In other words, in this limit the approximation that we made is actually exact. This is well known result in the theory of the Fermi liquid [194].

In addition, the result of Eq.(451) implies that there is a particle-hole (density) collective mode which at low momentum has energy $\hbar\omega_c$, without corrections. This result is known as Kohn's theorem [249] which states that in a Galilean invariant system the 2DEG must have an exact eigenstate at the cyclotron frequency. More intuitively, the center of mass of the 2DEG executes a cyclotron motion regardless of whether the particles are free, interacting or not, fermions, bosons, or anyons. This is a general result. Notice that the composite fermions do not satisfy Kohn's theorem as they cyclotron mode would be at the effective cyclotron frequency ω_c^{eff} , and that the leading quantum fluctuations have restored the magnetic symmetry.

In the limit of low energy $\omega \rightarrow 0$ and low momentum $\mathbf{q} \rightarrow 0$ of the kernels Π_0^{CF} , Π_1^{CF} and Π_2^{CF} become

$$\Pi_0^{CF}(0, 0) = \frac{p}{2\pi} \frac{M}{B_{\text{eff}}} \equiv \varepsilon^{CF}, \quad \Pi_1^{CF}(0, 0) = \pm \frac{p}{2\pi} \equiv \sigma_{xy}^{CF}, \quad \Pi_2^{CF}(0, 0) = -\frac{1}{2\pi} \frac{p^2}{M} \equiv -\chi^{CF} \quad (452)$$

where ε^{CF} is the “dielectric constant” of the composite fermions, χ^{CF} is their effective “permeability”, and σ_{xy}^{CF} is the (integer) composite fermion Hall conductivity. We then find that low-energy effective action of the statistical gauge field \mathcal{A}_μ is

$$\begin{aligned} S_{\text{eff}}[\mathcal{A}_\mu] = & \frac{1}{2} \left(\sigma_{xy}^{CF} + \frac{1}{2\pi n} \right) \int d^3x \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu \mathcal{A}^\lambda(x) \\ & - \frac{e}{2\pi n} \int d^3x \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu A^\lambda(x) + \frac{e^2}{4\pi n} \int d^3x \epsilon_{\mu\nu\lambda} A^\mu(x) \partial^\nu A^\lambda(x) \\ & + \int d^3x \frac{1}{2} (\varepsilon^{CF} \mathcal{E}_i^2(x) - \chi^{CF} \mathcal{B}^2(x)) \\ & - \frac{1}{8\pi^2 n^2} \int d^3x \int d^3y (\mathcal{B}(x) - eB(x)) V(x-y) (\mathcal{B}(y) - eB(y)) \end{aligned} \quad (453)$$

Notice that, except possibly for the last term, the effective action is a local. This is possible because the mean field theory state is gapped.

In what follows we will focus only on the leading terms of the effective action and neglect the subleading terms, the least two terms of the effective action which have the form of a Maxwell term and an additional (possibly non-local) term. The remaining leading terms have a Chern-Simons form and a BF form whose effective Lagrangian is

$$\mathcal{L}_{\text{eff}}[\mathcal{A}_\mu] = \frac{1}{4\pi} \left(\pm p + \frac{1}{n} \right) \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu \mathcal{A}^\lambda(x) - \frac{e}{2\pi n} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu A^\lambda(x) + \frac{e^2}{4\pi n} \epsilon_{\mu\nu\lambda} A^\mu(x) \partial^\nu A^\lambda(x) \quad (454)$$

If now now integrate ot the statistical gauge field \mathcal{A}_μ we find the the electromagnetic field A_μ has the effective Lagrangian

$$\mathcal{L}_{\text{eff}}[A_\mu] = \frac{\sigma_{xy}}{2} \epsilon_{\mu\nu\lambda} A_\mu(x) \partial^\nu A_\lambda(x) \quad (455)$$

from where we find that the Hall conductivity σ_{xy} of the 2DEG is

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar} \frac{p}{pn \pm 1} \quad (456)$$

which is the Hall conductivity of the Jain states (in standard units). Here, as before, n is an even integer.

We will now return to the effective Lagrangian of Eq.(454). As it stands, since the Chern-Simons term has fractional level, it is not invariant under large gauge transformations and, as such, cannot be defined on a closed manifold. To remedy this problem we can now write this Lagrangian in terms of a theory with two dynamical gauge fields \mathcal{A}_μ and \mathcal{B}_μ as follows:

$$\begin{aligned} \mathcal{L}_{\text{eff}}[\mathcal{A}_\mu, \mathcal{B}_\mu] = & \frac{p}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu \mathcal{A}^\lambda(x) - \frac{n}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu(x) \partial^\nu \mathcal{B}^\lambda(x) \\ & + \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu(x) \partial^\nu \mathcal{B}^\lambda(x) - \frac{e}{2\pi} \epsilon_{\mu\nu\lambda} \mathcal{B}^\mu(x) \partial^\nu \mathcal{A}^\lambda(x) \end{aligned} \quad (457)$$

This is a topological quantum field theory for all the Jain states [250]. This theory can be defined on any manifold no matter its topology. It is easy to see that upon integrating out the field \mathcal{B}_μ we recover the effective field theory for \mathcal{A}_μ of Eq.(454). Also, in the special case of the Laughlin states, for which $p = 1$, we can integrate out the field \mathcal{A}_μ , and arrive to the effective topological field theory for the field \mathcal{B}_μ of Eq.(435) that we derived in the composite boson theory in section 10.3.7.

Let us rewrite the Lagrangian of Eq.(457) in a more compact yet general form. To this end we define a multicomponent gauge field \mathcal{A}_μ^I with $I = 1, \dots, L$, an L -component *charge* vector t , and an $L \times L$ second rank symmetric matrix K_{IJ} . Wen and Zee have given a general classification of a large class of fractional quantum Hall states, said to be *abelian*, whose topological field theory is defined by the Lagrangian [251, 252]

$$\mathcal{L} = \frac{1}{4\pi} K_{IJ} \epsilon^{\mu\nu\lambda} \mathcal{A}_\mu^I(x) \partial^\nu \mathcal{A}_\lambda^J(x) - \frac{e}{2\pi} t_I \epsilon^{\mu\nu\lambda} A_\mu(x) \partial^\nu \mathcal{A}_\lambda^I(x) \quad (458)$$

In the case of the theory of Eq.(457) which has two components, $L = 2$, with $\mathcal{A}_\mu^1 = \mathcal{A}_\mu$ and $\mathcal{A}_\mu^2 = \mathcal{B}_\mu$. The two-component vector is $t = (0, 1)$ and the matrix 2×2 matrix is

$$K = \begin{pmatrix} -p & 1 \\ 1 & n \end{pmatrix} \quad (459)$$

Wen and Zee showed that, quite generally, a theory of the K -matrix form has a vacuum degeneracy on a torus of $|\det K|$ and, on a surface of genus g , the degeneracy is $|\det K|^g$. They also showed that the Hall conductivity is $\sigma_{xy} = v e^2/h$ where the filling fraction is

$$v = \sum_{I,J=1}^L t_I K_{IJ}^{-1} t_J \quad (460)$$

The properties of the quasiparticles can be determined by computing the appropriate correlator. In the composite fermion theory, whose Lagrangian is given in Eq.(437), the *gauge-invariant* composite fermion propagator is

$$G_{CF}(x - y; \gamma(x, y)) = \left\langle \psi^\dagger(x) \exp(i \int_{\gamma(x,y)} dz_\mu (eA_\mu(z) + \mathcal{A}^\mu(z))) \psi(y) \right\rangle \quad (461)$$

which, as in any gauge theory, it is gauge-invariant but path-dependent. Here γ is an oriented open path with endpoints that the space-time locations x and y . The composite fermion propagator has information about many of the properties of the quasiparticles, including their electric charge. To find what is the statistics of the quasiparticles we need to compute a two-particle correlator (a four-point function) which is defined in terms of two open oriented paths, say $\gamma_1(x, y)$ and $\gamma_2(u, w)$, with endpoints at the locations of the two particles, respectively.

The computation of these correlators is done using a feynam sum over trajectories with different weights which depend on the gauge field configurations. To do these calculations is, in general, a complicated problem. However, in a gapped state, the quasiparticles are massive and in the low energy regime these expressions are dominated by a classical trajectory on an oriented path $\tilde{\gamma}(x, y)$ with the same endpoints x and y . The end result reduces to the computation of the expectation value of a Wilson loop operator

$$W[\Gamma] = \left\langle \exp(i \oint_{\Gamma} dz_\mu (eA_\mu(z) + \mathcal{A}^\mu(z))) \right\rangle_{\mathcal{A}_\mu} \quad (462)$$

on the *closed* oriented path $\Gamma = \gamma \cup \tilde{\gamma}^-$ (where $\tilde{\gamma}^-(y, x)$ has the opposite orientation as $\tilde{\gamma}(x, y)$ and runs from y to x). The explicit dependence on the external electromagnetic field yields the value of the charge of the quasiparticle through the Aharonov-Bohm effect. Likewise, the two-particle correlator is reduced, in the asymptotic low energy regime, to the computation of an expectation value of two Wilson loops in the effective Chern-Simons theory, as was done in section 9.7, which yields the fractional statistics of the quasiparticles. The interested reader can find details in chapter 13 of Ref.[9].

A system described by a theory of the form of Eq.(458) has different types of quasiparticles. In general we can assign an integer-valued quantum number $\ell_I \in \mathbb{Z}$ as the charge of the quasiparticle with respect to the gauge field \mathcal{A}_I^μ . The general coupling of a quasiparticle current has the form $\mathcal{L}_{qp} = j_\mu \ell_I \mathcal{A}_I^\mu$. Then, the charge $Q[\ell]$ and the statistics $\delta[\ell]$ of a general quasiparticle defined by the integer-valued L -component vector ℓ are

$$Q[\ell] = -et_I K_{IJ}^{-1} \ell_J, \quad \frac{\delta[\ell]}{\pi} = \ell_I K_{IJ}^{-1} \ell_J + 1 \quad (463)$$

where the $+1$ is required to refer the statistics to fermions (not bosons) [250]. These results are general and hold for any abelian FQH state [251].

In the case of the Jain states, the charge vector is $t = (0, 1)$ and the quasiparticle is labeled by the vector $\ell = (1, 0)$ which yield the results for the charge Q and the statistics δ

$$Q = \frac{-e}{np+1}, \quad \delta = \pi \left(\frac{-n}{np+1} + 1 \right) \quad (464)$$

which are the quantum numbers of the quasiparticles of the Jain states. For the Laughlin state at $\nu = 1/3$ we obtains $Q = -e/3$ and $\delta = \pi/3$ (as we should), which for the first Jain state at $\nu = 2/5$ the charge Q and the statistics δ are $Q = -e/5$ and $\delta = 3\pi/5$. With some caveats, the predictions for the quasiparticle charge in the FQH states with $\nu = 1/3$ and $\nu = 2/5$ have been confirmed in noise measurements at a quantum point contact by R. de Picciotto and coworkers [253] and by L. Samindayar and coworkers [254]. Fractional statistics has been measured experimentally in both FQH states at $\nu = 1/3$ and $\nu = 2/5$ in (Fabry-Perot) interferometers by J. Nakamura, S. Liang, G. Gardner, and M. Manfra [233].

10.3.9. The Compressible States

At large values $p \rightarrow \infty$ the Jain sequences converge to the even-denominator fractions $\nu_\infty(n) = 1/n$. In this limit, $B_{\text{eff}} \rightarrow 0$ and $\omega_c^{\text{eff}} \rightarrow 0$. In other words, at the filling fractions $\nu_\infty(n)$ the average effective field experienced by the composite fermions is zero. This mean field theory then would predict that this state is essentially a Fermi liquid of composite fermions which is a compressible state. In addition, one can compute the effective charge of the composite fermion and find that it is $e/(np \pm 1)$, which also vanishes in the compressible limit. It is straightforward to see that the Fermi sea of the composite fermions is a disk with Fermi momentum $p_F = (2\nu_\infty(n))^{1/2} \hbar/\ell_0$ and that the Fermi energy is $E_F = \nu_\infty(n) \hbar/(M\ell_0^2)$. Halperin, Lee, and Read [232] did an in-depth analysis of the consequences of this compressible state which are largely in agreement with experiment.

In our discussion of the incompressible FQ states, both in the composite boson and in the composite fermion approaches in sections 10.3.7 and 10.3.8 we saw that the mean field approximation violated symmetries that were restored by quantum fluctuations. This fact allowed us the derive effective filed theories which are asymptotically exact in the low energy IR regime. What allowed us to obtain these exact results is the existence of a gap. However, the compressible states are gapless and in some sense should be regarded theories of a quantum critical system.

In fact, if we reexamine the field theory of composite fermions of section 10.3.8 we see that, in the compressible limit $p \rightarrow \infty$, the action of Eq.(437) describes a system of fermions at finite density coupled to the *fluctuations* \mathcal{A}_μ of the statistical gauge field with a Chern-Simons term, but without an external uniform field (which has been canceled by the flux of the average statistical gauge field). This theory has two salient features. one is that the only trace left of the explicitly broken time-reversal symmetry is in the presence of the Chern-Simons term. This means that while the mean field theory looks like a Fermi liquid, the consequences of the broken time reversal invariance only enters in the quantum fluctuations. In addition, in this regime this theory also breaks the particle-hole symmetry of a half-filled Landau level (in the high magnetic field regime).

However, these contributions, already at one-loop order, typically contain IR divergencies. These IR divergencies are due to singular forward scattering processes of the composite fermions

by the gauge fields. They are most severe in the computation of the fermion self energy and depend on the form of the electron-electron interactions. For instance if the interaction potential $V(R)$ is short range, the one-loop contribution to the fermion self-energy has an imaginary part with the behavior $\Sigma''(\omega) \sim \omega^{2/3}$ which is *much larger* than the real part $\Sigma'(\omega)$ as $\omega \rightarrow 0$ and the Fermi energy E_F is approached [232]. This behavior, which is found in many metallic systems at quantum criticality [255, 256, 69] implies that the width of the quasiparticle is asymptotically much larger than the energy as the Fermi surface is approached. This one-loop result means that perturbation theory fails in the IR and that actual behavior may be non-perturbative. This problem is present even in the case of unscreened $\frac{e^2}{R}$ (Coulomb) interactions where the singular behavior is milder with $\Sigma''(\omega) \sim \omega \ln \omega$ (known as Marginal Fermi Liquid scaling [257]).

At any rate, these IR singular behaviors imply that the quasiparticle picture is inadequate and that in these regimes there may be no quasiparticles. These systems are generally known as “Strange Metals”. Many strongly interacting physical systems of interest, ranging from high T_c superconductors to (largely conjectured) spin liquids with spinon Fermi surfaces to even quark-gluon plasmas share these types of singular IR behaviors [257, 258]. In spite of a large body of theoretical work, it has remained largely unsolved problem. Perturbative renormalization group methods used to justify the Landau theory of the Fermi liquid [259, 260] generally fail in these regimes, although in some cases long range interactions have allowed some degree of control [261, 262]. Remarkably, the AdS/CFT Gauge/Gravity Duality [263] has brought new insights into strange metal behaviors [264]. We will see below that relativistic dualities have given additional insights into the physics of compressible states.

10.3.10. Fractional Quantum Hall Wave Functions and Conformal Field Theory

The Laughlin wave function and its generalizations share the remarkable feature of being essentially universal. Aside from the dependence on the magnetic length ℓ_0 in the gaussian factors which ensure that the function is integrable at long distances (a feature inherited from the single-particle states in the Landau level), the FQH “ideal” wave functions do not have any length scales. In addition, in almost all cases, the ideal wave functions are the exact ground state wave functions of some suitable ultra-local Hamiltonian projected onto the lowest Landau level. Another feature, closely related to their universality and analyticity, is that these wave functions look similar to correlation functions in two-dimensional critical (chiral) critical systems. We will now see that these features are not accidental and point to a deep relation between quantum Hall states and Conformal Field Theory (CFT).

Although the relation between the simpler case of the Laughlin state was known to several people, such as Sergio Fubini [265], this deep connection with *rational* CFT was articulated in considerable generality in a somewhat earlier paper by Gregory Moore and Nicholas Read [266]. This approach showed its full power in the formulation of the non-abelian quantum Hall states.

Let us begin with the simpler case of the Laughlin states with $v = 1/m$, where m is odd for fermions and even for bosons. We will consider a U(1) *Euclidean* CFT of a compact boson (a scalar field) $\phi(x)$ closely related to our discussion of bosonization in 1+1 dimensions in section 7.4. This theory has been extensively studied in the CFT literature [103, 102, 104]. The Euclidean action for this theory is

$$S = \frac{1}{8\pi} \int d^2x (\partial_\mu \phi)^2 \quad (465)$$

This theory is compactified by the condition that the only allowed observables are invariant under the discrete symmetry

$$\phi(x) \mapsto \phi(x) + 2\pi R n \quad (466)$$

where R is the compactification radius and n is an arbitrary integer. The observables that satisfy this condition are vertex operators of the form

$$V_p(x) = \exp\left(i \frac{p}{R} \phi(x)\right) \quad (467)$$

where p is an integer.

The correlators of this theory are products of holomorphic (right-moving in Minkowski spacetime) and antiholomorphic (left-moving in Minkowski spacetime) factors. In the context of the FQH states we will be interested in a chiral theory, whose correlators are holomorphic (analytic) functions in complex coordinates $z = x_1 + ix_2$. Formally, the field $\phi(z, \bar{z})$ is decomposed into a sum of a holomorphic field $\phi(z)$ and an antiholomorphic field $\phi(\bar{z})$, so that the propagators is

$$\langle \phi(z, \bar{z}) \phi(w, \bar{w}) \rangle = -\ln(z-w) - \ln(\bar{z}-\bar{w}) \quad (468)$$

In what follows we will focus on the chiral(holomorphic) field $\phi(z)$ whose propagator is

$$\langle \phi(z)\phi(w) \rangle = -\ln(z-w) \quad (469)$$

The correlator of the chiral current operator

$$j(z) = \frac{i}{R} \partial_z \phi(z) \quad (470)$$

is given by

$$\langle i\partial_z \phi(z) i\partial_w \phi(w) \rangle \sim \frac{1}{(z-w)^2} \quad (471)$$

which tells us that this operator has (chiral) scaling dimension $\Delta = 1$. Similarly the correlator of two *chiral* vertex operators

$$V_p(z) = \exp\left(i\frac{p}{R}\phi(z)\right) \quad (472)$$

is

$$\langle V_p(z) V_{-p}(w) \rangle = \frac{1}{(z-w)^{p^2/R^2}} \quad (473)$$

whose (chiral) scaling dimension is $\Delta_p = p^2/(2R^2)$. These expressions are very similar to the operators that we used in section 5.1.1 to describe vortices in 2D superfluids (and planar magnets). The main difference is that here the fields and their correlators are holomorphic where in section 5.1.1 the correlators are the product of a holomorphic and antiholomorphic factors.

For general p and R the correlator of Eq.(473) has a branch cut from z to w . This means that the *monodromies* of the operators, i.e. dragging an operator around the other, induces a change in the *phase* of the correlator equal to $2\pi p^2/R^2$. This behavior is reminiscent of the analytic structure of the quasiparticle wave function in the FQH states and of the *braiding* operation between anyons.

Moore and Read [266] showed that the following correlator in a U(1) compactified boson CFT (with compactification radius $R = \sqrt{m}$) is equal to the Laughlin wave function at filling fraction $\nu = 1/m$,

$$\Psi_m(z_1, \dots, z_N) = \left\langle \left(\prod_{i=1}^N \exp\left(i\sqrt{m}\phi(z_i)\right) \right) \exp\left(-i \int d^2 z' \sqrt{m} \rho_0 \phi(z')\right) \right\rangle \quad (474)$$

$$= \prod_{i < j} (z_i - z_j)^m \exp\left(-\frac{1}{4} \sum_{i=1}^N |z_i|^2\right) \quad (475)$$

where we have used units in which the magnetic length is $\ell_0 = 1$.

In section 5.1.1 we saw that correlators of vertex operators are non-vanishing only if their total charge (vorticity) is zero. The reason for this condition is that the (Euclidean) action is invariant under arbitrary uniform shifts of the field ϕ . Now, in the correlator on the right hand side of Eq.(475) we have a product of N vertex operators, each with the same charge \sqrt{m} . Then, under an arbitrary shift by α of the field $\phi(z)$, the product of vertex operators change by a phase equal to $N\sqrt{m}\alpha$. On the other hand the other operator in the correlator, which describes a continuum background charge, changes by a phase $-\sqrt{m}\rho_0 L^2 \alpha$, where L^2 is the area. Since ρ_0 is the areal particle density, N/L^2 , the two changes of the phase cancel each other exactly. Thus, the operator in the expectation value of Eq.(475) is charge neutral and has a non-vanishing expectation value.

The other important feature of the correlator of Eq.(475) is the choice of compactification radius, $R = \sqrt{m}$, and of the vertex operator $V_p(z)$ with $p = m$:

$$V_m(z) = \exp(i\sqrt{m}\phi(z)) \quad (476)$$

whose correlation function is

$$\langle V_m(z) V_{-m}(w) \rangle = \frac{1}{(z-w)^m} \quad (477)$$

Which is invariant under a 2π rotation (since $m \in \mathbb{Z}$). However under a rotation by π (i.e. an *exchange*) it changes by $(-1)^m$. Hence, for m odd it changes sign, and the vertex operator V_m behaves as a *fermion*, while for m even the vertex operator behaves as a *boson*. We will call the vertex operator $V_m \equiv V_e$ the electron operator.

In other words, the correlator of Eq.(475) is the expectation value of N vertex operators each describing an *electron* (for m odd) and a neutralizing background charge. Then, an elementary calculation yields the expression of the Laughlin wave function.

In this formulation the ground state wave function for the integer QH state at $\nu = 1$, which has the form of a Vandermonde determinant shown in Eq.(377), is interpreted as the correlator of N electron operators in the U(1) CFT with compactification radius $R = 1$. In this simple case, the electron operator is the vertex operator $V_1(z) = \exp(i\phi(z))$, and there are no other operators (aside from the current $j(z) = i\partial_z\phi$). The propagator of this vertex operator is

$$\langle V_1(z)V_{-1}(w) \rangle = \frac{1}{z-w} \quad (478)$$

which indeed represents a fermion with electric charge e .

We will now see that the vertex operator of Eq.(472) with $p = 1$ (and $R = \sqrt{m}$)

$$V_1(z) = \exp\left(\frac{i}{\sqrt{m}}\phi(z)\right) \quad (479)$$

represents the Laughlin quasi-hole. The correlator of this vertex operator is

$$\langle V_1(z) V_{-1}(w) \rangle = \frac{1}{(z-w)^{1/m}} \quad (480)$$

which has a branch cut stretching from z to w and, as a result, under a rotation by $\pm\pi$ its phase changes by $\pm\pi/m$, just as the Laughlin anyons do. To see that this operator indeed is related to the Laughlin quasi-hole we will compute the effect of inserting the vertex operator $V_1(z)$ in the correlator of Eq.(475) and find

$$\begin{aligned} & \left\langle \exp\left(\frac{i}{\sqrt{m}}\phi(w)\right) \left(\prod_{i=1}^N \exp(i\sqrt{m}\phi(z_i)) \right) \exp\left(-i \int d^2 z' \sqrt{m} \rho_0 \phi(z')\right) \right\rangle = \\ &= \prod_{i=1}^N (z_i - w) \prod_{i < j} (z_i - z_j)^m \exp\left(-\frac{1}{4} \sum_{i=1}^N |z_i|^2 - \frac{1}{4m} |w|^2\right) \\ &= \Psi_m(w; z_1, \dots, z_N) \end{aligned} \quad (481)$$

which is, indeed, the wave function for the Laughlin quasi-hole of Eq.(392). The insertion of an additional vertex operator $V_1(u)$ at u inside the expectation value of Eq.(481) yields the two-quasi-hole wave function of Eq.(397), proposed by Halperin [234] and discussed in section 10.3.5, with the same branch cut shown in Eq.(398) and, therefore carry fractional statistics π/m .

The operator product expansion discussed in section 3.3.2 of the vertex operators of this CFT yields new insight on the quasiholes. Indeed the OPE of two vertex operators of charges p and q is

$$\lim_{w \rightarrow u} V_p(w)V_q(u) = \lim_{w \rightarrow u} \frac{1}{(w-u)^{\Delta_p + \Delta_q - \Delta_{[p+q]_m}}} V_{[p+q]_m}(u) \quad (482)$$

where $[p]_m$ is the integer p modulo a multiple of m , i.e. if $p = mr + s$ (with r a non-negative integer and $0 \leq s < m$, then $[p]_m = s$). Eq.(482) is understood in the sense that the contribution of all additional operators to the right hand side vanish as $w \rightarrow u$. In other words, the vertex operators are primary fields of this CFT and there are only m primary fields. The physical process described by the OPE is called *fusion*.

A CFT with a finite number of primaries is called a rational CFT [103, 102]. This result is, of course, the same statement that we made in section 10.3.7 that the FQH state has m distinct anyons (vortices). This result also means that a vertex operator of charge m , i.e. an electron, is indistinguishable from the *identity operator*, V_0 . In this sense, the allowed primaries are fields which are *local* respect to the electron operator V_m . From the point of view of the FQH state, an operator that creates or destroys an electron (at fixed filling fraction) has no effect, as the FQH state is a fluid made of electrons. This also means that all physical operators must braid trivially with the electron operator. In a CFT an operator of this type is said to generate and extended symmetry algebra [267, 268].

Furthermore, the OPE of the chiral current $j(z)$ with the vertex operator $V_1(z)$ is

$$\lim_{w \rightarrow u} j(w)V_1(u) = \frac{1/m}{z-w} V_1(u) \quad (483)$$

which implies that the vertex operator represents a state with charge $1/m$ (in units of the electric charge e).

The CFT approach to construct FQH states has become a very powerful tool. We will see in section 10.3.11 that FQH states on an open geometry (e.g. a disk) have edge states which can be understood a chiral CFTs. In addition to providing a deeper understanding of more general abelian (multi-component) FQH states with a K -matrix structure, new classes of of FQH states with non-abelian (multi-dimensional) representations of the Braid Group has been proposed. The anyons of these states are of particular interest since they have been proposed originally by Kitaev in 1997 [169] that anyons can be used as physical qubits for topologically-protected quantum computation [169, 269, 270, 271, 170].

So far we discussed the case of abelian anyons. Abelian anyons have the property that the fusion of two anyons is another anyon. Similarly, a braiding operation between two anyons is equivalent (up to a phase factor) to another anyon. Mathematically this means that abelian anyons are one-dimensional representations of the Braid group. The fractional statistics of an anyon is then a label of the representation of the Braid group. As we saw, the same holds under fusion.

However, the Braid group generally admits multi-dimensional representations. In this more general case the fusion (or braiding) of two anyons is represented by a linear combination (superposition) of anyon states. Linear combinations of anyon states are represented by unitary matrices of rank greater than 1. Since matrices generally do not commute braiding and/or fusion processes correspond to a multiplication of these matrices. Such unitary processes are then regarded as quantum gates. This concept is the physical basis of topological quantum computation. It is currently the cutting edge of research in the field.

We will now discuss the simplest non-abelian FQH states, the Moore-Read (MR) states [266]. The Moore-Read wavefunctions are

$$\Psi_{MR}(z_1, \dots, z_N) = \text{Pf} \left(\frac{1}{z_i - z_j} \right) \prod_{i < j} (z_i - z_j)^n \exp \left(-\frac{1}{4\ell_0^2} \sum_{i=1}^N |z_i|^2 \right) \quad (484)$$

which describes a FQH of electrons at filling fraction $\nu = 1/n$, with n an *even* integer. Here $\text{Pf}(1/(z_i - z_j))$ is the Pfaffian of the matrix $1/(z_i - z_j)$.

This state was motivated by the discovery of an unexpected FQH state in the $N = 1$ Landau level with an *even denominator* filling fraction $\nu = 5/2$ by Willett and coworkers in 1987 [272, 273, 274]. Until then (and since then) this is the only FQH state with an even denominator filling fraction. The Pfaffian factor has a simple pole when two coordinates approach each other. However, provided $n \geq 1$ the “Laughlin factor” in the MR wavefunction cancels the singularity. We will shortly discuss the CFT construction of the Moore-Read state. The poles in the MR wavefunction suggest that in this state the electrons can get closer to each other than in a Laughlin state. This observation suggests that there may be some form of attraction (or suppression of repulsion) between the electrons in the MR state and motivated the notion that the observed even-denominator plateau may be physically related to some sort of pairing interaction. Shortly after the Moore-Read proposal was made, a paired wave function at $\nu = 1/2$ with a Pfaffian factor was also proposed by Greiter, Wen and Wilczek [275, 276, 277] who suggested that it reflects electrons pairing in the $\ell = 1$ angular momentum channel in time-reversal breaking condensate with symmetry $p_x + ip_y$.

Although, following the logic behind the Kohn-Luttinger mechanism for paring in angular momentum state $\ell \geq 1$ with repulsive interactions [278, 279, 280] it may possible to get a paired state even in the lowest Landau level (although there is no evidence for it, so far), the $N = 1$ Landau level may be more hospitable to such a state. Indeed, in the $N = 1$ Landau level the single-particle states i have angular momentum greater or equal than 1, and their probability distributions are suppressed both at short and long distances; these states look like smoke-rings. The matrix elements of the Coulomb potential in the $N = 1$ Landau level are parametrically suppressed at short distances (compared with the states in the lowest, $N = 0$, Landau level). In this scenario, the $\nu = 5/2$ plateau is interpreted as a $\nu = 1/2$ fully polarized state of the $N = 1$ landau level, with an $N = 0$ Landau level filled with electrons with both spin polarizations in a state with $\nu = 2$. There is numerical evidence that an instability into a p-wave paired state such as the MR state is favored if the short-distance repulsion between the *composite fermions* is softened [281].

In fact, in 1983 Halperin [282] considered generalizations of the Laughlin state in which due to very strong attractive interactions electrons would form clusters which then condense into a Laughlin-type state. The simplest example was Laughlin state of bosons (paired electrons)

a paired state at $\nu = 1/2$ whose Laughlin quasi-hole carries charge $e/4$. Extensive numerical calculations are more compatible with the MR state than with the Halperin state of pairs [283]. We will see shortly that the Halperin state is an abelian relative of the MR state.

Moore and Read derived the wave function of Eq.(484) by considering a CFT with two sectors: a chiral boson CFT with compactification radius $R = \sqrt{n}$, and a chiral Majorana fermion CFT, i.e. the chiral part of the 2D Ising CFT, discussed in section 4.2. The chiral Ising CFT has central charge $c = 1/2$ and several primary fields: the identity I , the twist field σ , and the (chiral) Majorana fermion χ with (chiral) scaling dimensions 0, 1/16, and 1/2, respectively. The propagator of the Majorana fermion, which is a free field, is

$$\langle \chi(z) \chi(w) \rangle = \frac{1}{z - w} \quad (485)$$

The primary field σ , the “twist field”, is non-local with respect to the Majorana fermion χ , and changes its boundary conditions from periodic to antiperiodic.

The fusion rules of the chiral Ising CFT are

$$\sigma \star \chi = I, \quad \sigma \star \sigma = I + \chi, \quad \sigma \star \chi = \chi \quad (486)$$

What will be important below is that the σ field has two fusion channels. As in the analysis of the Laughlin states, the compactified boson n primaries (anyons), the vertex operators $V_p(x) = \exp(ip\phi(x)/n)$, with $n = 0, 1, \dots, n$ and there are n types of anyons, and a (charge) current $J = \frac{i}{\sqrt{n}}\partial_z\phi(z)$.

The first task is to identify the electron operator which has to be a fermion and has to carry electric charge. This means that it is a composite of an operators (with Fermi statistics) the (neutral) chiral ising CFT and a vertex operator whose charge is an integer multiple of e . The desired electron operator is $\psi_e(z) = \chi(z) \exp(i\sqrt{n}\phi(z))$ whose correlator is $\langle \chi(z)V_n(z)\chi(w)V_{-n}(z) \rangle = 1/(z-w)^{1+n}$, which is a fermion with charge e .

The n -point function of the chiral Majorana fermions is

$$\langle \chi(z_1) \dots \chi(z_N) \rangle = \text{Pf} \left(\frac{1}{z_i - z_j} \right) \quad (487)$$

which follows fro applying Wick’s theorem. Then we see that the MR wave function is the expectation value

$$\Psi_{MR}(z_1, \dots, z_N) = \langle \chi(z_1) \dots \chi(z_N) \rangle \times \left\langle \left(\prod_{i=1}^N \exp(i\sqrt{n}\phi(z_i)) \right) \exp \left(-i \int d^2 z' \sqrt{n} \rho_0 \phi(z') \right) \right\rangle \quad (488)$$

Next we need to identify the primary fields of the tensor product of the chiral Ising CFT, \mathbb{Z}_2 , and the chiral U(1) CFT with compactification radius $R = \sqrt{n}$. The allowed primary field must be local with respect to the electron operator, ψ_e . This leaves us with four primary fields: a) the identity I , b) the non-abelian vortex $\sigma(z) \times \exp\left(\frac{i}{2\sqrt{n}}\phi(z)\right)$ with charge $e/(2n)$ and non-abelian statistics, c) the charge-neutral Majorana fermion χ , and d) the abelian vortex (the Laughlin quasi-hole) $\exp\left(\frac{i}{\sqrt{n}}\phi(z)\right)$, with charge e/n and abelian statics $\delta = \pi/n$.

The he new feature here is the non-abelian fractional statistics of the non-abelian vortex (sometimes called a “half-vortex”). Its non-abelian character is a consequence of the fact that the fusions of two twist fields has two channels, see Eq.(486). This implies that the wave function of four non-abelian vortices can be expressed as a linear combination of two so-called *conformal blocks*. In other words, this wave function belongs to a degenerate two-dimensional Hilbert space, each component labeled by a conformal block [102]. A braiding operation between two non-abelian vortices is a monodromy of the wave function that induces a unitary transformation U in this Hilbert space

$$U = \frac{1}{\sqrt{2}} \exp \left[i\pi \left(\frac{1}{8} + \frac{1}{4n} \right) \right] \times \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (489)$$

Therefore, the degenerate Hilbert space of four quasiholes provides a two-dimensional representation of the Braid Group. This observation is the basis for regarding a state of four non-abelian quasiholes as a qubit. Furthermore, Nayak and Wilczek showed that a state of $2p$ quasiholes spans a 2^{p-1} -dimensional degenerate Hilbert space [284]. This means that, for large p , the degeneracy per quasihole is $\sqrt{2}$. In other words, this degeneracy is not due to a local degree of freedom attached to each quasihole but that it is shared in a non-local fashion!

In what sense are the Moore-Read states paired? In an insightful paper Read and Green [285] used a BCS theory approach to investigate the pairing properties of a condensate of composite fermions in the $p_x + ip_y$ channel. The mean-field BCS Hamiltonian is

$$H_F = \int d^2k \left[(\varepsilon(\mathbf{k}) - \mu) \psi^\dagger(\mathbf{k}) \psi(\mathbf{k}) + \frac{1}{2} (\Delta^*(\mathbf{k}) \psi(-\mathbf{k}) \psi(\mathbf{k}) + \Delta(\mathbf{k}) \psi^\dagger(\mathbf{k}) \psi^\dagger(-\mathbf{k})) \right] \quad (490)$$

where $\psi(\mathbf{k})$ is the composite fermions field, $\varepsilon(\mathbf{k}) = \frac{\mathbf{k}^2}{2M}$, and $\Delta(\mathbf{k})$ is the gap function. For a $p_x + ip_y$ condensate the gap function the gap function transforms under rotations as an $\ell = -1$ angular momentum eigenstate, and for $\mathbf{k} \rightarrow 0$ it behaves as $\Delta(\mathbf{k}) = (k_x - ik_y)\Delta$, where Δ is a constant pairing amplitude.

The BCS ground state has the form

$$|G\rangle = \prod_{\mathbf{k}} (u(\mathbf{k}) + v(\mathbf{k}) \psi^\dagger(\mathbf{k}) \psi^\dagger(-\mathbf{k})) |0\rangle \quad (491)$$

where $|u(\mathbf{k})|^2 = |v(\mathbf{k})|^2 = 1$. The amplitudes $u(\mathbf{k})$ and $v(\mathbf{k})$ satisfy the Bogoliubov-de Gennes Equation (BdG)

$$\begin{pmatrix} \xi(\mathbf{k}) & -\Delta^*(\mathbf{k}) \\ -\Delta(\mathbf{k}) & -\xi^*(\mathbf{k}) \end{pmatrix} \begin{pmatrix} u(\mathbf{k}) \\ v(\mathbf{k}) \end{pmatrix} \equiv E(\mathbf{k}) \hat{\mathbf{n}}(\mathbf{k}) \cdot \boldsymbol{\sigma} \begin{pmatrix} u(\mathbf{k}) \\ v(\mathbf{k}) \end{pmatrix} = E(\mathbf{k}) \begin{pmatrix} u(\mathbf{k}) \\ v(\mathbf{k}) \end{pmatrix} \quad (492)$$

where $\xi(\mathbf{k}) = \varepsilon(\mathbf{k}) - \mu$, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices, and $\hat{\mathbf{n}}(\mathbf{k}) = (-\text{Re}\Delta(\mathbf{k}), \text{Im}\Delta(\mathbf{k}), \xi(\mathbf{k}))$ is a unit vector defined for every \mathbf{k} . The eigenvalues $E(\mathbf{k})$ and the eigenvectors $(u(\mathbf{k}), v(\mathbf{k}))$ are

$$E(\mathbf{k}) = \sqrt{\xi^2(\mathbf{k}) + |\Delta(\mathbf{k})|^2}, \quad \frac{v(\mathbf{k})}{u(\mathbf{k})} = -\frac{E(\mathbf{k}) - \xi(\mathbf{k})}{\Delta^*(\mathbf{k})} \quad (493)$$

The spinor amplitudes are $|u(\mathbf{k})|^2 = \frac{1}{2}(1 + \frac{\xi(\mathbf{k})}{E(\mathbf{k})})$ and $v(\mathbf{k}) = \frac{1}{2}(1 - \frac{\xi(\mathbf{k})}{E(\mathbf{k})})$.

In the low momentum limit and in real space the BdG Equations become

$$\begin{aligned} i\partial_t u &= -\mu u + \Delta^* i(\partial_x + i\partial_y)v \\ i\partial_t v &= \mu v + \Delta i(\partial_x - i\partial_y)u \end{aligned} \quad (494)$$

which is just the Dirac Equation in 2+1 dimensions, with μ playing the role of the mass, and with the restriction that the spinor (u, v) obeys the Majorana condition

$$(u, v) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \quad (495)$$

The Majorana condition is obeyed in all superconductors and reflects that fact that in these condensates only the fermion parity $(-1)^{N_F}$ (with N_F being the number of fermions) is conserved, while the fermion number N_F is not.

This is a good BCS state in that it is fully gapped and it is chiral. Assuming that the pair fields in the p_x and p_y channels are equal, they showed that the BCS ground state $|G\rangle$ has the form

$$|G\rangle \sim \exp \left(\frac{1}{2} \sum_{\mathbf{k}} g(\mathbf{k}) \psi^\dagger(\mathbf{k}) \psi^\dagger(-\mathbf{k}) \right) |0\rangle \quad (496)$$

Projected onto a state with N fermions with real space coordinates \mathbf{x}_i the wave function is a Pfaffian

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \langle \mathbf{x}_1, \dots, \mathbf{x}_N | G \rangle = \text{Pf}(g(\mathbf{x}_i - \mathbf{x}_j)) \quad (497)$$

The long distance behavior of this wave function depends on whether the chemical potential $\mu > 0$ (this is the weak-pairing or BCS regime), or $\mu < 0$ (which is the strong pairing BEC-like regime). While in the case of an s-wave superconductor these two regimes are smoothly connected by a crossover, in the $p_x + ip_y$ case they are separated by a quantum phase transition at $\mu = 0$. In the weak pairing regime the function $g(\mathbf{k})$ has the asymptotic long distance form, as $\mathbf{k} \rightarrow 0$,

$$g(\mathbf{k}) \simeq -\frac{2\mu}{(k_x + ik_y)\Delta^*} \quad (498)$$

where the pair field behaves as $\Delta(\mathbf{k}) \simeq (k_x - ik_y)\Delta$. In real space (in complex coordinates) the function $g(z)$ becomes

$$g(z) = \left(\frac{i\mu}{\pi\Delta^*} \right) \frac{1}{z} \quad (499)$$

which is the form used in the Pfaffian wave function. On the other hand, in the strong pairing regime, $\mu < 0$, $g(\mathbf{k})$ has the asymptotic behavior

$$g(\mathbf{k}) \simeq -A \frac{(k_x - ik_y)}{a_0^{-2} + \mathbf{k}^2} \quad (500)$$

where A and a_0 are functions of Δ and μ ,

$$A = \frac{2|\mu|M\Delta}{2|\mu| + m|\Delta|^2}, \quad a_0 = \frac{1}{2|\mu|} \sqrt{\frac{2|\mu|}{M} + |\Delta|^2} \quad (501)$$

In real space, at distances $|x| \gg a_0$ the Fourier transform of $g(\mathbf{k})$ decays exponentially, and as $a_0 \rightarrow \infty$, $\mu \rightarrow 0$ and the Fourier transform $g(z)$ has the power-law behavior

$$g(z) \simeq \left(\frac{i|\Delta|}{2\pi\Delta^*} \right) \frac{1}{z|z|} \quad (502)$$

This means that $\mu = 0$ is a quantum critical point that separates the weak pairing phase from the strong pairing phase, which have distinct properties.

Read and Green then used a topological argument, originally formulated by Grisha Volovik [286] in the context of superfluid ${}^3\text{He}$ films. To see this we notice that the three component unit vector $\hat{\mathbf{n}}(\mathbf{k})$ takes values on the surface of a sphere S_2 . In addition, since $v(\mathbf{k}) \rightarrow 0$ for $\mathbf{k} \rightarrow \infty$ (and, hence $u(\mathbf{k}) \rightarrow 1$), we can wrap the momentum space \mathbf{k} onto a sphere S_2 . Thus $\hat{\mathbf{n}}(\mathbf{k})$ are smooth functions of $S_2 \mapsto S_2$. such maps are classified into the homotopy classes of $\pi_2(S_2) \simeq \mathbb{Z}$ given in terms of the integer-valued topological invariant

$$Q = \frac{1}{4\pi} \int d^2 k \hat{\mathbf{n}}(\mathbf{k}) \cdot \partial_{k_x} \hat{\mathbf{n}}(\mathbf{k}) \times \partial_{k_y} \hat{\mathbf{n}}(\mathbf{k}) \quad (503)$$

In the strong pairing phase $\mu < 0$ and $\xi(\mathbf{k}) > 0$ and $\hat{\mathbf{n}}(\mathbf{k})$ takes values only on the northern hemisphere of the target space S_2 . Such maps can be deformed continuously to the North Pole and, hence, are topologically trivial, $Q = 0$. On the other hand, in the weak pairing phase, $\mu > 0$, $\xi(\mathbf{k})$ takes both positive and negative values. In this phase $\hat{\mathbf{n}}(\mathbf{k})$ is topologically non-trivial and the topological invariant Q takes non-trivial values ± 1 .

The upshot of this analysis is that, in the weak pairing phase, the paired FQH state is, at this mean field level, a two-dimensional *topological superconductor*. Read and Green [285] further showed that this pairing state has vortices with an interesting fermionic spectrum. A vortex is a state where at long distances the (complex) amplitude of the $p_x + ip_y$ condensate behaves as $\Delta \exp(i\varphi)$, where φ is the azimuthal angle, which winds by 2π on a circumference of large radius R . The BdG equation, Eq.(494), has zero mode solutions. In polar coordinates (r, φ) the BdG Equation the zero modes satisfy

$$\begin{aligned} \Delta i e^{i\varphi} \left(\partial_r + \frac{i}{r} \partial_\varphi \right) v &= \mu u \\ \Delta i e^{-i\varphi} \left(\partial_r - \frac{i}{r} \partial_\varphi \right) u &= -\mu v \end{aligned} \quad (504)$$

The normalizable zero mode spinor solution is

$$\begin{pmatrix} u(r, \varphi) \\ v(r, \varphi) \end{pmatrix} = \frac{f(r)}{\sqrt{r}} \begin{pmatrix} \frac{1}{\sqrt{i}} e^{i\varphi/2} \\ \frac{1}{\sqrt{-i}} e^{-i\varphi/2} \end{pmatrix} \quad (505)$$

where $f(r)$ is given by

$$f(r) \sim \exp \left(- \int_0^r dr' \frac{\mu(r')}{|\Delta|} \right) \sim \exp(-\mu r / |\Delta|) \quad (506)$$

Under a 2π rotation this spinor solution is double-valued

$$(u(r, \varphi + 2\pi), v(r, \varphi + 2\pi)) = -(u(r, \varphi), v(r, \varphi)) \quad (507)$$

which follows from the global phase invariance. In fact, in all paired states, topological or not, under a global transformation of the pair field by a phase θ , the (composite) fermion must transform with a phase of $\theta/2$,

$$\Delta(\mathbf{x}) \mapsto e^{i\theta} \Delta(\mathbf{x}), \quad \psi(\mathbf{x}) \mapsto e^{i\theta/2} \psi(\mathbf{x}), \quad \psi^\dagger(\mathbf{x}) \mapsto e^{-i\theta/2} \psi^\dagger(\mathbf{x}) \quad (508)$$

In other words, in a paired state the fermion is non-local to the vortex. Hence, under a 2π phase transformation of the pair field, the fermions must change sign, and the spinor zero mode solution must be double-valued. This means that this state has a branch cut.

The structure of the vortex and of the fermion zero mode is closely related to the problem of solitons with fractional charge that we discussed in section 8.3. In fact, Roman Jackiw and Paolo Rossi [287] investigated a closely related problem of a theory of Dirac field in 2+1 dimensions coupled to a charged scalar field through a Yukawa coupling with the form of a Majorana mass. They showed that such a theory admits vortex solutions with fermion zero modes. In a subsequent paper Erik Weinberg [288] showed that these zero modes are counted by an index theorem which relates the number of zero modes to the vorticity. More recently, Nishida, Santos and Chamon [289] that the relativistic theory of Jackiw and Rossi in the non-relativistic approximation reduces to the theory of a $p_x + ip_y$ superconductor.

There is, however, a subtle but profound difference between the fermion zero modes of the one-dimensional fractionally charged solitons and the fermion zero modes of a superconductor. In the case of the 1D solitons, the zero modes can be either occupied or empty rendering a soliton charge of $-e/2$ or $+e/2$. However, in the case of the superconductor, the BdG fermions are charge-neutral since their charge has gone into the condensate. Thus, in a superconductor fermion number is not conserved, and only the fermion parity is conserved. This means that the field operator associated with the vortex zero mode, which we will denote by γ , must be a self-adjoint operator, $\gamma^\dagger = \gamma$.

Ivanov [290] showed that this behavior is the origin of the non-abelian fractional statistics in this system. Indeed, if one considers a configuration with $2n$ vortices, each will carry a Majorana zero mode γ_i . Since these are fermion operators they satisfy the usual anticommutator algebra, $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. We can (arbitrarily) group the $2n$ Majorana fermion operators into n pairs. For each pair we can define a complex Dirac operator $\psi_j = (\gamma_{2j} + i\gamma_{2j+1})/2$ (and its adjoint), which satisfies the usual fermionic algebra, $\{\psi_j, \psi_k^\dagger\} = \delta_{jk}$. Each Dirac fermion can be either in an empty state or in an occupied state. Thus a system of $2n$ vortices supports a degenerate Hilbert space of dimension 2^{n-1} , which agrees with the results of Nayak and Wilczek [284]. However, the assignment of Majorana operators to specific pairs is actually arbitrary, which amounts to a particular definition of the branch cuts. Changing the assignment of the operators into pairs is then equivalent to a rearrangement of the cuts. In 2006 Michael Stone and Suk Bum Chung [291] showed that these vortices obey the braiding and fusion properties of *Ising* anyons. These properties follow from the branch cut configurations which affect the monodromies of the vortices. It is important to stress that the vortices with zero modes are the non-abelions. Majorana fermions are fermions and, as such, are abelian.

In addition to the Moore-read Pfaffian state, the CFT approach has led to the formulation of new non-trivial FQH states. Read and Rezayi [292] proposed a series of so-called *cluster* states, which generalize the concept of paired states. They investigated a particular type of cluster states in which the Pfaffian factor of the MR state is replaced by a correlator of *parafermion* primary fields of a \mathbb{Z}_k CFT of Zamolodchikov and Fateev [293]. Parafermions were originally introduced by Fradkin and Kadanoff [152] (see also Ref.[294]) as a generalization of the fermions of the 2D Ising model to the \mathbb{Z}_k clock model. In this model one can define several types of parafermions each consisting of the fusion of a charge operator and a magnetic (disorder) operator. These operators obey an algebra of the type $AB = \exp(ip2\pi/k)BA$, where the integer p depends on the electric and magnetic charges of the parafermions, which is reminiscent of fractional statistics. In close analogy with the Majorana fermions of the 2D Ising model, \mathbb{Z}_k CFT describes the behavior of these systems their self-dual points.

The \mathbb{Z}_k CFT is actually much richer and has more primary fields than the \mathbb{Z}_2 case of the Ising model. As a CFT, the \mathbb{Z}_k theory is the same as the CFT on the coset $SU(2)-k/U(1)$. A coset means that a $U(1)$ sector has been projected out of the spectrum. Here we will focus on a special case of the \mathbb{Z}_3 CFT. This CFT has a parafermion primary field, which we will call ψ , and a non-abelian primary that we will call τ . Read and Rezayi [292] proposed to replace the Pfaffian factor of the MR state with a correlator of N parafermions of a \mathbb{Z}_3 CFT and a Laughlin factor with an exponent of $n + 2/3$. Such states require that the number of electrons N be a multiple of 3. Thus, these states can be viewed as states in which the electrons cluster in groups of 3. They also considered the more general case of the \mathbb{Z}_k CFT in which case N is a multiple of k . The resulting filling fractions are $\nu = k/(mk + 2)$ (with $m \geq 0$). The Read-Rezayi $k = 3$ parafermion state is a leading candidate for the observed FQH plateau at $\nu = 12/5 = 2 + 2/5$ [295, 296] (the competing state being the 2/5 Jain state). There is strong numerical support for the 12/5 state

being a \mathbb{Z}_3 parafermion [297].

The quasiholes obtained by inserting the primary fields τ in the correlator of parafermions have interesting properties which stem from their basic fusion rule

$$\tau \star \tau = I + \tau \quad (509)$$

Read and Rezayi [292] showed that the number of conformal blocks for $3n$ quasiholes, i.e. the number of degenerate states, in the parafermion theory is the Fibonacci number F_{3n-2} , where $F_1 = 1$, $F_2 = 2$, $F_3 = 3$, $F_4 = 5$. In general $F_m = F_{m-1} + F_{m-2}$. For $m \rightarrow \infty$, the rate of increase approaches the limit $\lim_{m \rightarrow \infty} F_m/F_{m-1} = (1 + \sqrt{5})/2$ which is known as the Golden Mean. In other words, for large m , the dimension of the Hilbert space increases exponentially as $((1 + \sqrt{5})/2)^m$. Aside from these interesting mathematical curiosities, the significance of this fusion rule is that these states, regarded as “qubits”, define unitary transformations that cover uniformly the Bloch sphere which is required for universal quantum computation [270].

At the level of topological quantum field theory the Moore-Read and Read-Rezayi states are related to the non-abelian Chern-Simons gauge theory with gauge group $SU(2)$ at Chern-Simons level k . In the more general case of the gauge group $SU(N)$ on a manifold \mathcal{M} action is given by

$$S_{\text{CS}}[A_\mu] = \frac{k}{4\pi} \int_{\mathcal{M}} d^3x \epsilon_{\mu\nu\lambda} \left(A_a^\mu \partial^\nu A_a^\lambda + \frac{2}{3} f_{abc} A_a^\mu A_b^\nu A_c^\lambda \right) \quad (510)$$

where $k \in \mathbb{Z}$ is the level, the gauge fields $A^\mu = A_a^\mu t_a$ take values in the algebra of $SU(N)$, with $\{t_a\}$ being the $N^2 - 1$ generators, and f_{abc} the structure constants of $SU(N)$. The expectation values of Wilson loop operators of this theory compute the Jones polynomials which are topological invariants of knot theory [165]. The connection with $SU(2)_k$ Chern-Simons gauge theory has been essential in formulating effective low energy quantum field theories for the non-abelian states [298, 299, 300, 301, 302, 303].

10.3.11. Edge States and Chiral Conformal Field Theory

We will now discuss the edge states of the FQH states. As we saw the FQH states are incompressible in the bulk and all bulk excitations are gapped. The edge of the region occupied by the fluid (in many cases that edge of the physical sample) is where the bulk gap collapses and hence where the system has low energy excitations.

The role of edges and their nature can be seen already in the simple case of the integer Hall effect treated as a system of free fermions in the lowest Landau level. In this state all single particle states are occupied, and the bulk ground state wave function is a Slater state which takes the form of the Vandermonde determinant of Eq.(377). The potential that keeps the electrons inside the sample increases monotonously near the edge. As we discussed in section 10.3.1, in this region the electrons experience an electric field E that pushes them into the sample, and in the presence of the perpendicular magnetic field B the electrons move *along the edge* at the drift velocity $v = c|E|/|B|$. At some spatial location, corresponding to the locus of a single particle Landau state, the potential crosses the Fermi energy and in that region potential is essentially a linear function of the coordinate and hence of momentum (or angular momentum depending on the gauge that is being used). In other words, the low energy states are one branch of a one-dimensional chiral excitation, such as in the right-moving states of our discussion of the chiral anomaly in section 7.2

In section 9.5 we showed that a Chern-Simons gauge theory on a manifold with a boundary projects onto a chiral CFT on that boundary. In that section we showed that an abelian Chern-Simons theory $U(1)_m$ integrates to the boundary, the 1+1-dimensional Minkowski spacetime $S^1 \times \mathbb{R}$, as a chiral $U(1)_m$ CFT for a compactified boson with compactification radius $R = 1$ whose action is given by

$$S[\varphi] = \int_{S^1 \times \mathbb{R}} d^2x \frac{1}{4\pi} [\partial_0 \phi \partial_1 \phi - v(\partial_1 \phi^2)] \quad (511)$$

where we rescaled the compactified scalar by a factor of \sqrt{m} to that its compactification radius $R = \sqrt{m}$ as in section 10.3.10. The only difference between the theory of Eq.(241) and what we did in section 10.3.10 is that this theory is in a 1+1-dimensional Minkowski spacetime $S^1 \times \mathbb{R}$ while before we were in Euclidean spacetime. In the case of the integer quantum Hall effect $v = 1$ (and hence $m = 1$) and the (time-ordered) correlator of the vertex operator, $V_1(x, t) = \exp(i\phi(x, t))$ is

$$\langle T \exp(i\phi(x, t)) \exp(-i\phi(0, 0)) \rangle = \frac{1}{x - vt - i\varepsilon} \quad (512)$$

which is, indeed, the propagator of a chiral free fermion.

On the other hand, for the Laughlin states at filling fraction $\nu = 1/m$ (and compactification radius $R = \sqrt{m}$) the propagator of the electron $\psi_e \sim \exp(i\sqrt{m}\phi)$ is

$$\langle T \exp(i\sqrt{m}\phi(x, t)) \exp(-i\sqrt{m}\phi(0, 0)) \rangle \sim \frac{1}{(x - vt - i\varepsilon)^m} \quad (513)$$

while the propagator of the quasi-hole $\exp\left(\frac{i}{\sqrt{m}}\phi\right)$ is

$$\langle T \exp\left(\frac{i}{\sqrt{m}}\phi(x, t)\right) \exp\left(-\frac{i}{\sqrt{m}}\phi(0, 0)\right) \rangle \sim \frac{1}{(x - vt - i\varepsilon)^{1/m}} \quad (514)$$

Therefore, the CFT of the edge is the same as the CFT of the ideal wavefunction with the only difference that the former is in Minkowski spacetime while the latter is in Euclidean spacetime.

In this sense there is a one-to-one correspondence between the bulk and the edge. We can also see how that works by considering a fundamental Wilson arc in the bulk along a path $\gamma(x, y)$ where x and y are at the boundary

$$\langle W[\gamma] \rangle = \langle \exp(i \int_{\gamma(x,y)} dx_\mu \mathcal{A}^\mu) \rangle \equiv \langle \exp\left(\frac{i}{\sqrt{m}}\phi(x)\right) \exp\left(-\frac{i}{\sqrt{m}}\phi(y)\right) \rangle \quad (515)$$

where on the right hand side we have rescaled the field by \sqrt{m} , as before. The scaling dimensions of the electron, the quasi-hole and the current are $\Delta_e = \frac{m}{2}$, $\Delta_{qh} = \frac{1}{2m}$ and $\Delta_{current} = 1$. These results will be important shortly.

The same structure applies to the multi-component abelian FQH states. The only difference is that in multi-component FQH states the edge consists, in general, of a charge field which couples to the electromagnetic field and one or more neutral edge states. An example is the theory of the non-abelian Pfaffian states at filling fraction $\nu = 1/n$. In this case the edge theory consists of a compactified chiral boson ϕ of radius $R = \sqrt{n}$ and a chiral Majorana (neutral) fermion χ . At a formal level the Lagrangian for the edge state(s) is a sum of two terms

$$\mathcal{L} = \frac{1}{4\pi} (\partial_x \phi \partial_t \phi - v_c (\partial_x \phi)^2) + \chi i (\partial_t - v_n \partial_x) \chi \quad (516)$$

where v_c and v_n are the velocities of the (charged) compactified chiral scalar ϕ and of the neutral Majorana field χ . Numerically (and experimentally) it is known that the charge mode is (substantially) faster than the neutral mode(s), $v_c > v_n$, and often by significant factors.

A superficial reading of this Lagrangian suggests that these degrees of freedom are decoupled. However this is not correct. Only operators from both sectors which are local with respect to the electron $\psi_e \sim \chi \exp(i\sqrt{n}\phi)$ are physically allowed. Here an operator is local with respect to the electron means that the operator braids trivially with the electron. This condition restricts the allowed observables. In the case of the fermionic MR state, with $n = 2$, the allowed operators are the non-abelion $\sigma \exp(ip/2\sqrt{2})$ with scaling dimension is $\Delta = 1/8$ and electric charge is $Q = 1/4$, the Majorana fermion with scaling dimension $\Delta = 1/2$ and no electric charge $Q = 0$, and the Laughlin quasiparticle with scaling dimension $\Delta = 1/4$ and charge $Q = 1/2$. Its electron operator has scaling dimension $\Delta = 3/2$ and charge $Q = 1$.

Conformal field theory has an important defining universal quantity called the central charge which is closely related to the energy-momentum tensor of the theory [102]. For pedagogical introductions to CFT see Refs.[304, 45] and chapter 21 of Ref.[10]. The energy-momentum tensor $T_{\mu\nu}$ is a locally conserved current, $\partial^\mu T_{\mu\nu} = 0$. Its local conservation implies the global conservation of energy and momentum. As such, the energy-momentum tensor is a fundamental observable of any quantum field theory. In a *conformal* field theory, the energy-momentum is also the generator of local scale and conformal transformations. In the special case of 1+1-dimensional systems, such as the edge states of the FQH fluids, the energy momentum tensor has special and crucially important properties. In addition of being locally conserved, conformal invariance requires that $T_{\mu\nu}$ must be traceless, $T^\mu_\nu = 0$, since its trace is the generator of dilations. In this case, if the theory is chiral, the energy momentum tensor has only one (right-moving) component $T = (T_{00} - T_{11})/2$. In complex coordinates of the Euclidean metric, the correlator of the energy momentum tensor $T = T_{zz}$ is [305]

$$\langle T(z)T(w) \rangle = \frac{c/2}{(z - w)^4} \quad (517)$$

where c is a universal quantity known as the central charge of the CFT. In the case of the compactified boson the central charge is $c = 1$ whereas for a massless Majorana fermion $c = 1/2$.

The central charge of the CFT enters in many observables of fundamental physical importance. For example, for a 1+1-dimensional system of length L , such as the edge state of a FQH droplet, the ground state energy E_{gnd} for large L has the behavior [306, 307]

$$E_{\text{gnd}} = \varepsilon_0 L - c \frac{\pi v}{6L} + O(1/L^2) \quad (518)$$

where ε_0 is the ground state energy density, which is a non-universal quantity, c is the central charge of the CFT and v is the velocity of the massless modes. The second term in Eq.(518) is known as the Casimir energy. Similarly, the free energy density $f(T)$ of a CFT has the low temperature Stephan-Boltzmann behavior (in one dimension)

$$f(T) = \varepsilon_0 + c \frac{\pi T^2}{6v} + O(T^3) \quad (519)$$

where we set the Boltzmann constant to unity. The low-temperature specific heat $c(T)$ is

$$c(T) = c \frac{\pi T}{3v} + O(T^2) \quad (520)$$

In a chiral CFT, such as the edge states of the FQH fluids, the central charge enters in the low temperature behavior of the (Hall) thermal conductivity κ_{xy} , [285]

$$\kappa_{xy} = c \frac{\pi T}{6\hbar} \quad (521)$$

Much of what is known about FQH states comes from experiments involving their edge states. In a series of exquisite experiments Granger, Eisenstein and Reno [308] measured κ_{xy} in the edge states of a $v = 1$ quantum Hall fluid and showed that the heat transport is indeed chiral, i.e. the edge state behaves as a heat “diode”. This effect was confirmed in the FQH states by Bid and coworkers [309] who, in addition, used this effect to detect the neutral modes in several Jain states at filling fractions 2/3 and 2/5 and in the non-abelian state at $v = 5/2$.

The simplest experimental probes are quantum point contacts and typically are of two types: inter-edge tunneling inside a FQH liquid or tunneling of electrons into the edge state of a FQH liquid. Since the bulk is gapped, only the edge states participate in these tunneling processes. In the general case we will have two edges and a tunnel process at a point contact. Here the two edges can be either the two edges of the same FQH state, in which case the process happens inside the FQH liquid and involves tunneling of quasiparticles, or the edges of different liquids, in which case this process is external and involves tunneling of electrons. Problems of these types were first investigated by Charles Kane and Matthew Fisher [310] which led to a considerable amount of work on these problems.

Let us consider first the case of tunneling into the edge state of a $v = 1/m$ Laughlin state from a Fermi liquid, which we will take to be a QH fluid in the $v = 1$ state. The point contact is at $x = 0$. The total Lagrangian is

$$\mathcal{L} = \mathcal{L}_{\text{edge}} + \mathcal{L}_{\text{FL}} - \Gamma e^{i\omega_0 t} \psi_{e,\text{edge}}^\dagger(0, t) \psi_{e,\text{FL}}(0, t) + \text{h.c.} \quad (522)$$

where $\omega_0 = eV/\hbar$, and V is the bias voltage between the two fluids, and Γ is a tunneling matrix element. The local electron spectral density (the density of states) $N(\omega)$ at energy ω is

$$N(\omega) = \text{Im} \lim_{x \rightarrow 0^+} \int_{-\infty}^{\infty} dt G_e(x, t) e^{i\omega t} = \text{const. } |\omega|^{m-1} \quad (523)$$

where $G_e(x, t) = \langle T\psi_e^\dagger(x, t)\psi_e(0, 0) \rangle$ is the electron correlator in the FQH edge, shown in Eq.(512). This result, combined with Fermi’s Golden Rule for a point contact with voltage bias V , predicts a tunneling current from a Fermi liquid (FL) into the chiral Luttinger liquid (CLL) of the edge state to be [311]

$$I(V) = 2\pi \frac{e}{\hbar} |\Gamma|^2 \int_{-eV}^0 dE N_{\text{CLL}}(E, T) N_{\text{FL}}(E + eV, T) \propto V^m \quad (524)$$

and a tunneling differential conductance $G(V)$

$$G(V) = \frac{dI}{dV} = 2\pi \frac{e}{\hbar} |\Gamma|^2 N_{\text{FL}}(0) N_{\text{CLL}}(V, T) \propto V^{m-1} \quad (525)$$

which is non-Ohmic and vanishes as $V \rightarrow 0$.

Early experiments by Albert Chang and coworkers , on a geometry that (most likely) had many point contacts, confirmed the predicted CLL behavior [312, 313], although there were discrepancies in the measured exponents. The behavior seen in more recent experiments by Cohen and coworkers [314], on a point contact in graphene, are consistent with the theoretical predictions [311]. Interestingly, these newer class of experiments [315, 314] also show evidence of an analog of Andreev reflection predicted to exist near the strong coupling fixed point of the theory of Sandler, Chamon and Fradkin [316] as a consequence of electron fractionalization.

Most experiments are done on a geometry call a Hall bar in which the QH fluid occupies a rectangular region with its length L larger than its width W . In this geometry, there are chiral edge states at opposite sides of the Hall bar propagating in opposite directions. One type of point contact consists in creating a constriction in the quantum Hall fluid by applying a gate, a bias potential on a narrow strip accross the Hall bar. This gate repels the electrons in the QH fluid, forcing the opposite edges to approach each other in the proximity of the gate. In the absence of the gate, momentum is conserved on each edge which forbids tunneling accross the Hall bar since the edge states have opposite Fermi momentum. However, the gate breaks translation invariance on both edges and tunneling between them is now allowed. In other words, the gate creates a point contact between the opposite propagating edge states leading to a tunneling process between the edges of the QH liquid. Thus, this is internal tunneling as oppose the process of tunneling between two different liquids that we discussed above.

The tunneling Lagrangian for this system has the same form as in Eq.(522) except that now the two edges are identical. The theory Kane and Fisher shows that in the factional case the most relevant process is the tunneling of FQH quasiparticles. The Fermi Golden Rule argument used in Eq.(525) to compute the differential conductance also applies in the present case except that now the two densities of states are the densities of states of the (Laughlin) quasiparticles (instead of electrons). The local quasiparticle density of states is $N_{qp}(\omega) \sim |\omega|^{\frac{2}{m}-1}$. Then, the differential tunneling conductance $G(V)$ in this case is

$$G(v) \sim V^{2(\frac{1}{m}-1)} \quad (526)$$

This behavior is also non-Ohmic but, unlike the case of electron tunneling of Eq.(522), the differential tunneling conductance now *diverges* as $V \rightarrow 0$. This behavior means that the $\Gamma \rightarrow 0$ fixed point is *unstable* and that the point contact flow to a *strong coupling fixed point* at $\Gamma \rightarrow \infty$. Early experiments by Milliken, Umbach and Webb in 1996 [317] showed indications of CLL behavior. The predicted behavior was confirmed in the experiments of Roddaro and coworkers [318, 319].

In the RG language, we can define a dimensionless tunneling amplitude g by $\Gamma = a^{\Delta-1} g$, where Δ is the scaling dimension of the tunneling operator and a is a UV cutoff. The “tree-level” beta function is readily found to be

$$\beta(g) = a \frac{\partial g}{\partial a} = (1 - \Delta) g + O(g^2) \quad (527)$$

which shows that if the scaling dimension of the operator of the tunneling particle is $\Delta < 1$, then this tunneling process is relevant, while is $\Delta > 1$ it is irrelevant.

Kane and Fisher argued that there should be a crossover from the IR unstable weak-coupling fixed point governed by quasiparticle tunneling to an IR stable strong coupling fixed point governed by electron tunneling. This crossover is reminiscent of the the crossover in quantum impurity problems such as the Kondo problem of a magnetic impurity in a metal [320, 32, 52, 53, 321]. The main difference is that the impurity coupling is marginal and the crossover scale, the Kondo temperature T_K , which is related to the coupling constant g as $T_K \sim \exp(-1/g)$, while in the FQH constriction the crossover scale is $T_K(g) \sim g^{1/(1-\Delta)}$. Kane and Fisher argued that this crossover can be viewed as a process that interpolated between a FQH fluid with a weak constriction to a regime in which the Hall bar splits into two pieces with weak electron tunneling between them.

It turns out that, after some manipulations, the model of the constriction can be mapped into a one-dimensional compactified boson φ on a semi-infinite line, $x \geq 0$, with a vertex operator acting at the boundary. The action of this system, known as boundary sine-Gordon is

$$S = \frac{1}{8\pi} \int_{-\infty}^{\infty} dt \int_0^{\infty} dx (\partial_\mu \varphi)^2 + \Gamma_{qp} \int_{-\infty}^{\infty} dt \cos\left(\sqrt{\frac{\nu}{2}} \varphi(0, t)\right) \quad (528)$$

This theory has two fixed points: a) the IR unstable fixed point at $\Gamma_{qp} = 0$ where the field Neumann boundary conditions at $x = 0$, $\partial_x \varphi = 0$, and b) an IR stable fixed point at $\Gamma_{qp} \rightarrow \infty$

where the field has Dirichlet boundary conditions, $\varphi = 2\pi\sqrt{2/v} n$ (with $n \in \mathbb{Z}$). It turns out that this is an integrable field theory. Paul Fendley, Andreas Ludwig and Hubert Saleur [322] used the thermodynamic Bethe Ansatz to calculate the differential tunneling conductance for the Laughlin state at $v = 1/3$ as a function of voltage V and temperature T and obtained the full weak to strong crossover RG flow. Detailed predictions of this theory have been verified experimentally by the work of Roddaro and coworkers [318, 319]. Point contact experiments were performed in the more challenging $v = 5/2$ FQH state by Miller and coworkers [323] and Radu and coworkers [324]. They found that the MR state is the one that best fits their experimental results.

The charge of particle can, in principle, be found directly by measuring the noise of a weak current. This process is known as shot noise. In the case of the constriction, the quasiparticle current operator is $I_{qp} = 2ev\Gamma_{qp} \sin(\sqrt{v/2}\phi + \omega_0^*t)$, where $\omega_0^* = evV/\hbar$. The noise spectrum, $S(\omega)$, of the tunneling current $_{qp}$ is [325]

$$S(\omega) = \int_{-\infty}^{\infty} dt \langle \{I_{qp}(t), I_{qp}(0)\} \rangle e^{i\omega t} \quad (529)$$

Using the expression of the quasiparticle correlator of Eq.(514), one readily finds that, to leading order in Γ_{qp} , the noise spectrum is

$$S(\omega) = ev\langle I_{qp} \rangle \left[\left(1 - \frac{\omega}{\omega_0^*}\right)^{2v-1} + \left(1 + \frac{\omega}{\omega_0^*}\right)^{2v-1} \right] \quad (530)$$

In the limit of zero frequency the noise spectrum takes the shot noise form

$$\lim_{\omega \rightarrow 0} S(\omega) = 2e^* \langle I_{qp} \rangle \quad (531)$$

where the expectation value of the tunneling current is

$$\langle I_{qp} \rangle = \frac{2\pi}{\Gamma(2v)} ev |\Gamma_{qp}|^2 \omega_0^{*2v-1} \quad (532)$$

Chamon, Freed and Wen [326] calculated the exact noise spectrum for the case of a $v = 1/2$ bosonic FQH state and were able to investigate the crossover as well, and Fendley, Ludwig and Saleur [327] used the Bethe Ansatz to construct a soliton basis to compute the DC shot noise.

These theoretical predictions were tested in tour-de-force experiments by de Picciotto and coworkers [253] and Samindayan and coworkers [254] who were able to measure the fractional charge of $e/3$ in the $v = 1/3$ state and, with some caveats, $e/5$ in the $v = 2/5$ state. Dolev and coworkers [328] went on to measure the charge from noise experiments in the nonabelian state at $v = 5/2$ and obtained results consistent with $e^* = e/4$ as predicted by theory.

We will now consider experiments on Hall bars with two quantum point contacts created by two narrow gates transversal to the bar and separated at some distance d each other. Quantum devices of this type are Fabry-Pérot quantum interferometers. The way they operate is as follows [329, 298]. A current is injected in the bottom edge. At the first quantum point contact (QPC) part if the current I_1 tunnels to the opposite edge and the other part I_2 goes on and tunnels at the second QPC. The FQH fluid is assumed to occupy uniformly the region between the two QPC's. There is some magnetic flux Φ in that region which also contains a number of localized vortices (quasiparticles). When both currents rejoin at the top edge they interfere and the interference has information on the charge of the particles that tunnels through the Aharonov-Bohm effect with the flux Φ . The interference also has information on the fractional statistics of the quasiparticles that tunneled through their braiding with the static vortices. Chamon, Freed, Kivelson, Sondhi and Wen proposed a setup of this type to measure the fractional statistics of the quasiparticles for the abelian states [329]. They showed that the total current $I_t = I_1 + I_2$ is given by

$$I_t = e^* |\Gamma_{\text{eff}}|^2 \frac{2\pi}{\Gamma(2v)} |\omega_0|^{2v-1} \text{sign}(\omega_0^*) \quad (533)$$

where Γ_{eff} is given by

$$|\Gamma_{\text{eff}}|^2 = |\Gamma_1|^2 + |\Gamma_2|^2 + (\Gamma_1 \Gamma_2^* + \Gamma_1^* \Gamma_2) F_v \left(\frac{\omega_0^* d}{v} \right) \quad (534)$$

Here Γ_1 and Γ_2 are the tunneling amplitudes at the two QPCs, $v = 1/m$ is the filling fraction, v is the velocity of the edge modes, d is the distance between the two QPCs and

$$F_v(x) = \sqrt{\pi} \frac{\Gamma(2v)}{\Gamma(v)} \frac{J_{v-1/2}(x)}{(2x)^{v-1/2}} \quad (535)$$

where $\Gamma(z)$ is the Euler Gamma function and $J_{\nu-1/2}(z)$ is the Bessel function of the first kind. In the presence of N_q localized quasiparticles in the area between the two QPCs, the contribution of the phases of the tunneling matrix elements get shifted to

$$\Gamma_1^* \Gamma_2 = \bar{\Gamma}_1^* \bar{\Gamma}_2 \exp \left[-2\pi i \left(\nu \frac{\Phi}{\phi_0} - \nu N_q \right) \right] \quad (536)$$

where ϕ_0 is the flux quantum. In Eq.(536) the first term in the phase shift is the Aharonov-Bohm effect of the tunneling quasiparticles, while the second term in the phase shift $2\pi\nu N_q$ is the contribution of the fractional statistics of the tunneling quasiparticle as its worldline braids with the N_q localized quasiparticles. This means that there is an interference contribution to the tunneling current (and also to the transmitted current) which is sensitive to both the charge and to the fractional statistics of the quasiparticles. This is this effect that is being measured in experiments.

Early attempts at doing this experiment were made by Camino and coworkers [330] but were difficult to interpret partly due to subtle reasons related to the difficulty in controlling the actual area of the region comprised between the two QPCs [331]. Technical advances in the fabrication of these devices led in 2020 to the first successful measurement of the fractional statistics of the quasiholes of the Laughlin state at $\nu = 1/3$ (and also at the Jain state at $\nu = 2/5$) by Nakamura, Liang, Gardner and Manfra [233].

The nonabelian case is more subtle both theoretically and experimentally. The theory of the abelian interferometer of Chamon and coworkers was generalized to the nonabelian FQH states by Fradkin, Nayak, Tsvelik and Wilczek in 1998 [298]. The structure of the interferometer is the same as in the abelian case but the interference effects are different. In addition, in the case of the MR state, aside from the nonabelian quasiparticle $\sigma \sim \chi \exp(i\phi/2\sqrt{2})$ has charge $e/4$ and nonabelian fractional statistics, the MR state has two more abelian anyons, the charge neutral Majorana fermion χ and the charge $e/2$ Laughlin quasiparticle. The number of anyons and their properties are very different in different nonabelian FQH states.

Ignoring for now the existence of more quasiparticles, let us focus now on the fundamental anyon which is nonabelian. Fradkin and coworkers showed [298] considered a nonabelian quasi-hole that is injected to the bottom edge, tunnels at the first QPC and arrives at the left end of the top edge in state $|\psi\rangle$. If a second such quasiparticle is now injected but now tunnels to the top edge at the second QPC., arriving at the left end of the top edge in the state $e^{i\alpha} B_{N_q} |\psi\rangle$, where α is the Aharonov-Bohm phase determined by the flux Φ piercing the interferometer and B_{N_q} si the braiding operator os the second quasiparticle that is circling around the N_q localized quasiparticles in the interferometer. Then the tunneling conductance measured at the left exit of the top edge has an interference contribution

$$\sigma_{xx} \propto |\Gamma_1|^2 + |\Gamma_2|^2 + \text{Re} \left[\Gamma_1^* \Gamma_2 e^{i\alpha} \langle \psi | B_{N_q} | \psi \rangle \right] \quad (537)$$

The matrix element $\langle \psi | B_{N_q} | \psi \rangle$ is given in terms of the expectation value of the Wilson loop operators of the tunneling quasiparticles braided with the Wilson loops of the N_q localized quasiparticles in the area of the interferometer. In 1989 Edward Witten [165] showed that this expectation value, which is computed in the nonabelian Chern-Simons gauge theory, is equal to a topological invariant of the braid known as the Jones polynomial $V_{N_q}(e^{i\pi/4})$. Therefore, the oscillatory component of the tunneling current (and of the conductance) measures a topological invariant!

Ref.[298] gave a general algorithm for the computation of this matrix element. However, in the particular case of the MR states, the non-abelian gauge theory associated with these states is the $SU(2)_2$ Chern-Simons gauge theory. In this case, an explicit calculation Bonderson, Kitaev and Shtengel [332] and by Stern and Halperin [333] leads to the result

$$\begin{aligned} \sigma_{xx} &\propto |\Gamma_1|^2 + |\Gamma_2|^2, & \text{for } N_q \text{ odd} \\ \sigma_{xx} &\propto |\Gamma_1|^2 + |\Gamma_2|^2 + 2 |\Gamma_1| |\Gamma_2| (-1)^{N_\psi} \cos \left(\alpha + \arg \left(\frac{\Gamma_2}{\Gamma_1} \right) + \frac{\pi}{4} N_q \right), & \text{for } N_q \text{ even} \end{aligned} \quad (538)$$

Here, $N_\psi = 1$ when the N_q quasiparticles fuse into the state ψ and $N_\psi = 0$ otherwise. The interference effect is absent for N_q odd since an odd number of σ quasiparticles cannot fuse into the identity state I . This simple even-odd effect is special for $SU(1)_2$. In the general case and, in particular in the $SU(2)_3$ case which applies to the $k = 3$ Read-Rezayi (“Fibonacci”) state, the expressions are more complex.

At any rate, even in the MR state the situation is more complicated for two reasons. One is that the charge $e/2$ abelian Laughlin quasiparticle is always able to tunnel thus spoiling the even-odd effect. the other complication is that the nonabelian FQH occurs in the $N = 1$ Landau level and the edge states of the nonabelian FQH state is surrounded by an abelian $\nu = 2$ state, which makes accessing the interesting edge states more difficult. Robert Willett has pioneered the fabrication and operation of the interferometer for the nonabelian state at $\nu = 5/2$ [334, 335, 336]. With some significant caveats although there is solid evidence of nonabelian braiding, more work remains to be done on this system.

11. Particle-Vortex Dualities in 2+1 dimensions

11.1. Electromagnetic Duality

The oldest form of duality in Physics is, perhaps, Dirac's observation that in the absence of electric charges and currents Maxwell's equations are invariant under the exchange of electric and magnetic fields, $\mathbf{E} \rightarrow \mathbf{B}$ and $\mathbf{B} \rightarrow -\mathbf{E}$ [114]. This observation led him to conjecture the existence of magnetic monopoles. In a relativistic invariant formulation, Maxwell's equations in free space can be expressed in terms of the electromagnetic field tensor $F_{\mu\nu}$ and the dual field tensor $F_{\mu\nu}^*$

$$\partial^\mu F_{\mu\nu} = 0, \quad \partial^\mu F_{\mu\nu}^* = 0 \quad (539)$$

where $F_{\mu\nu}^* = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$, and where $\epsilon_{\mu\nu\rho\sigma}$ is the fourth-rank antisymmetric Levi-Civita tensor. The first Maxwell equation in Eq.(539) is just the wave equation in free Minkowski spacetime. The second equation in Eq.(539) is known as the Bianchi identity. The Bianchi identity is a *constraint* which implies that there are no magnetic monopoles and that the second rank antisymmetric field strength tensor can be expressed in terms of the vector potential A_μ as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

This is an example of what in differential geometry is called *Hodge duality*, which relates a vector or, more generally a tensor field to its Hodge dual. In general in D dimensions the Hodge dual of a (fully antisymmetric) tensor of rank p is an antisymmetric tensor of rank $D - p$. When contracted with the oriented infinitesimal element of a p -dimensional hypersurface, $dx_1 \wedge dx_2 \wedge \dots \wedge dx_p$, an antisymmetric tensor of rank p , $F_{\mu_1\mu_2\dots\mu_p}$, defines a p -differential form, called a p -form. Thus, differential forms embody the physically intuitive notions of circulation of a vector field, flux of a second rank tensor, etc. We will see below that duality transformations are closely related to these geometric notions of duality.

11.2. Particle-Vortex Duality in 2+1 dimensions

In this section we will extend the particle-vortex duality discussed in 2D in section 5.1.1 (see Eq.(75)) to 3D. In the field theory interpretation, 2D is a 1+1-dimensional Minkowski spacetime and the XY model is a representation of a complex scalar field of unit modulus. In the duality, the particles are the particle-like excitations of the complex scalar field. The high temperature phase of the XY model is viewed as a partition function of a set of oriented loops that carry charge. the loops are the worldlines of the particles of the XY model.

11.2.1. The 3D XY Model

This picture can be seen as follows. Consider an XY model on a D-dimensional hypercubic lattice whose sites are labelled by $\{\mathbf{r}\}$. At each site there is a periodic variable $\theta(\mathbf{r}) \in [0, 2\pi]$. The partition function is

$$Z_{\text{XY}} = \prod_{\mathbf{r}} \int_0^{2\pi} \frac{d\theta(\mathbf{r})}{2\pi} \exp\left(\frac{1}{T} \sum_{\mathbf{r}} \sum_{j=1,\dots,D} (\cos \Delta_j \theta(\mathbf{r}))\right) \quad (540)$$

where $\Delta_j \theta(\mathbf{r}) = \theta(\mathbf{r} + \mathbf{e}_j) - \theta(\mathbf{r})$, with $j = 1, \dots, D$, is the lattice difference, and T is the temperature (in the classical statistical mechanical picture). Since the interaction on each link is a periodic function of the phase difference $\Delta_j \theta(\mathbf{r})$, we can expand the Gibbs weight (for each link!) in a Fourier series or, what is the same, in the integer-valued representations of the group U(1) (which is the global symmetry of the XY model) to obtain

$$Z_{\text{XY}} = \prod_{\mathbf{r}} \int_0^{2\pi} \frac{d\theta(\mathbf{r})}{2\pi} \sum_{\ell_j(\mathbf{r})=-\infty}^{\infty} \exp\left(-\sum_{\mathbf{r}, j} \frac{T}{2} \ell_j^2(\mathbf{r}) + i \sum_{\mathbf{r}} \theta(\mathbf{r}) \Delta_j \ell_j(\mathbf{r})\right) \quad (541)$$

where we used a Gaussian approximation for the modified Bessel function $I_n(z)$, and $\Delta_j \ell_j(\mathbf{r}) = \sum_{j=1}^D (\ell_j(\mathbf{r}) - \ell_j(\mathbf{r} - \mathbf{e}_j))$ (where \mathbf{e}_j is the lattice unit vector along the direction j) denotes the lattice divergence. Integrating-out the phase variables $\theta(\mathbf{r})$ we obtain

$$Z_{XY} = \prod_{\mathbf{r}} \sum_{\ell_j(\mathbf{r})=-\infty}^{\infty} \prod_{\mathbf{r}} \delta(\Delta_j \ell_j(\mathbf{r})) \exp \left(- \sum_{\mathbf{r}} \sum_{j=1}^D \frac{T}{2} \ell_j^2(\mathbf{r}) \right) \quad (542)$$

Therefore, the partition function is given by a sum over loops of conserved currents $\ell_j(\mathbf{r})$ defined on the links of the lattice, with a weight on each link $\propto \exp(-\ell_j^2/2\beta)$. These loops are the (lattice) worldlines of the particles of the complex scalar field. In the phase where this representation is convergent, the complex scalar field is massive, the XY model is gapped, and these particles have short range interactions. This picture is true in all dimensions, 3D included.

On the other hand, in 2D the vortices are point-like events in Euclidean spacetime which interact with each other through long range, logarithmic, interactions. In the field theory language, in 2D the point-like vortices are *instantons* which govern the low temperature phase of the XY model. In spacetime dimensions $D > 2$ the vortices become extended objects: vortex loops (or strings) in 3D, closed vortex surfaces in 4D, etc. In 3D the vortex loops are magnetic flux tubes which interact with each other through a Biot-Savart type interaction, namely bits of vortices interact with each other with a 3D Coulomb interaction much in the same way as with loops of current in magnetostatics,

$$Z_{3DXY} = \sum_{\{m_j(\tilde{\mathbf{r}})\}} \prod_{\tilde{\mathbf{r}}} \delta(\Delta_j m_j(\tilde{\mathbf{r}})) \exp \left(-\frac{2\pi^2}{T} \sum_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'} \sum_{j=1}^3 m_j(\tilde{\mathbf{r}}) G_0(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}') m_j(\tilde{\mathbf{r}}') - \alpha \sum_{\tilde{\mathbf{r}}, j} m_j^2(\tilde{\mathbf{r}}) \right) \quad (543)$$

where $\{\tilde{\mathbf{r}}\}$ labels the sites of the dual (cubic) lattice, the variables $m_j(\tilde{\mathbf{r}})$ take values on the integers, α is a (short-distance) vortex core energy, and $G_0(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}')$ is the 3D lattice propagator (Green function) which at long distances has the standard form

$$G_0(\mathbf{x} - \mathbf{x}') = \frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|} \quad (544)$$

In Eq.(543) the vortex loops are represented by the integer-valued conserved currents $m_j(\tilde{\mathbf{r}})$, which are naturally interpreted as the world lines of magnetic charges.

We could have also reached the same result by following the line of reasoning that we used in section 5.1.1. Indeed, let $\theta(x)$ be the phase field of a complex scalar field $\phi(x)$ deep in its broken symmetry state where the amplitude of the field $\phi(x)$ can be taken to be approximately fixed. The partition function in this phase reduces to

$$Z[a_\mu] = \int \mathcal{D}\theta \exp \left(-\frac{1}{2g} \int d^3x (\partial_\mu \theta - a_\mu)^2 \right) \quad (545)$$

where g is a coupling constant which in the XY model is proportional to the temperature. Here, much as in Eq.(70), the gauge field a_μ represents the vortices. Indeed, in 3D the vorticity is represented by a locally conserved vector field ω_μ

$$\omega_\mu = \epsilon_{\mu\nu\lambda} \partial_\nu a_\lambda = 2\pi \sum_k m_\mu^k \delta(x - x_k) \quad (546)$$

where m_μ^k are the integer-valued vortex (magnetic) currents we used above. As before, the periodic nature of the phase field $\theta(x)$ implies that the vortex currents must be quantized and be integer-valued.

Here too we can perform a Hubbard-Stratonovich transformation in terms of a vector field b_μ to find a dual theory which now is

$$Z(a_\mu) = \int \mathcal{D}b_\mu \mathcal{D}\theta \exp \left(-\frac{g}{2} \int d^3x b_\mu^2 - i \int d^3x b_\mu (\partial_\mu \theta - a_\mu) \right) \quad (547)$$

Upon integrating out the phase field θ , which acts as a Lagrange multiplier which imposes the constraint $\partial_\mu b_\mu = 0$. This constraint implies that we can write the field b_μ in terms of a dual gauge field ϑ_μ

$$b_\mu = \epsilon_{\mu\nu\lambda} \partial_\nu \vartheta_\lambda \quad (548)$$

Therefore, we can rewrite the partition function as

$$Z(a_\mu) = \int \mathcal{D}\vartheta \exp\left(-\frac{g}{4} \int d^3x f_{\mu\nu}^2 + i \int d^3x \omega_\mu \vartheta_\mu\right) \quad (549)$$

where $f_{\mu\nu} = \partial_\mu \vartheta_\nu - \partial_\nu \vartheta_\mu$ is the field strength of the dual gauge field ϑ_μ . Notice that the vortex current ω_μ is minimally coupled to the dual gauge field ϑ_μ . If we now integrate-out the gauge field ϑ_μ we obtain an expression for the weight in the path integral for a configuration of vortices m_μ^k which is identical to the result of Eq.(543).

What we have shown above is that in 3D the Goldstone phase of a complex scalar field with coupling constant g is the dual of a Maxwell gauge theory with coupling constant $1/g$. Moreover, the periodicity of the phase field implies that the dual gauge field ϑ_μ is compact in the sense that its fluxes must obey flux quantization. It is straightforward to show that a charge operator $V_n = \exp(in\theta)$ with electric charge n in the scalar field theory is represented in the dual gauge theory by a magnetic monopole of magnetic charge n .

In summary we showed that the 3D XY model (equivalent to the theory of the complex scalar field) can be written in terms of two different models of loop configurations. We saw that the XY model is a theory of particle (electric) loops in the high temperature phase and of vortex (magnetic) loops in the low temperature phase. However the particle loops have short range interactions while the vortex loops have long range interactions. Thus, the 3D XY model is not self-dual. In spite of having a representation in terms of vortex loops, the 3D XY model is in reality has a very different behavior than the 2D system. The Kosterlitz-Thouless theory describes the phase transition in terms of a vortex-antivortex unbinding transition upon which the vortices proliferate. In addition, in 2D the Goldstone phase is actually a line of fixed points, there is never true long range order but, instead, power-law correlations of the physical observables. In the Kosterlitz-Thouless theory the disordered phase arises the strong fluctuations of the phase field due to the proliferation of vortices while, at the local level, the order parameter still has a finite magnitude. A consequence of this behavior is that the superfluid density has a universal jump at the Kosterlitz-Thouless transition [337].

In contrast in the 3D XY model the Goldstone phase is a true phase with long range order. The theory has a continuous phase transition (the thermodynamic superfluid transition) at which the superfluid density vanishes continuously. This means that the phase transition of the 3D XY is better described by the Wilson-Fisher fixed point of a complex scalar field [30, 48], rather than by a vortex proliferation picture of the Kosterlitz-Thouless theory [84, 85]. In particular, since this theory does not have magnetic monopoles, the vortex loops cannot proliferate as they do in 2D. Instead, as the phase transition is approached, the vortex loops grow large in size but also become fractal-like objects whose effective core size diverges as the transition is approached. In fact, qualitative vortex proliferation arguments readily lead to the incorrect conclusion that the transition should be (strongly) first order [338].

In spite of these differences, particle-vortex duality still plays an important role in 3D by relating the 3D XY model to another theory which is its dual under electromagnetic duality. Here electromagnetic duality is understood as the exchange of the electric worldlines (electric loops) and the magnetic loops (vortex loops) while exchanging strong and weak coupling, $T \leftrightarrow 1/T$, in this case between different theories. This duality was investigated by Michael Peskin [339], by Paul Thomas and Michael Stone [338], and by Chandan Dasgupta and Bertrand Halperin [340].

11.2.2. Scalar QED in 3D

These authors considered a theory of a superconductor represented by a complex scalar field $\varphi(x)$, coupled to a fluctuating electromagnetic (Maxwell) field, also known as the abelian Higgs model, or scalar electrodynamics (scalar QED) a_μ , with $\mu = 1, 2, 3$. The partition function of this theory is

$$Z_{\text{SC}} = \int \mathcal{D}\varphi \mathcal{D}\varphi^* \mathcal{D}a_\mu \exp\left(-\int d^3x \mathcal{L}[\varphi, \varphi^*, a_\mu]\right) \quad (550)$$

where

$$\mathcal{L} = |D_\mu(a)\varphi|^2 + m^2|\varphi|^2 + u|\varphi|^4 + \frac{1}{4e^2}f_{\mu\nu}^2 \quad (551)$$

where the covariant derivative is $D_\mu(a) = \partial_\mu + iq a_\mu$, with the integer q being the charge of the scalar field (in units of the coupling constant e), and $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$ is the field strength. This theory is known as (Euclidean) scalar electrodynamics (or, the abelian Higgs model).

The scalar electrodynamics was extensively studied using the perturbative renormalization group within the epsilon expansion (near 4 dimensions) which leads to the conclusion that it has a weakly (fluctuation-induced) first order phase transition [109, 341]. Scalar QED of an N -component scalar field coupled to a Maxwell gauge field has a continuous phase transition with a non-trivial fixed point for $N \geq N_c \simeq 183$ (!) [342]. Since these results rely on perturbation theory (or in the large- N limit) it was long suspected that the physics may be different in $D = 3$. We will see that particle-vortex duality provides the answer to this question [340].

We begin by writing the lattice version of Eq.(550) which is obtained by coupling the XY model of Eq.(540) to a dynamical (Euclidean) Maxwell field

$$Z_{\text{SC}} = \prod_r \int_0^{2\pi} \frac{d\theta(\mathbf{r})}{2\pi} \int_{-\infty}^{\infty} \frac{da_\mu(\mathbf{r})}{2\pi} \exp(-S(\theta, a_\mu)) \quad (552)$$

where the action $S(\theta, a_\mu)$ is

$$S(\theta, a_\mu) = -\frac{1}{T} \sum_{r,\mu} \cos(\Delta_\mu \theta(\mathbf{r}) - qa_\mu(\mathbf{r})) + \frac{1}{4e^2} \sum_{r,\mu,\nu} (\Delta_\mu a_\nu(\mathbf{r}) - \Delta_\nu a_\mu(\mathbf{r}))^2 \quad (553)$$

In this action we assumed that the abelian gauge fields do not have monopole configurations.

The partition function of Eq.(550) admits a representation as a sum over loops. We can then used the same line of reasoning that led to Eq.(542) and write the partition function as a sum over loops of the worldlines of the particles (charges) of the complex scalar field

$$\frac{Z_{\text{SC}}}{Z_0^{\text{gauge}}} = \prod_r \sum_{\ell_\mu(\mathbf{r})=-\infty}^{\infty} \delta(\Delta_\mu \ell_\mu(\mathbf{r})) \exp\left(-\sum_{r,\mu} \frac{T}{2} \ell_\mu^2(\mathbf{r})\right) \langle \exp\left(iq \sum_{r,\mu} \ell_\mu(\mathbf{r}) a_\mu(\mathbf{r})\right) \rangle_a \quad (554)$$

where $\langle O[a] \rangle_a$ is the expectation value of the operator $O[a]$ over the gauge fields a_μ , and Z_0^{gauge} is the partition function of the free gauge fields. After computing this free-field expectation value we obtain the result

$$\frac{Z_{\text{SC}}}{Z_0^{\text{gauge}}} = \sum_{\{\ell_\mu(\mathbf{r})\}=-\infty}^{\infty} \delta(\Delta_\mu \ell_\mu(\mathbf{r})) \exp\left(-\sum_{r,\mu} \frac{T}{2} \ell_\mu^2(\mathbf{r}) - \frac{q^2 e^2}{2} \sum_{r,\mu} \sum_{r',\nu} \ell_\mu(\mathbf{r}) G_{\mu\nu}(\mathbf{r} - \mathbf{r}') \ell_\nu(\mathbf{r}')\right) \quad (555)$$

where

$$G_{\mu\nu}(\mathbf{r} - \mathbf{r}') = \langle a_\mu(\mathbf{r}) a_\nu(\mathbf{r}') \rangle \quad (556)$$

is the Euclidean propagator of the free gauge fields. Since the configurations of the worldlines, the currents represented by $\ell_\mu(\mathbf{r})$, are conserved and satisfy the local constraint $\Delta_\mu \ell_\mu = 0$, the second term in the exponent of Eq.(555) is gauge invariant. Hence, we can use the propagator in the Feynman gauge

$$G_{\mu\nu}(\mathbf{r} - \mathbf{r}') = \delta_{\mu\nu} G_0(\mathbf{r} - \mathbf{r}') \quad (557)$$

where $G_0(\mathbf{r} - \mathbf{r}')$ is the 3D lattice propagator which has the same Coulomb form at long distances already given in Eq.(544).

Therefore we can write the partition function of the scalar QED model (the superconductor) as a sum over worldline loop configurations of the particles of the scalar field in the equivalent form

$$\frac{Z_{\text{SC}}}{Z_0^{\text{gauge}}} = \sum_{\{\ell_\mu(\mathbf{r})\}=-\infty}^{\infty} \delta(\Delta_\mu \ell_\mu(\mathbf{r})) \exp\left(-\sum_{r,\mu} \frac{T}{2} \ell_\mu^2(\mathbf{r}) - \frac{q^2 e^2}{2} \sum_{r,r',\mu} \ell_\mu(\mathbf{r}) G_0(\mathbf{r} - \mathbf{r}') \ell_\mu(\mathbf{r}')\right) \quad (558)$$

which is the same as the partition function of the 3D XY model in the broken symmetry state, given in Eq.(543), with the identification of the particle loops of the abelian Higgs model with the vortices of the 3D XY model and a relation between the coupling constants. Hence we obtain the duality between the broken symmetry phase of the 3DXY model (“low T”) and the symmetric phase of the abelian Higgs model (“high T”)

$$\ell_\mu \leftrightarrow m_\mu, \quad q^2 e^2 \leftrightarrow \frac{2\pi^2}{T}, \quad \frac{T}{2} \leftrightarrow \alpha \quad (559)$$

where the quantities on the left hand side of the identifications refer to the 3D abelian Higgs model and the quantities on the right hand side to the 3D XY model.

We can now repeat the same analysis but in the broken symmetry of the abelian Higgs model. In this phase the gauge field a_μ becomes massive by the Higgs mechanism (or, what is the same, by the Meissner effect of the superconductor), and in this phase the vortex loops m_μ of the complex scalar field have short range interactions. This phase then is mapped onto the unbroken phase of the 3DXY model with the screened vortex loops of the abelian Higgs model identified with the particle loops of the 3D XY model, and with the same identifications between the coupling constants of the two theories given in Eq.(559). The identification between the two partition functions implies that the phase transitions must be the same. In other words, the duality implies that the 3D abelian Higgs model has a continuous phase transition with the same fixed point as that of the complex scalar field in 3D. The only difference is that the phases are reversed and, for this reason, this is sometimes called an “inverted” 3D XY transition [340].

11.2.3. The duality mapping

We can summarize these results with the following identification of two Lagrangians (in real time, Minkowski signature)[208]

$$|\partial_\mu(B)\phi|^2 - m^2|\phi|^2 - u|\phi|^4 \leftrightarrow |\partial_\mu(a)\varphi|^2 + m^2|\varphi|^2 - u|\varphi|^4 - \frac{1}{4e^2}f_{\mu\nu}^2 + \frac{1}{2\pi}\epsilon_{\mu\nu\lambda}a_\mu\partial_\nu B_\lambda \quad (560)$$

The left hand side of this equivalence is the Lagrangian of the 3D complex scalar field ϕ and B_μ is a background electromagnetic gauge field. The right hand side is the Lagrangian of the abelian Higgs model with a scalar field φ and a U(1) gauge field a_μ . Notice that the mass terms of the two sides are inverted reflecting the reverse order of their two phases. This identification also shows that the current of the scalar field $j_\mu = -\frac{i}{2}(\phi^*\partial_\mu\phi - \phi\partial_\mu\phi^*)$ is identified as

$$j_\mu \leftrightarrow \frac{1}{2\pi}\epsilon_{\mu\nu\lambda}\partial_\nu a_\lambda \quad (561)$$

The field operator of the complex scalar field creates worldlines (as opposed to loops) of charged particles. In the dual abelian Higgs theory this operator is identified with an operator that creates a magnetic monopole of the U(1) gauge field a_μ with unit magnetic charge. It is easy to see that in the broken symmetry phase of the abelian Higgs model a monopole-antimonopole pair creation operator decays exponentially with distance due to the flux expulsion effect (the Meissner effect of a superconductor). This means that the monopoles are confined with a linear potential. This is the same behavior of the complex scalar field ϕ in its symmetric phase where the propagator of the complex scalar field decays exponentially with distance. Conversely, in the broken symmetry phase of the complex scalar field theory, the field operator ϕ condenses and its propagator approaches the value $|\langle\phi\rangle|^2$ at long distances. This is the same behavior one readily finds for the monopole operator of the abelian Higgs model in the symmetric phase .

11.3. Bosonization in 2+1 dimensions

The particle-vortex duality that we discussed in section 11.2 essentially has been understood since the 1980s. In section 7.4 we discussed in detail the problem of bosonization in 1+1 dimensions which is a mapping between a massless Dirac fermion and a massless compactified scalar field. As we noted, this mapping is a powerful theoretical tool to investigate non-perturbatively the structure of many theories in 1+1 dimensions of great physical interest. This success has motivated a sustained effort to extend these concepts to higher dimensions where the problem in many ways is more subtle.

Fermi systems in dimensions higher differ from the 1+1-dimensional case in two significant ways. Due to the kinematic restriction of one space dimension a massless relativistic fermion and a theory of fermions (relativistic or not) at finite density are mostly equivalent to each other. The reason is quite simple. Consider a free massless Dirac theory in 1+1 dimensions. As we saw, it is equivalent to a theory of two chiral fermions, the right and left moving component of the spinor. If we add a chemical potential μ , the right and left moving states will be filled up to μ , which is the Fermi energy. While in space dimensions $d > 1$ a finite chemical potential leads to a finite fermi surface, in $d = 1$ space dimensions the Fermi “surface” is just two points, for the right and left moving fermion modes respectively. The Dirac Hamiltonian at finite chemical potential is

$$\mathcal{H} = \psi^\dagger(-i)\alpha\partial_x\psi - \mu\psi^\dagger\psi \quad (562)$$

Under a chiral transformation with angle $\theta(x)$, the field operators change as $\psi_R \rightarrow e^{i\theta}\psi_R$ and $\psi_L \rightarrow e^{-i\theta}\psi_L$ and the Hamiltonian becomes

$$\mathcal{H} = \psi^\dagger(-i)\alpha\partial_x\psi - (\mu - \partial_x\theta) \psi^\dagger\psi \quad (563)$$

Clearly if we choose $\partial_x\theta = \mu$ (or, what is the same, $\theta = \mu x$) the explicit dependence on the chemical potential has been cancelled and the transformed Dirac Hamiltonian is the same as the Hamiltonian at zero chemical potential. However under this transformation the mass operators transform as $\bar{\psi}\psi \rightarrow \cos(2\mu x)\bar{\psi}\psi$ and $i\bar{\psi}\gamma_5\psi \rightarrow \sin(2\mu x)i\bar{\psi}\gamma_5\psi$. In the non-relativistic context this change amounts to a modulation of the charge density with wave vector $Q = 2p_F(\mu)$.

In contrast, in space dimensions $d > 1$ the massless Dirac system and a system of fermions at finite density (relativistic or not) are no longer equivalent to each other as the latter system has a Fermi surface (of co-dimension $d - 1$) while the former system does not. This difference has profound effects on their dynamics: the system of fermions at finite density is, at least in the weak coupling regime, expected to be a Fermi liquid (except for an instability to a superconducting state even for infinitesimal attractive interactions) while the massless Dirac theory is stable as all local interactions are irrelevant.

11.3.1. Bosonization of the Dirac theory in 2+1 dimensions

We will now turn to the problem of Bose-Fermi mappings in relativistic systems in 2+1 dimensions. This problem is of great interest both in Condensed Matter Physics and in High Energy Physics.

Bosonization in 2+1 dimensions is a more subtle problem than in 1+1 dimensions that we discussed in section 7.4. There we saw, in 1+1 dimensions bosonization consists of a set of operator identities relating two free field theories, a theory of massless Dirac fermions and a compactified massless free scalar field $\phi(x)$. The success of that program is largely based on the fact that both theories are massless (and hence are scale and conformally invariant), on the chiral anomaly, and on the relation between the gauge current of the Dirac theory with the topological current of the compactified boson.

The physics in 2+1 dimensions (and in general) is very different. For instance, instead of the chiral anomaly in 2+1 dimensions theories of relativistic fermions have a parity anomaly, discussed in section 10.1.6. We will also see that although in 2+1 dimensions there are theories of free massless Dirac fermions and free massless compactified bosons they are no longer dual to each other. For these and other reasons it has been difficult to extend the bosonization program to 2+1 dimensions.

One of the first bosonization constructions for a theory of massive (relativistic) Dirac fields was put forth by Alexander Polyakov in 1988 [168, 167]. For reasons that no longer relevant, Polyakov considered a theory of CP^1 complex scalar fields (in its unbroken phase) coupled to a $U(1)_1$ Chern-Simons gauge theory. Nevertheless, his main argument, with some corrections associated with the parity anomaly, still holds correct.

Here we will follow the work of Hart Goldman and myself [343] and consider instead a theory of a single massive complex field coupled to a $U(1)_1$ Chern-Simons gauge theory whose Lagrangian density is

$$\mathcal{L} = |D_\mu(\mathcal{A}_\mu)\phi|^2 - m^2|\phi|^2 - \lambda|\phi|^4 + \frac{1}{4\pi}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu\partial^\nu\mathcal{A}^\lambda \quad (564)$$

In essence, this theory is a relativistic version of what we did in section 10.3.7 where we mapped non-relativistic fermions to (also non-relativistic) bosons whereas here the theory is relativistic and we are mapping bosons to fermions. In his work Polyakov actually considered the problem of the propagator of a free massive scalar field coupled to the Chern-Simons gauge field. In this limit, the propagator $G(x, x')$ is the transition amplitude from x to x' (in Minkowski spacetime). In the case in which $\phi(x)$ is a free massive field this propagator can be expressed in the form of a Feynman path-integral as a sum over all open oriented paths $\{P_{x,x'}\}$ with endpoints at x and x' which in Euclidean spacetime is

$$G_{\gamma(x,x')} = \sum_{\{P_{x,x'}\}} \exp(-mL(P_{x,x'})) \left\langle \exp\left(i \int_{P_{x,x'}} dz_\mu \mathcal{A}^\mu(z)\right) \right\rangle_{U(1)_1} \quad (565)$$

where the expectation value is over the Chern-Simons gauge fields \mathcal{A}_μ of Eq.(564).

Let us consider now the partition function of the bosons which is a sum over all closed paths $\{P\}$, which we will assume to be non-intersecting (which requires a short distance repulsion)

$$Z = \sum_{\{P\}} \exp(-mL(P)) \left\langle \exp\left(i \oint_P dz_\mu \mathcal{A}^\mu(z)\right) \right\rangle_{U(1)_1} \quad (566)$$

The expectation value is given by

$$\left\langle \exp\left(i \int_P dz_\mu \mathcal{A}^\mu(z)\right) \right\rangle_{U(1)_1} = \exp\left(-\frac{1}{2} \oint_P \oint_P dx_\mu dy_\nu \langle \mathcal{A}_\mu(x) \mathcal{A}_\nu(y) \rangle\right) \equiv \exp(i\pi W(P)) \quad (567)$$

where $W(P)$ is the *writhe* of the closed path P . The computation of the writhe requires a regularization. One possible regularization is to thicken the path which is equivalent to including a Maxwell term for the gauge field in the Lagrangian. In general, the writhe of the path P is

$$W(P) = SL(P) - T(P) \quad (568)$$

where $SL(P)$ is an integer-valued topological invariant called the self-linking number and $T(P)$ is the twist (or torsion) of the path. The twist $T(P)$ is a Berry phase which depends on the coordinates of the space in which the path P is embedded, and it is not a topological invariant.

We will now compute the Berry phase $T(P)$. Let $\hat{e}(s)$ be a unit vector tangent to the path P and where we parametrized the closed path P by a coordinate $s \in [0, L]$. The closed path P is the boundary of a surface Σ we can take to be a disk. We will write the Berry phase by extending $\hat{e}(s)$ smoothly to the interior of Σ as $\hat{e}(s, u)$, where $0 \leq u \leq 1$ and $\hat{e}(s, u=1) = \hat{e}(s)$ and $\hat{e}(s, u=0) = \hat{e}_0 \equiv \text{constant}$. With these definition the Berry phase is

$$T(P) \equiv W(\hat{e}) = \frac{1}{2\pi} \int_0^L ds \int_0^1 du \hat{e} \cdot \partial_s \hat{e} \times \partial_u \hat{e} \quad (569)$$

which is defined modulo an integer.

In this formulation, in the bosonic theory the amplitude for a path of length L with tangent vector $\hat{e}(s)$ with endpoints at x and x' is

$$G(x - x') = \int_0^\infty dL \int \mathcal{D}\hat{e}(s) \delta(1 - |\hat{e}|^2) \delta\left(x - x' - \int_0^L ds \hat{e}(s)\right) \exp(-|m|L \pm i\pi W(\hat{e})) \quad (570)$$

In momentum space we find

$$G(p) = \int_0^\infty dL \int \mathcal{D}\hat{e} \delta(1 - |\hat{e}|^2) \exp(-|m|L \pm i\pi W(\hat{e})) \exp(ip^\mu \int_0^L ds \hat{e}_\mu(s)) \quad (571)$$

By inspection of Eq.(569) we see that this is the same expression that we looked in the theory of the path integral for spin in section 4.3.2. for spin $S = 1/2$ particle in a magnetic field $b_\mu = \pm 2p_\mu$.

The equation of motion for \hat{e} is

$$\partial_s \hat{e}_\mu = \pm 2\epsilon_{\mu\nu\lambda} \hat{e}^\lambda = i[H, \hat{e}_\mu] \quad (572)$$

where $H = \mp p_\mu \hat{e}^\mu$ is the Hamiltonian. Upon quantization, \hat{e}_μ satisfies the commutation relations

$$[\hat{e}_\mu, \hat{e}_\nu] = 2i\epsilon_{\mu\nu\lambda} \hat{e}^\lambda \quad (573)$$

This means that we should make the identification $\hat{e}_\mu \rightarrow \sigma_\mu$ where σ_μ are the three 2×2 Pauli matrices. Up to rescaling of the mass $m \rightarrow M$, upon performing the path integral of Eq.(571) we find that $G(P)$ is the (Euclidean) Dirac propagator [168]

$$G(p) = \frac{1}{ip_\mu \sigma^\mu - M} \quad (574)$$

Using these results the partition function of Eq.(566) becomes

$$Z_{\text{fermion}} = \det[i\partial\!\!\!/ - M] = \int \mathcal{D}J \delta(\partial_\mu J^\mu) \exp(-|m|L[J] - i\text{sgn}(M)\pi\Phi[J]) \quad (575)$$

where $\Phi[J] = W[J]$, defined in Eq.(568).

We will return to this construction shortly below when we include the effects of the parity anomaly.

Early attempts at deriving a mapping for a theory of Dirac fermions in 2+1 dimensions were based on their behavior in the presence of gauge fields and the associated parity anomaly, discussed in section 10.1.6. These early theories are essentially a hydrodynamic description of the Dirac theory deep in the massive phase. With minor differences, the same results were derived independently (and simultaneously) by two groups, Fidel Schaposnik and myself [344] and Cliff Burgess and Francisco Quevedo [345]. This approach was reexamined more recently by A. Chan, T. Hughes, S. Ryu and myself in 2016 in a derivation of a hydrodynamic effective field theory for topological insulators in different dimensions [346, 347]. These derivations are similar in spirit to what we discussed in section 10.3.6 for the fractional quantum Hall effect. We will follow the approach of Ref.[346] for the special case of a Chern insulator 2+1 dimensions.

The Dirac theory, in any dimension, is globally gauge invariant. By Noether's theorem this means that it has a locally conserved current, $j_\mu = \bar{\psi} \gamma_\mu \psi$ which thus satisfies $\partial_\mu j^\mu = 0$. Here too, this means that we can write $j_\mu = \epsilon_{\mu\nu\lambda} \partial^\nu b^\lambda$, where b_μ is a gauge field. We expect that the effective action of the gauge field b_μ should be local, gauge invariant and, in this case, relativistic invariant. In 2+1 dimensions the effective action should break time reversal invariance and hence it should have a Chern-Simons term.

We will consider the free massive Dirac theory coupled to a background probe gauge field A_μ . The Lagrangian is $\mathcal{L} = \bar{\psi}(i(D) - M)\psi$, where $D_\mu = \partial_\mu + iA_\mu$. As we saw in section 10.1.6, gauge invariance and locality require that we have an even number of Dirac fermions. Here we will assume that one of the Dirac fermions is massive and acts as a regulator. The partition function

$$Z[A_\mu] = \int \mathcal{D}\bar{\psi} \mathcal{D}\psi \exp(iS_F[\bar{\psi}, \psi, A_\mu]) \quad (576)$$

By definition the expectation value of a product of current operators j_μ can be obtained by functional differentiation of the partition function with respect to A_μ .

The partition function is gauge invariant, i.e.

$$Z[A_\mu] = Z[A_\mu + a_\mu] \quad (577)$$

where the vector field a_μ is a pure gauge, $a_\mu = \partial_\mu \phi$ and $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu = 0$. Hence, a_μ is said to be flat. Therefore, up to a normalization

$$\begin{aligned} Z[A_\mu] &= \int \mathcal{D}[a_\mu]_{\text{pure}} Z[A_\mu + a_\mu] \\ &= \int \mathcal{D}a_\mu \delta(f_{\mu\nu} = 0) Z[A_\mu + a_\mu] \\ &= \int \mathcal{D}a_\mu \mathcal{D}b_\mu Z[A_\mu + a_\mu] \exp\left(-\frac{i}{2} \int d^3x \epsilon^{\mu\nu\lambda} b_\mu f_{\nu\lambda}\right) \end{aligned} \quad (578)$$

where in the last equality we introduced a representation of the delta function and the vector field b_μ plays the role of a Lagrange multiplier field. Using the invariance of the integration measure under $a_\mu \rightarrow a_\mu - A_\mu$ we find

$$Z[A_\mu] = \int \mathcal{D}a_\mu \mathcal{D}b_\mu Z[a_\mu] \exp\left(-\frac{i}{2} \int d^3x \epsilon^{\mu\nu\lambda} b_\mu (f_{\nu\lambda} - F^{\nu\lambda})\right) \quad (579)$$

where $F_{\mu\nu}$ is the field strength of the external field A_μ . From these identities and by differentiation of Eq.(579) it follows that a general expectation value of products of currents is

$$\langle j_\mu(x) j_\nu(y) \dots \rangle = \frac{\delta}{\delta A_\mu(x)} \frac{\delta}{\delta A_\nu(y)} \dots Z[A_\mu] = \langle \epsilon_{\mu\alpha\beta} \partial^\alpha b^\beta \epsilon_{\nu\gamma\delta} \partial^\gamma b^\delta \dots \rangle \quad (580)$$

which means that, at the operator level, we can identify [344]

$$j_\mu(x) \Leftrightarrow \epsilon_{\mu\nu\lambda} \partial^\nu b^\lambda \quad (581)$$

This result is actually general. The only difference is the nature of the field b in different dimensions: in 1+1 is a pseudo-scalar, in 2+1 is a vector (gauge) field, in 3+1 is an anti-symmetric (Kalb-Ramond) tensor field, etc.

Returning to the partition function of Eq.(579) we find, using the result of Eq.(324) applied to the partition function $Z[a_\mu]$, that $Z[A_\mu]$ is given by

$$Z[A_\mu] = \int \mathcal{D}a_\mu \mathcal{D}b_\mu \exp(i \int d^3x \mathcal{L}_{\text{eff}}[a_\mu, b_\mu, A_\mu]) \quad (582)$$

where the effective Lagrangian is

$$\mathcal{L}_{\text{eff}} = -\epsilon^{\mu\nu\lambda} b_\mu \partial_\nu a_\lambda + \frac{s}{4\pi} \epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda - \frac{1}{4g_{\text{eff}}} f_{\mu\nu} f^{\mu\nu} + A_\mu \epsilon^{\mu\nu\lambda} \partial_\nu b_\lambda + \dots \quad (583)$$

where $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$. Here $s = 1$ in the Chern insulator, $s = 0$ in the trivial insulator and $s = 1/2$ at the quantum critical point, and g_{eff} an effective coupling constant (with dimensions of length $^{-1}$). Here we included the effects of the parity anomaly. We recognize that the first term of the effective Lagrangian of Eq.(583) is a BF term.

In Ref.[346] a similar result is also derived for a 3+1-dimensional topological insulator. The main differences are that there is no parity anomaly but an axial anomaly and that the Lagrange multiplier field is a Kalb-Ramond field,

$$\mathcal{L}_{\text{eff}}[a_\mu, b_{\mu\nu}, A_\mu] = \epsilon^{\mu\nu\lambda\rho} b_{\mu\nu} \partial_\lambda a_\rho + \frac{\theta}{8\pi^2} \epsilon^{\mu\nu\lambda\rho} \partial_\mu a_\nu \partial_\lambda a_\rho - \frac{1}{4g^2} f_{\mu\nu} f^{\mu\nu} + A_\mu \epsilon^{\mu\nu\lambda\rho} \partial_\nu b_{\lambda\rho} + \dots \quad (584)$$

where $\theta = \pi$ in the topological insulator and $\theta = 0$ in the trivial insulator.

Although these results are correct, they do not give a full bosonization mapping. In particular, these results do not identify a fixed point for the dual bosonic theory and, along with it, a full mapping of the observables. Progress on this problem has only been achieved in the past few years both in the high energy literature [348, 349, 350, 351, 352, 199, 353, 354], and in the condensed matter physics literature [355, 356, 357, 358, 359]. Much of that new insight was presented in a 2016 insightful paper by Nathaniel Seiberg, T. Senthil, Chong Wang and Edward Witten [208]. Many of the dualities are actually conjectures supported by strong consistency checks. A derivation of the basic bosonization duality based on loop models was constructed by Hart Goldman and myself [343].

The basic conjectured bosonization duality is a mapping of a free Dirac fermion to a gauge complex scalar field with a Chern-Simons term [352, 208]. We can think of the Dirac theory as either being defined entirely in 2+1 dimensions or as being defined at the boundary of a non-trivial 3+1 dimensional topological insulator. In the first scenario we need to take into account the contribution of the fermionic doublers (this is what happens in the case of a lattice theory) or as the Pauli-Villars heavy fermionic regulators. In both cases, the additional heavy degrees of freedom cancel the anomaly of the 2+1-dimensional Dirac fermion, see section 10.1.6. In the second scenario, there is only one Dirac fermion at the boundary and the bulk θ -term cancels the anomaly of the boundary theory, see section 10.2.3. With these provisos, the Lagrangian \mathcal{L}_A of the free massive (or massless) Dirac fermion in 2+1 dimensions coupled to the electromagnetic gauge field A_μ is

$$\mathcal{L}_A = \bar{\psi}(i\cancel{D}(A_\mu) - M)\psi - \frac{1}{8\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu A^\lambda \quad (585)$$

where M is the Dirac mass. We will call this Theory A. Here $\cancel{D}_\mu(A) = \gamma^\mu(\partial_\mu - iA_\mu)$ and A_μ is a background (non-dynamical) gauge field. The last term in Eq.(585) is the Chern-Simons term with the 1/2-quantized coefficient, the usual short-hand for the η -invariant term of Eq.(342) needed to cancel the parity anomaly.

From the effective field theories we discussed above we know that the fermionic current maps to the curl of a gauge field, see Eq(581). Then, the conjectured bosonic dual is given by the Lagrangian of Eq(564) with an extra term for the (dual) coupling to the electromagnetic gauge field. We will call this Theory B whose Lagrangian \mathcal{L}_B is

$$\mathcal{L}_B = |D_\mu(\mathcal{A}_\mu)\phi|^2 - m^2|\phi|^2 - \lambda|\phi|^4 + \frac{1}{4\pi} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \partial^\nu \mathcal{A}^\lambda + \frac{1}{2\pi} \epsilon_{\mu\nu\lambda} A^\mu \partial^\nu \mathcal{A}^\lambda \quad (586)$$

To check the consistence we first observe that both theories are anomaly free. By functional differentiation of the two partition functions we check that we get the correct mapping for the fermionic current, $j_\mu \leftrightarrow \epsilon_{\mu\nu\lambda} \partial^\nu \mathcal{A}^\lambda$. This mapping implies that the charge density of Theory A maps onto the gauge field flux of Theory B, which is electromagnetic duality. We see that this bosonization is a relativistic version of flux attachment.

Let us now identify the mapping of the phases of both theories. If the Dirac mass $M < 0$, Theory A describes an anomalous quantum Hall insulator. Integrating out the massive fermions we find that the effective action for the electromagnetic gauge field A_μ is a $U(1)_1$ Chern-Simons theory. This means that in this phase $\sigma_{xy} = -1/(2\pi)$ (in units in which $e = \hbar = c = 1$). Looking now at Theory B we see that if $m^2 > 0$, the scalar field is in the unbroken phase, and it is massive. In this phase we set $\phi = 0$ and find that the low energy resulting theory is just a $U(1)_1$ Chern-Simons gauge theory coupled to the curl of the electromagnetic field. Upon integrating out the \mathcal{A}_μ gauge field we find that the effective action of A_μ is also a $U(1)_1$ Chern-Simons term, which implies that $\sigma_{xy} = -1/(2\pi)$.

Conversely, for $M > 0$ the effective electromagnetic action of the fermionic theory is a Maxwell term (the Chern-Simons term canceled). This phase is a trivial insulator. Looking now at Theory B we see that for $m^2 < 0$ this theory is in its Higgs phase where $\langle \phi \rangle \neq 0$ and the gauge field \mathcal{A}_μ now has a mass term, $\propto \mathcal{A}_\mu^2$. In the low energy limit $\mathcal{A}_\mu \rightarrow 0$, and the Hall conductivity vanishes, consistent to what is expected from Theory A.

On the other hand, for $M = 0$ Theory A is at a quantum critical point with a very simple CFT structure but a non-vanishing Hall conductivity $\sigma_{xy} = 1/(4\pi)$. It is natural to map this CFT to a Wilson-Fisher (WF) fixed point of Theory B. At the WF one sets the (renormalized!) mass $m_R^2 = 0$ (not the *bare* mass). In the absence of the Chern-Simons gauge field this WF fixed point is well understood from high quality (five loop!) epsilon expansion calculations (enhanced with Borel resummation) [49], and by more recent numerical Conformal Bootstrap methods [360]. However, not much is known of the *gauged* version of the CFT of Theory B since it cannot be accessed by either methods. If the mapping to Theory A is correct it should have a Hall conductivity of $\sigma_{xy} = 1/(4\pi)$. This conjectured value has not yet been confirmed.

Consider now a monopole operator of the gauge field \mathcal{A}_μ of Theory B. We will denote this operator $\mathcal{M}_{\mathcal{A}}$. The Chern-Simons term is not gauge invariant in a monopole background. To put it differently it has charge 1 under the Chern-Simons gauge field \mathcal{A}_μ . It also has charge 1 under the electromagnetic field A_μ . Likewise the scalar field ϕ has charge 1 under the Chern-Simons gauge field \mathcal{A}_μ . The operator $\phi^\dagger \mathcal{M}_{\mathcal{A}}$ is gauge-invariant under the Chern-Simons gauge field \mathcal{A}_μ (i.e. it has charge 0). This composite operator has electromagnetic charge 1 (which it inherited from the monopole). Moreover, it has spin-1/2 required by the Wu-Yang construction [151]. In other words, the operator $\phi^\dagger \mathcal{M}_{\mathcal{A}}$ has the *same* quantum numbers as the Dirac fermion and are identified by this conjecture. Notice, however, that the free massless Dirac field has scaling dimension $\Delta_\psi = 1$ in 2+1 dimensions. The conjecture implies the composite operator $\phi^\dagger \mathcal{M}_{\mathcal{A}}$ should also have scaling dimensions 1 at the (gauged) WF fixed point of Theory B. Many of these conjectures have been verified in non-abelian versions of Theory A and Theory B: a theory of free massless Dirac fermions coupled to a Chern-Simons gauge theory with gauge group $SU(N)_k$ and a theory of complex scalars coupled to a non-abelian Chern-Simons gauge theory with gauge group $SU(k)_N$, both in the limits $N \rightarrow \infty$ and $k \rightarrow \infty$ with N/k fixed [348, 349, 350, 351, 352].

The particle-vortex duality discussed in section 11.2 combined with the fermion-vortex duality we sketched here provide a web of dualities. These identifications constitute a powerful non-perturbative tool which has led to many significant developments. For instance, Goldman and myself [361] used the web of dualities to explain the experimentally observed self-duality at fractional quantum Hall plateau transitions, which was a long standing puzzle. It has also provided a powerful new tool to derive effective field theories of non-abelian fractional quantum Hall states [301] and even to propose novel states with a single (Fibonacci) anyon [303].

11.3.2. Bosonization of the Fermi Surface

A form of bosonization has been developed for the case of systems of fermions with a Fermi surface. Dense Fermi systems at sufficiently weak coupling are well described by the Landau theory of the Fermi liquid [13, 3, 194, 260, 259]. In this regime, the system of fermions has a collective mode, a bound state of particles and holes, which at long wavelengths is well described by the random phase approximation (RPA) [362]. However, in space dimensions $d > 1$ there is no kinematical restriction and quasi-particles and quasi-holes may move in their separate ways. The result is that, in addition to the collective modes, there is a low-energy spectrum of renormalized but essentially free quasi-particles. Because of the existence of this quasi-particle spectrum the collective modes generally (but not always) decay into particle-hole pairs resulting in a finite lifetime of the collective modes.

Superficially the particle-hole collective modes are similar to the scalar field of the bosonized theory in $d = 1$. To an extent it has been possible to “bosonize the Fermi surface” and to treat it as a quantum mechanical object [363, 364, 365, 366, 367, 368]. In this approach one essentially

regards each direction normal to the Fermi surface as a one-dimensional chiral fermion, including the current algebra structure, subject to global constraints that cancels the anomalies. This point of view ensures that the total fermion number in the Fermi sea is conserved. Here I will only present a short summary of the main ideas.

As in the one-dimensional case one defines a filled Fermi sea state $|FS\rangle$. This is a state of non-interacting fermions filling up all one-particle states up to the Fermi energy E_F . For concreteness we consider a system in two space dimensions. In this case, neglecting lattice effects, the Fermi surface is a circumference of radius p_F , the Fermi momentum. At fixed fermion number, the excitations are particle-hole pairs. The operator $n_k(\mathbf{q}) = c^\dagger(\mathbf{k} + \frac{\mathbf{q}}{2}) c(\mathbf{k} - \frac{\mathbf{q}}{2})$ creates a particle-hole pair with relative momentum \mathbf{q} with total momentum \mathbf{k} . In the low energy regime \mathbf{k} is a point on the Fermi surface (with $|\mathbf{k}| = p_F$) and \mathbf{q} is small momentum compared with p_F . In what follows we will label the point \mathbf{k} on the Fermi surface by the angle θ of the arc spanned by \mathbf{k} and an arbitrary origin on the Fermi surface.

We will normal-order the particle-hole creation operator with respect to the filled Fermi sea and define $\delta(\mathbf{q}, \theta) := n_k(\mathbf{q}) - \langle FS|n_k(\mathbf{q})|FS\rangle$. Haldane [363], Houghton and Marston [364, 367], and Castro Neto and Fradkin [365, 366] showed that in real space that these normal ordered density operators obey the equal-time commutation relations

$$[\delta n(\mathbf{x}, \theta), \delta n(\mathbf{x}', \theta')] = -\frac{1}{2\pi} \mathbf{k}_F(\theta) \cdot \nabla \delta^2(\mathbf{x} - \mathbf{x}') \delta(\theta - \theta') \quad (587)$$

where $\mathbf{k}_F(\theta)$ is a unit vector normal to the Fermi surface at angle θ . This is the algebra of the quantum fluctuations of the Fermi surface. We can further define at each point θ on the Fermi surface a Bose field $\varphi(\mathbf{x}, \theta)$ such that

$$\delta n(\mathbf{x}, \theta) = N(0) \mathbf{v}_F(\theta) \cdot \nabla \varphi(\mathbf{x}, \theta) \quad (588)$$

where $N(0)$ is the density of one-particle states at the Fermi surface and $\mathbf{v}_F(\theta)$ is the Fermi velocity at the location θ . The scalar field $\varphi(\mathbf{x}, \theta)$ is a chiral boson at θ which parametrizes the quantum fluctuations of the Fermi surface. The quantum dynamics of the chiral bosons $\varphi(\mathbf{x}, \theta)$ is governed by the action

$$S = \frac{1}{2} N(0) \int_0^{2\pi} \frac{d\theta}{2\pi} \int d^2x dt \left(-\partial_t \varphi(x, \theta) \mathbf{v}_F(\theta) \cdot \nabla \varphi(x, \theta) - (\mathbf{v}_F(\theta) \cdot \nabla \varphi(x, \theta))^2 \right) \\ + \frac{1}{2} N(0) \int_0^{2\pi} \frac{d\theta}{2\pi} \int_0^{2\pi} \frac{d\theta'}{2\pi} \int d^2x d^2x' dt F(\mathbf{x} - \mathbf{x}', \theta - \theta') \mathbf{v}_F(\theta) \cdot \nabla \varphi(x, \theta) \mathbf{v}_F(\theta') \cdot \nabla \varphi(x, \theta') \quad (589)$$

where $x = (\mathbf{x}, t)$, and $F(\mathbf{x} - \mathbf{x}', \theta - \theta')$ are the Fermi liquid parameters that parametrize the effective forward scattering interactions among the fermion quasiparticles on the Fermi surface [13]. The equation of motion predicted by this (quadratic) action is equivalent to the linearized quantum Boltzmann equation familiar from the Landau theory of the Fermi liquid. A non-linear extension has been introduced recently by Delacrétaiz and coworkers [368]. Recent work on the role of quantum anomalies in dense Fermi systems has yielded new insights on these problems [369].

In the regime, where the Landau theory of the Fermi liquid is expected to work [13], bosonization of the Fermi surface approach has reproduced the previously known results. There remain many open problems in this approach (and others) in the vicinity of quantum phase transitions. In spite of intense research using many different approaches, quantum phase transitions in metallic systems are not yet fully understood beyond perturbation theory [255, 256]. Non-perturbative ideas such as deconfined quantum criticality have been proposed [370] and significant work has been done using large- N methods [371, 372]. A particularly important metallic quantum phase transition (and perhaps the simplest) occurs near the Pomeranchuk instability of the fermi liquid. A quantum phase transition to an electron nematic state [373] has been predicted and studied within the Hertz-Millis approach. It has also been studied using higher dimensional bosonization [373, 374, 375]. Quantum Monte Carlo simulations [376] have shown that the vicinity of a nematic quantum critical point can trigger a superconducting state. Nematic Fermi fluids have been found in many physical systems of interest ranging from high temperature superconductors (such as cuprates and iron superconductors), to electron gases in large magnetic fields [377, 378, 379, 380].

12. Conclusions

In this chapter I have attempted to cover the role of Quantum Field Theory in modern Condensed Matter Physics. Its role and influence is vast and deep. In spite of the length of this chapter, I have not done justice to its role in many important areas of Condensed Matter. In particular, I have not touched on the theory of quantum spin liquids, fractons, and particularly topological phases in three space dimensions, among others. An area in which the ideas and concepts of QFT have a huge impact , both conceptually and as a tool, is on the problem of quantum entanglement. The concept of quantum entanglement as applied to condensed matter systems, both in equilibrium and out of equilibrium, is a major area of current development. It has become a crucial tool for characterizing systems at quantum criticality as well as topological phases of matter. Nevertheless, I hope that the reader will find this chapter both insightful and a useful resource.

Acknowledgments

This work was supported in part by the US National Science Foundation through grant No. DMR 1725401 at the University of Illinois.

References

- [1] J. R. Schrieffer, *Theory of Superconductivity*, Frontiers in Physics, Addison-Wesley, New York, New York, 1964.
- [2] Y. Nambu, G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I*, Physical Review 122 (1961) 345.
- [3] A. A. Abrikosov, L. P. Gorkov, I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.
- [4] A. L. Fetter, J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, New York, New York, 1971.
- [5] S. Doniach, E. H. Sondheimer, *Green's Functions for Solid State Physicists*, Frontiers in Physics, W. A. Benjamin Inc., New York, New York, 1974.
- [6] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, New York, 1965.
- [7] R. P. Feynman, *Statistical Mechanics, A Set of Lectures*, Frontiers in Physics, W. A. Benjamin Inc., Reading, Massachusetts, 1972.
- [8] L. D. Faddeev, *Introduction to Functional Methods*, in: *Methods of Field Theory, Proceedings of the Les Houches Summer School 1975, Session XXVIII*, R. Balian and J. Zinn-Justin editors, North Holland, Amsterdam, 1976, pp. 1–40.
- [9] E. Fradkin, *Field Theories of Condensed Matter Physics*, Second Edition, Cambridge University Press, Cambridge, U.K., 2013, (expanded 2nd edition of the originally published book in 1991).
- [10] E. Fradkin, *Quantum Field Theory: An Integrated Approach*, Princeton University Press, Princeton, N. J., 2021.
- [11] A. Altland, B. D. Simons, *Condensed Matter Field Theory*, 2nd Edition, Cambridge University Press, Cambridge U.K., 2010.
- [12] D. Pines, P. Nozières, *The Theory of Quantum Liquids*, Vol. I, W. A. Benjamin Inc., New York, New York, 1966.
- [13] G. Baym, C. J. Pethick, *Landau Fermi Liquid Theory*, John Wiley & Sons, New York, NY, 1991.
- [14] L. P. Kadanoff, G. Baym, *Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and Non-Equilibrium Problems*, Frontiers in Physics, W. A. Benjamin Inc., New York, New York, 1962.
- [15] A. Kamenev, *Field theory of Non-Equilibrium Systems*, 1st Edition, Cambridge University Press, Cambridge, U.K., 2011.
- [16] P. C. Martin, *Measurements and Correlation Functions*, Gordon and Breach Science Publishers Ltd., New York/London/Paris, 1968.
- [17] D. C. Mattis, *The Theory of Magnetism*, Harper & Row, New York, New York, 1965.
- [18] L. D. Landau, *On the Theory of Phase Transitions*, Zh. Eksp. Teor. Fiz. 7 (1937) 19.
- [19] L. D. Landau, E. M. Lifshitz, *Statistical Physics*, Vol. 5 of Course of Theoretical Physics, Pergamon Press, Oxford, U.K., 1959.
- [20] L. Onsager, *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*, Physical Review 65 (1944) 117.
- [21] T. D. Schultz, D. C. Mattis., E. H. Lieb, *Two-Dimensional Ising Model as a Soluble Problem of Many Fermions*, Rev. Mod. Phys. 36 (1964) 856.
- [22] E. Lieb, T. Schultz, D. C. Mattis, *Two soluble models of an antiferromagnetic chain*, Annals of Physics (N. Y.) 16 (1961) 407.
- [23] P. Pfeuty, *The one-dimensional Ising model with a transverse field*, Annals of Physics 57 (1970) 79.
- [24] C. N. Yang, *The Spontaneous Magnetization of a Two-Dimensional Ising Model*, Physical Review 85 (1952) 808.
- [25] L. P. Kadanoff, *Scaling Laws for Ising Models near T_c* , Physics 2 (1966) 263.
- [26] L. P. Kadanoff, *Spin-spin correlations in the two-dimensional Ising model*, Il Nuovo Cimento B 44 (1966) 276.
- [27] E. Efrati, Z. Wang, A. Kolan, L. P. Kadanoff, *Real-space renormalization in Statistical Mechanics*, Rev. Mod. Phys. 86 (2014) 647–667.
- [28] K. G. Wilson, *Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture*, Physical Review B 4 (1971) 3174–3183.
- [29] K. G. Wilson, *Renormalization Group and Critical Phenomena. II. Phase-Space Cell Analysis of Critical Behavior*, Physical Review B 4 (1971) 3184–3205.
- [30] K. G. Wilson, M. E. Fisher, *Critical Exponents in 3.99 Dimensions*, Phys. Rev. Lett. 28 (1972) 240.

- [31] K. G. Wilson, J. B. Kogut, The Renormalization Group and the ϵ Expansion, Physics Reports C 12 (1974) 75.
- [32] K. G. Wilson, The Renormalization Group: Critical phenomena and the Kondo problem, Rev. Mod. Phys. 47 (1975) 773.
- [33] K. G. Wilson, The Renormalization Group and Critical Phenomena, Rev. Mod. Phys. 55 (1983) 583.
- [34] A. Z. Patashinskii, V. L. Pokrovskii, Behavior of Ordered Systems Near the Transition Point, Soviet Physics JETP 23 (1966) 292, Zh. Eksp. i Teor. Fiz. **50**, 439 (1966).
- [35] M. E. Fisher, The Theory of Equilibrium Critical Phenomena, Reports on Progress in Physics 30 (1967) 615.
- [36] M. E. Fisher, The renormalization group in the theory of critical behavior, Reviews of Modern Physics 46 (1974) 597–615.
- [37] L. P. Kadanoff, W. Götzte, D. Hamblen, R. Hecht, E. A. S. Lewis, V. V. Palciauskas, M. Rayl, J. Swift, D. Aspnes, J. Kane, Static Phenomena Near Critical Points: Theory and Experiment, Rev. Mod. Phys. 39 (1967) 395–431.
- [38] M. Gell-Mann, F. E. Low, Quantum Electrodynamics at Small Distances, Physical Review 95 (1954) 1300.
- [39] N. N. Bogoliubov, D. V. Shirkov, Introduction to the Theory of Quantized Fields, Interscience, New York, New York, 1959.
- [40] L. P. Kadanoff, Operator Algebra and the Determination of Critical Indices, Phys. Rev. Lett. 23 (1969) 1430.
- [41] K. G. Wilson, Non-Lagrangian Models of Current Algebra, Phys. Rev. 179 (1969) 1499.
- [42] A. M. Polyakov, Conformal Symmetry of Critical Fluctuations, JETP Letters 12 (1970) 381, [ZhETF Pis. red. **12**, 538 (1970)].
- [43] A. M. Polyakov, Non-Hamiltonian approach to quantum field theory, Soviet Physics JETP 39 (1974) 10, [Zh. Eksp. Teor. Fiz. **66**, 23 (1974)].
- [44] L. P. Kadanoff, H. Ceva, Determination of an Operator Algebra for the Two-Dimensional Ising Model, Phys. Rev. B 3 (1971) 3918.
- [45] J. Cardy, Scaling and Renormalization in Statistical Physics, Cambridge Lecture Notes in Physics, Cambridge University Press, Cambridge, U.K., 1996.
- [46] G. 't Hooft, M. Veltman, Regularization and Renormalization of Gauge Fields, Nuclear Physics B 44 (1972) 189.
- [47] C. G. Bollini, J. J. Giambiagi, Dimensional Renormalization: The Number of Dimensions as a Regularizing Parameter, Il Nuovo Cimento B 12 (1972) 20.
- [48] D. J. Amit, Field Theory, the Renormalization Group and Critical Phenomena, McGraw Hill, New York, NY, 1980.
- [49] J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, 4th Edition, International Series of Monographs in Physics, Oxford University Press, Oxford, U.K., 2002.
- [50] D. J. Gross, F. Wilczek, Ultraviolet Behavior of Non-Abelian Gauge Theories, Phys. Rev. Lett. 30 (1973) 1343–1346.
- [51] P. W. Anderson, A poor man's derivation of the scaling laws of the Kondo problem, J. Phys. C: Solid State Phys. 3 (1970) 2436.
- [52] N. Andrei, Diagonalization of the Kondo Hamiltonian, Physical Review Letters 45 (1980) 379.
- [53] P. B. Wiegmann, Exact solution of the Kondo problem, JETP Letters 31 (1980) 392.
- [54] H. D. Politzer, Reliable Perturbative Results for Strong Interactions?, Phys. Rev. Lett. 30 (1973) 1346.
- [55] K. G. Wilson, Confinement of Quarks, Physical Review D 10 (1974) 2445.
- [56] J. Kogut, L. Susskind, Hamiltonian formulation of Wilson's lattice gauge theories, Phys. Rev. D 11 (1975) 395.
- [57] M. Creutz, L. Jacobs, C. Rebbi, Monte Carlo computations in lattice gauge theories, Physics Reports 95 (4) (1983) 201–282.
- [58] A. M. Polyakov, Interaction of goldstone particles in two dimensions. Applications to ferromagnets and massive Yang-Mills fields, Physics Letters B 59 (1975) 79.
- [59] S. H. Shenker, J. Tobochnik, Monte Carlo renormalization group analysis of the classical Heisenberg model in two dimensions, Phys. Rev. B 22 (1980) 4462.
- [60] E. Brézin, J. Zinn-Justin, Renormalization of the Nonlinear σ Model in $2 + \epsilon$ Dimensions: Application to the Heisenberg Ferromagnets, Phys. Rev. Lett. 36 (1976) 691.
- [61] D. H. Friedan, Nonlinear Models in $2 + \epsilon$ Dimensions, Annals of Physics 163 (1985) 318.
- [62] P. W. Anderson, Absence of Diffusion in Certain Random Lattices, Phys. Rev. 109 (1958) 1492.
- [63] E. Abrahams, P. W. Anderson, D. C. Licciardello, T. V. Ramakrishnan, Scaling Theory of Localization: Absence of Quantum Diffusion in Two Dimensions, Phys. Rev. Lett. 42 (1979) 673.
- [64] F. J. Wegner, The mobility edge problem: Continuous symmetry and a conjecture, Zeitschrift für Physik B Condensed Matter 35 (1979) 207.
- [65] A. J. McKane and M. Stone, Localization as an Alternative to Goldstone's Theorem, Annals of Physics 131 (1981) 36–55.
- [66] Shinobu Hikami, Anderson localization in a non-linear σ -model representation, Physical Review B 24 (1981) 2671–2679.
- [67] K. G. Wilson, Quantum Field Theory Models in Less Than 4 Dimensions, Phys. Rev. D 7 (1973) 2911.
- [68] S. L. Sondhi, S. M. Girvin, J. P. Carini, D. Shahar, Continuous quantum phase transitions, Rev. Mod. Phys. 69 (1997) 315.
- [69] S. Sachdev, Quantum Phase Transitions, 2nd Edition, Cambridge University Press, Cambridge, UK, 2011.
- [70] E. Fradkin, L. Susskind, Order and disorder in gauge systems and magnets, Phys. Rev. D 17 (1978) 2637.
- [71] J. B. Kogut, An introduction to lattice gauge theory and spin systems, Rev. Mod. Phys. 51 (1979) 659.
- [72] P. M. Chaikin, T. C. Lubensky, Principles of Condensed Matter Physics, Cambridge University Press, Cambridge, U.K., 1995.
- [73] E. Ardonne, P. Fendley, E. Fradkin, Topological order and conformal quantum critical points, Annals of Physics 310 (2004) 493.
- [74] A. Perelomov, Generalized Coherent States and Their Applications, Springer-Verlag, Berlin, 1986.
- [75] E. Fradkin, M. Stone, Topological terms in one- and two-dimensional quantum Heisenberg antiferromagnets, Phys. Rev. B 38 (1988) 7215(R).
- [76] P. B. Wiegmann, Multivalued functionals and geometrical approach for quantization of relativistic particles and strings, Nuclear Physics B 323 (1989) 311.
- [77] M. V. Berry, Quantal Phase Factors Accompanying Adiabatic Changes, Proceedings of the Royal Society of London A 392 (1984) 45.

- [78] B. Simon, Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase, *Phys. Rev. Lett.* 51 (1983) 2167.
- [79] D. Domb, N. Read, Absence of the Hopf invariant in the long-wavelength action of two-dimensional antiferromagnets, *Phys. Rev. B* 38 (1988) 7181.
- [80] F. D. M. Haldane, $O(3)$ non-linear σ model and the topological distinction between integer and half-integer-spin antiferromagnets in two dimensions, *Phys. Rev. Lett.* 61 (1988) 1029.
- [81] S. Chakravarty, B. I. Halperin, D. R. Nelson, Low-temperature behavior of two-dimensional quantum antiferromagnets, *Phys. Rev. Lett.* 60 (1988) 1057.
- [82] N. D. Mermin, The topological theory of defects in ordered media, *Rev. Mod. Phys.* 51 (1979) 591.
- [83] A. A. Abrikosov, On the magnetic properties of superconductors of the second group, *Soviet Physics JETP* 5 (1957) 1174, *Zh. Eksp. i Teor. Fiz.* 32, 1442 (1957).
- [84] J. M. Kosterlitz, D. J. Thouless, Order, metastability and phase transitions in two-dimensional systems, *J. Phys. C: Solid State Phys.* 6 (1973) 1181.
- [85] J. M. Kosterlitz, The critical properties of the two-dimensional XY model, *J. Phys. C: Solid State Phys.* 7 (1974) 1046.
- [86] V. L. Berezinskii, Destruction of Long-Range Order in One-Dimensional and Two-Dimensional Systems having a Continuous Symmetry Group I. Classical Systems, *Soviet Physics JETP* 32 (1971) 493.
- [87] V. L. Berezinskii, Destruction of Long-Range Order in One-Dimensional and Two-Dimensional Systems having a Continuous Symmetry Group II. Quantum Systems, *Soviet Physics JETP* 34 (1972) 610.
- [88] J. V. José, L. P. Kadanoff, S. Kirkpatrick, D. R. Nelson, Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model, *Phys. Rev. B* 16 (1977) 1217.
- [89] B. I. Halperin, D. R. Nelson, Theory of Two-Dimensional Melting, *Physical Review Letters* 41 (1978) 121.
- [90] A. P. Young, Melting and the vector Coulomb gas in two dimensions, *Physical Review B* 19 (1979) 1855.
- [91] J. Toner, D. R. Nelson, Smectic, cholesteric, and Rayleigh-Benard order in two dimensions, *Physical Review B* 23 (1981) 316.
- [92] D. R. Nelson, J. Toner, Bond-orientational order, dislocation loops, and melting of solids and smectic-A liquid crystals, *Physical Review B* 24 (1981) 363.
- [93] S. Coleman, *Aspects of Symmetry*, Cambridge University Press, Cambridge U.K., 1985.
- [94] R. Rajaraman, *Solitons and Instantons*, North-Holland, Amsterdam, The Netherlands, 1985.
- [95] H. B. Nielsen, P. Olesen, Vortex-line models for dual strings, *Nuclear Physics B* 61 (1973) 45.
- [96] G. 't Hooft, Symmetry Breaking through Bell-Jackiw Anomalies, *Physical Review Letters* 37 (1976) 8.
- [97] A. M. Polyakov, Quark confinement and topology of gauge theories, *Nucl. Phys. B* 120 (1977) 429.
- [98] A. M. Polyakov, *Gauge Fields and Strings*, Harwood Academic Publishers, London, 1987.
- [99] C. G. Callan, R. F. Dashen, D. J. Gross, The structure of the gauge theory vacuum, *Physics Letters B* 63 (1976) 334.
- [100] A. O. Caldeira, A. J. Leggett, Quantum Tunneling in a Dissipative System, *Annals of Physics* 149 (1983) 374–456.
- [101] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. Fisher, A. Garg, W. Zwerger, Dynamics of the dissipative two-state system, *Reviews of Modern Physics* 59 (1987) 1–85.
- [102] P. Di Francesco, P. Mathieu, D. Sénéchal, *Conformal Field Theory*, Springer-Verlag, Berlin, 1997.
- [103] P. Ginsparg, Applied Conformal Field Theory, in: E. Brézin, J. Zinn-Justin (Eds.), *Champs, Cordes et Phénomènes Critiques/ Fields, Strings and Critical Phenomena*, Proceedings of the Les Houches Summer School 1988, Session XLIX, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1989, pp. 1–168.
- [104] J. Polchinski, *String Theory*, Cambridge University Press, Cambridge U.K., 1998.
- [105] L. P. Kadanoff, Lattice Coulomb gas representations of two-dimensional problems, *Journal of Physics A: Mathematical and Theoretical* 11 (1978) 1399.
- [106] B. Nienhuis, Two-dimensional critical phenomena and the Coulomb gas, in: C. Domb, M. Green, J. L. Lebowitz (Eds.), *Phase Transitions and Critical Phenomena*, Vol. 11, Academic Press, London, U. K., 1987, p. 1.
- [107] N. D. Mermin, H. Wagner, Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models, *Phys. Rev. Lett.* 17 (1966) 1133–1136.
- [108] P. C. Hohenberg, Existence of Long-Range Order in One and Two Dimensions, *Phys. Rev.* 158 (1967) 383–386.
- [109] S. Coleman, There are no Goldstone bosons in two dimensions, *Communications in Mathematical Physics* 31 (1973) 259–264.
- [110] F. A. Schaposnik, Pseudoparticles and confinement in the two-dimensional Abelian Higgs model, *Physical Review D* 18 (1978) 1183.
- [111] G. 't Hooft, Computation of the quantum effects due to a four-dimensional pseudoparticle, *Phys. Rev. D* 14 (1976) 3432–3450.
- [112] H. Georgi, S. Glashow, Unity of all elementary-particle forces, *Physical Review Letters* 32 (1974) 438.
- [113] A. Polyakov, Compact gauge fields and the infrared catastrophe, *Physics Letters B* 59 (1975) 82 – 84.
- [114] P. Dirac, Quantised singularities in the electromagnetic field, *Proc. Roy. Soc. London* 133 (1931) 60.
- [115] F. D. M. Haldane, 'θ' physics' and quantum spin chains, *J. Appl. Phys.* 57 (1985) 3359.
- [116] H. Bethe, Theory of metals. I. Eigenvalues and eigenfunctions of the linear atomic chain, *Z. Physik* 71 (1931) 205.
- [117] C. N. Yang, C. P. Yang, Thermodynamics of a One-Dimensional System of Bosons with Repulsive Delta-Function Interaction, *J. Math. Phys.* 10 (1969) 1115.
- [118] I. Affleck, F. D. M. Haldane, Critical theory of quantum spin chains, *Phys. Rev. B* 36 (1987) 5291.
- [119] E. Witten, Non-Abelian Bosonization in Two Dimensions, *Comm. Math. Phys.* 92 (1984) 455.
- [120] V. G. Knizhnik, A. B. Zamolodchikov, Current Algebra and Wess-Zumino Model in Two Dimensions, *Nuclear Physics B* 247 (1984) 83.
- [121] I. Affleck, T. Kennedy, E. H. Lieb, H. Tasaki, Valence Bond Ground States in Isotropic Quantum Antiferromagnets, *Comm. Math. Phys.* 115 (1988) 477.
- [122] M. Hagiwara, K. Katsumata, I. Affleck, B. I. Halperin, J. P. Renard, Observation of $S = 1/2$ degrees of Freedom in an $S = 1$ Linear-Chain Heisenberg Antiferromagnet, *Physical Review Letters* 65 (1990) 3181.
- [123] H. A. Kramers, G. H. Wannier, Statistics of the Two-Dimensional Ferromagnet. Part I, *Phys. Rev.* 60 (1941) 252.
- [124] F. J. Wegner, Duality in Generalized Ising Models and Phase Transitions without Local Order Parameters, *J. Math. Phys.* 12 (1971) 2259.
- [125] S. Elitzur, R. B. Pearson, J. Shigemitsu, Phase structure of discrete Abelian spin and gauge systems, *Physical*

- Review D 19 (1979) 3698.
- [126] A. Ukawa, P. Windey, A. H. Guth, Dual variables for lattice gauge theories and the phase structure of $Z(N)$ systems, Physical Review D 21 (1980) 1013.
- [127] F. Haldane, Luttinger liquid theory of one-dimensional quantum fluids. I. Properties of the Luttinger model and their extension to the general 1D interacting spinless Fermi gas, J. Phys. C: Solid State Phys. 14 (1981) 2585.
- [128] D. C. Mattis, E. H. Lieb, Exact Solution of a Many-Fermion System and its Associated Boson Field, Journal of Mathematical Physics 6 (1965) 304.
- [129] A. Luther, V. J. Emery, Backward Scattering in the One-Dimensional Electron Gas, Physical Review Letters 33 (1974) 589.
- [130] S. Mandelstam, Soliton operators for the quantized sine-Gordon equation, Physical Review D 11 (1975) 3026.
- [131] S. Coleman, Quantum sine-Gordon equation as the massive Thirring model, Phys. Rev. D 11 (1975) 2088.
- [132] D. J. Gross, A. Neveu, Dynamical symmetry breaking in asymptotically free field theories, Phys. Rev. D 10 (1974) 3235.
- [133] S. E. Adler, The Axial-Vector Vertex in Spinor Electrodynamics, Physical Review 177 (1969) 2426.
- [134] J. S. Bell, R. Jackiw, The PCAC Puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ -model, Nuovo Cimento A 60 (1969) 47.
- [135] X. L. Qi, T. L. Hughes, S. C. Zhang, Topological field theory of time-reversal invariant insulators, Phys. Rev. B 78 (2008) 195424.
- [136] J. Schwinger, Field Theory Commutators, Phys. Rev. Lett. 3 (1959) 296.
- [137] E. Witten, Some Properties of the $(\bar{\psi}\psi)^2$ Model in Two Dimensions, Nucl. Phys. B 142 (1978) 285.
- [138] W. P. Su, J. R. Schrieffer, A. J. Heeger, Solitons in Polyacetylene, Phys. Rev. Lett. 42 (1979) 1698.
- [139] R. Jackiw, J. R. Schrieffer, Solitons with fermion number 1/2 in condensed matter and relativistic field theories, Nuclear Physics B 190 (1981) 253.
- [140] R. Jackiw, C. Rebbi, Solitons with fermion number 1/2, Physical Review D 13 (1976) 3398.
- [141] J. Goldstone, F. Wilczek, Fractional Quantum Numbers on Solitons, Phys. Rev. Lett. 47 (1981) 986.
- [142] H. Takayama, Y. R. Lin-Liu, K. Maki, Continuum model for solitons in polyacetylene, Phys. Rev. B 21 (1980) 2388–2393.
- [143] D. K. Campbell, A. R. Bishop, Solitons in polyacetylene and relativistic-field-theory models, Phys. Rev. B 24 (1981) 4859–4862.
- [144] D. Campbell, A. Bishop, Soliton excitations in polyacetylene and relativistic field theory models, Nuclear Physics B 200 (1982) 297–328.
- [145] J. E. Hirsch, E. Fradkin, Effect of quantum fluctuations on the Peierls instability: A Monte Carlo study, Phys. Rev. Lett. 49 (1982) 402–405.
- [146] E. Fradkin, J. E. Hirsch, Phase diagram of one-dimensional electron-phonon systems. I. The Su-Schrieffer-Heeger model, Phys. Rev. B 27 (1983) 1680.
- [147] M. F. Atiyah, V. K. Patodi, I. M. Singer, Spectral Asymmetry and Riemannian Geometry. I, Math. Proc. Cambridge Philos. Soc. 77 (1975) 43.
- [148] P. Dirac, The Principles of Quantum Mechanics, Oxford University Press, Oxford, U.K., 1930.
- [149] L. D. Landau, E. M. Lifshitz, Quantum Mechanics: Non-Relativistic Theory, 1st Edition, Vol. Volume 3 of Course of Theoretical Physics, Pergamon Press, Oxford, U.K., 1957.
- [150] S. Weinberg, The Quantum Theory Of Fields, 1st Edition, Vol. I-III, Cambridge University Press, Cambridge, U.K., 2005.
- [151] T. T. Wu, C. N. Yang, Dirac monopole without strings: Monopole harmonics, Nuclear Physics B 107 (1976) 365.
- [152] E. Fradkin, L. P. Kadanoff, Disorder variables and para-fermions in two-dimensional statistical mechanics, Nuclear Physics B 170 (1980) 1.
- [153] J. M. Leinaas, J. Myrheim, On the theory of identical particles, II Nuovo Cimento B (1971-1996) 37 (1977) 1–23.
- [154] M. G. G. Laidlaw, C. M. DeWitt, Feynman Functional Integrals for Systems of Indistinguishable Particles, Physical Review D 3 (1971) 1375–1378.
- [155] Y. Aharonov, D. Bohm, Significance of Electromagnetic Potentials in the Quantum Theory, Phys. Rev. 115 (1959) 485–491.
- [156] F. Wilczek, Magnetic Flux, Angular Momentum, and Statistics, Phys. Rev. Lett. 48 (1982) 1144.
- [157] F. Wilczek, Quantum Mechanics of Fractional-Spin Particles, Phys. Rev. Lett. 49 (1982) 957–959.
- [158] A. S. Goldhaber, Connection of Spin and Statistics for Charge-Monopole Composites, Phys. Rev. Lett. 36 (1976) 1122–1125.
- [159] D. Finkelstein, J. Rubinstein, Connection between Spin, Statistics, and Kinks, Journal of Mathematical Physics 9 (1968) 1762–1779.
- [160] T. Skyrme, A Unified Theory of Mesons and Baryons, Nuclear Physics 31 (1962) 556–569.
- [161] Y.-S. Wu, General Theory for Quantum Statistics in Two Dimensions, Phys. Rev. Lett. 52 (1984) 2103–2106.
- [162] F. Wilczek, A. Zee, Linking numbers, spin, and statistics of solitons, Phys. Rev. Lett. 51 (1983) 2250.
- [163] S. Deser, R. Jackiw, S. Templeton, Three-Dimensional Massive Gauge Theories, Physical Review Letters 48 (1982) 975.
- [164] S. Deser, R. Jackiw, S. Templeton, Topologically Massive Gauge Theories, Annals of Physics 140 (1982) 372–411.
- [165] E. Witten, Quantum Field Theory and the Jones Polynomial, Commun. Math. Phys. 121 (1989) 351.
- [166] G. T. Horowitz, Exactly Soluble Diffeomorphism Invariant Theories, Communications in Mathematical Physics 125 (1989) 417.
- [167] J. Grundberg, T. H. Hansson, A. Karlhede, On Polyakov's Spin Factors, Nuclear Physics B 347 (1990) 420–440.
- [168] A. M. Polyakov, Fermi-Bose Transmutations Induced by Gauge Fields, Mod. Phys. Lett. A 3 (1988) 325.
- [169] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, Annals of Physics 303 (2003) 2, ([arXiv:quant-ph/9707021](https://arxiv.org/abs/quant-ph/9707021)).
- [170] S. Das Sarma, M. Freedman, C. Nayak, S. H. Simon, A. Stern, Non-abelian anyons and topological quantum computation, Rev. Mod. Phys. 80 (2008) 1083.
- [171] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, The electronic properties of graphene, Rev. Mod. Phys. 81 (2009) 109–162.
- [172] X. L. Qi, S. C. Zhang, Topological insulators and superconductors, Rev. Mod. Phys. 83 (2011) 1057.
- [173] L. Susskind, Lattice Fermions, Physical Review D 16 (1977) 3031.
- [174] H. B. Nielsen, M. Ninomiya, Absence of neutrinos on a lattice (I). Proof by homotopy theory, Nuclear Physics B

- 185 (1981) 20–40.
- [175] H. B. Nielsen, M. Ninomiya, Absence of neutrinos on a lattice (II). Intuitive topological proof, Nuclear Physics B 193 (1981) 173–194.
- [176] H. B. Nielsen, M. Ninomiya, The Adler-Bell-Jackiw Anomaly and Weyl fermions in a Crystal, Physics Letters B 130 (1983) 389.
- [177] D. Friedan, A Proof of the Nielsen-Ninomiya Theorem, Communications in Mathematical Physics 85 (1982) 481–490.
- [178] G. W. Semenoff, Condensed-Matter Simulation of a Three-Dimensional Anomaly, Physical Review Letters 53 (1984) 2449.
- [179] D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, Quantized Hall Conductance in a Two-Dimensional Periodic Potential, Physical Review Letters 49 (1982) 405.
- [180] I. Affleck, J. B. Marston, Large- N limit of the Heisenberg-Hubbard Model: Implications for High- T_c Superconductors, Phys. Rev. B 37 (1988) 3774.
- [181] X. G. Wen, F. Wilczek, A. Zee, Chiral spin states and superconductivity, Phys. Rev. B 39 (1989) 11413–11423.
- [182] F. Bloch, Über der quantenmechanik der elektronen in kristallgittern, Zeitschrift für physik 52 (1929) 555–600.
- [183] C. Kittel, Introduction to Solid State Physics, 1st Edition, John Wiley & Sons, New York, New York, 1996.
- [184] N. Ashcroft, D. N. Mermin, Solid State Physics, Holt, Reinhart and Winston, New York, New York, 1976.
- [185] B. Bradlyn, L. Elcoro, J. Cano, M. G. Vergniory, Z. Wang, C. Felser, M. I. Aroyo, B. A. Bernevig, Topological quantum chemistry, Nature 547 (2017) 298–305.
- [186] J. Cano, B. Bradlyn, Band Representations and Topological Quantum Chemistry, Annual Review of Condensed Matter Physics 12 (2021) 225–246.
- [187] M. Kohmoto, Topological invariant and the quantization of the Hall conductance, Annals of Physics 160 (1985) 343.
- [188] A. P. Schnyder, S. Ryu, A. Furusaki, A. W. W. Ludwig, Classification of topological insulators and superconductors in three spatial dimensions, Phys. Rev. B 78 (2008) 195125.
- [189] A. Kitaev, Periodic table for topological insulators and superconductors, in: M. Feigelman (Ed.), Advances in Theoretical Physics: Landau Memorial Conference, Vol. 1134, American Institute of Physics, AIP Conference Proceedings, College Park, Maryland, 2009, p. 22.
- [190] J. E. Moore, L. Balents, Topological invariants of time-reversal-invariant band structures, Phys. Rev. B 75 (2007) 121306.
- [191] E. Fradkin, M. Kohmoto, Quantized Hall effect and geometric localization of electrons on lattices, Phys. Rev. B 35 (1987) 6017–6023.
- [192] D. R. Hofstadter, Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields, Physical Review B 14 (1876) 2239.
- [193] P. C. Martin, Measurements and Correlation Functions, Gordon & Breach, Philadelphia, Pennsylvania, 1967.
- [194] P. Nozières, D. Pines, The Theory of Quantum Liquids: Normal Fermi Liquids, W. A. Benjamin Inc., New York, New York, 1966.
- [195] S. Ryu, C. Mudry, C.-Y. Hou, C. Chamon, Masses in graphene-like two-dimensional electronic systems: Topological defects in order parameters and their fractional exchange statistics, Physical Review B 80 (2009) 205319.
- [196] F. D. M. Haldane, Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the “Parity Anomaly”, Phys. Rev. Lett. 61 (1988) 2015–2018.
- [197] X. Chen, Z.-C. Gu, Z.-X. Liu, X.-G. Wen, Symmetry-Preserved Topological Orders in Interacting Bosonic Systems, Science 338 (2012) 1604–1606.
- [198] T. Senthil, Symmetry-protected topological phases of quantum matter, Annual Review of Condensed Matter Physics 6 (2015) 299–324.
- [199] E. Witten, Fermion path integrals and topological phases, Reviews of Modern Physics 88 (2016) 035001.
- [200] A. N. Redlich, Gauge Noninvariance and Parity Nonconservation of Three-Dimensional Fermions, Physical Review Letters 52 (1984) 18.
- [201] A. N. Redlich, Parity violation and gauge noninvariance of the effective gauge field action in three dimensions, Phys. Rev. D 29 (1984) 2366.
- [202] K. Fujikawa, Path-Integral Measure for gauge-Invariant Fermion Theories, Physical Review Letters 42 (1979) 1195.
- [203] R. E. Gamboa Saraví, M. A. Muschietti, F. A. Schaposnik, J. E. Solomin, Chiral Symmetry and Functional Integral, Annals of Physics 157 (1984) 360.
- [204] G. E. Volovik, Zeros in the fermion spectrum in superfluid systems as diabolical points, Journal of Experimental and Theoretical Physics Letters 46 (1987) 98.
- [205] V. M. Yakovenko, Chern-Simons Terms and n Field in Haldane’s Model for the Quantum Hall Effect without Landau Levels, Physical Review Letters 65 (1990) 251.
- [206] M. F. Golterman, K. Jansen, D. B. Kaplan, Chern-Simons currents and chiral fermions on the lattice, Physics Letters B 301 (1993) 219.
- [207] X.-L. Qi, Y.-S. Wu, S.-C. Zhang, Topological quantization of the spin Hall effect in two-dimensional paramagnetic semiconductors, Physical Review B 74 (2006) 085308.
- [208] N. Seiberg, T. Senthil, C. Wang, E. Witten, A Duality Web in 2+1 Dimensions and Condensed Matter Physics, Annals of Physics 374 (2016) 395 – 433.
- [209] L. Fu, C. L. Kane, Topological insulators with inversion symmetry, Phys. Rev. B 76 (2007) 045302.
- [210] C. L. Kane, E. J. Mele, Z_2 Topological Order and the Quantum Spin Hall Effect, Phys. Rev. Lett. 95 (2005) 146802.
- [211] B. A. Bernevig, T. L. Hughes, S. C. Zhang, Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells, Science 314 (2006) 1757.
- [212] J. Gooth, B. Bradlyn, S. Honnali, C. Schindler, N. Kumar, J. Noky, Y. Qi, C. Shekhar, Y. Sun, Z. Wang, B. A. Bernevig, C. Felser, Axionic charge-density wave in the Weyl semimetal (TaSe₄)₂I, Nature 575 (2019) 315–319.
- [213] A. M. Essin, J. E. Moore, D. Vanderbilt, Magnetic Polarizability and Axion Electrodynamics in Crystalline Insulators, Physical Review Letters 102 (2009) 146805.
- [214] P. Hosur, S. Ryu, A. Vishwanath, Chiral topological insulators, superconductors, and other competing orders in three dimensions, Physical Review B 81 (2010) 045120.

- [215] F. Wilczek, Two applications of axion electrodynamics, *Phys. Rev. Lett.* 58 (1987) 1799–1802.
- [216] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, The Advanced Book Program, Perseus Books Publishing L.L.C., Reading, Massachusetts, 1995.
- [217] E. Witten, Dyons of charge $e\theta/2\pi$, *Physics Letters B* 86 (1979) 283.
- [218] C. G. Callan, J. A. Harvey, Anomalies and Fermion Zero Modes on Strings and Domain Walls, *Nucl. Phys. B* 250 (1985) 427.
- [219] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, M. Z. Hasan, A tunable topological insulator in the spin helical Dirac transport regime, *Nature* 460 (2009) 1101–1105.
- [220] K. von Klitzing, G. Dorda, M. Pepper, New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance, *Physical Review Letters* 45 (1980) 494.
- [221] D. C. Tsui, H. L. Stormer, A. C. Gossard, Two-Dimensional Magnetotransport in the Extreme Quantum Limit, *Physical Review Letters* 48 (1982) 1559.
- [222] X. Du, I. Skachko, F. Duerr, A. Luican, E. Y. Andrei, Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene, *Nature* 462 (2009) 192–195.
- [223] A. A. Zibrov, C. Kometter, H. Zhou, E. M. Spanton, T. Taniguchi, K. Watanabe, M. P. Zaletel, A. F. Young, Tunable interacting composite fermion phases in a half-filled bilayer-graphene Landau level, *Nature* 549 (2017) 360–364.
- [224] L. Landau, Diamagnetismus der Metalle, *Zeitschrift für Physik* 64 (1930) 629–637.
- [225] Q. Niu, D. J. Thouless, Y.-S. Wu, Quantized Hall conductance as a topological invariant, *Physical Review B* 31 (1985) 3372.
- [226] R. B. Laughlin, Quantized Hall conductivity in two dimensions, *Physical Review B* 23 (1981) 5632–5633.
- [227] H. Levine, S. B. Libby, A. Pruisken, Electron Delocalization by a Magnetic Field in Two Dimensions, *Physical Review Letters* 51 (1983) 1915.
- [228] A. Pruisken, On localization in the theory of the quantized hall effect: A two-dimensional realization of the θ -vacuum, *Nuclear Physics B* 235 (1984) 277.
- [229] S. Kivelson, M. Roček, Consequences of Gauge Invariance for Fractionally Charged Quasi-Particles, *Physics Letters B* 85 (156).
- [230] J. K. Jain, Incompressible quantum Hall states, *Physical Review B* 40 (1989) 8079.
- [231] J. K. Jain, Composite-fermion approach for the fractional quantum Hall effect, *Physical Review Letters* 63 (1989) 199.
- [232] B. I. Halperin, P. A. Lee, N. Read, Theory of the half-filled Landau level, *Phys. Rev. B* 47 (1993) 7312.
- [233] J. Nakamura, S. Liang, G. C. Gardner, M. J. Manfra, Direct observation of anyonic braiding statistics, *Nature Physics* 16 (2020) 931–936.
- [234] B. I. Halperin, Statistics of Quasiparticles and the Hierarchy of Fractional Quantized Hall States, *Physical Review Letters* 52 (1984) 1583.
- [235] D. Arovas, J. R. Schrieffer, F. Wilczek, Fractional Statistics and the Quantum Hall Effect, *Phys. Rev. Lett.* 53 (1984) 722.
- [236] J. Fröhlich, A. Zee, Large-scale physics of the quantum Hall Fluid, *Nuclear Physics B* 364 (1991) 517.
- [237] V. Pasquier, F. Haldane, A dipole interpretation of the $v = 1/2$ state, *Nuclear Physics B* 516 (1998) 719.
- [238] N. Read, Lowest-Landau-level theory of the quantum Hall effect: The Fermi-liquid-like state of bosons at filling factor one, *Phys. Rev. B* 58 (1998) 16262.
- [239] L. Susskind, The Quantum Hall Fluid and Non-Commutative Chern-Simons Theory, unpublished (2001). [arXiv:arXiv:hep-th/0101029](https://arxiv.org/abs/hep-th/0101029).
- [240] A. P. Polychronakos, Quantum Hall States as matrix Chern-Simons theory, *Journal of High Energy Physics* 2001 (2001) 011.
- [241] E. Fradkin, V. Jejjala, R. G. Leigh, Non-commutative Chern-Simons for the Quantum Hall system and Duality, *Nuclear Physics B* 642 (2002) 483–500.
- [242] Z. Dong, T. Senthil, Noncommutative field theory and composite Fermi liquids in some quantum Hall systems, *Phys. Rev. B* 102 (2020) 205126.
- [243] H. Goldman, T. Senthil, Lowest Landau level theory of the bosonic Jain states, *Phys. Rev. B* 105 (2022) 075130.
- [244] S. C. Zhang, T. H. Hansson, S. Kivelson, Effective-Field-Theory Model for the Fractional Quantum Hall Effect, *Phys. Rev. Lett.* 62 (1989) 82.
- [245] F. Haldane, E. H. Rezayi, Periodic Laughlin-Jastrow wave functions for the fractional quantized Hall effect, *Physical Review B* 31 (1985) 2529.
- [246] X. G. Wen, Q. Niu, Ground-state degeneracy of the fractional quantum Hall states in the presence of a random potential and on high-genus Riemann surfaces, *Phys. Rev. B* 41 (1990) 9377.
- [247] A. López, E. Fradkin, Fractional quantum Hall effect and Chern-Simons gauge theories, *Phys. Rev. B* 44 (1991) 5246.
- [248] A. López, E. Fradkin, Response Functions and Spectrum of Collective Excitations of Fractional Quantum Hall Effect Systems, *Physical Review B* 47 (1993) 7080.
- [249] W. Kohn, Cyclotron Resonance and de Haas-van Alphen Oscillations of an Interacting Electron Gas, *Physical Review* 123 (1961) 1242.
- [250] A. López, E. Fradkin, Universal structure of the edge states of the fractional quantum Hall states, *Physical Review B* 59 (1999) 15323.
- [251] X.-G. Wen, A. Zee, Classification of Abelian quantum Hall states and matrix formulation of topological fluids, *Physical Review B* 46 (1992) 2290.
- [252] X. G. Wen, Topological Orders and Edge Excitations in Fractional Quantum Hall States, *Advances in Physics* 44 (1995) 405.
- [253] R. de Picciotto, M. Reznikov, M. Heiblum, V. Umansky, G. Bunin, D. Mahalu, Direct observation of a fractional charge, *Nature* 389 (1997) 162.
- [254] L. Saminadayar, D. C. Glattli, Y. Jin, B. Etienne, Observation of the $e/3$ Fractionally Charged Laughlin Quasiparticle, *Physical Review Letters* 79 (1997) 2526.
- [255] J. A. Hertz, Quantum critical phenomena, *Phys. Rev. B* 14 (1976) 1165–1184.
- [256] A. J. Millis, Effect of a nonzero temperature on quantum critical points in itinerant fermion systems, *Physical*

- Review B 48 (1993) 7183–7196.
- [257] C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, A. E. Ruckenstein, Phenomenology of the normal state of Cu-O high-temperature superconductors, Phys. Rev. Lett. 63 (1989) 1996–1999.
- [258] J. Polchinski, Low-energy dynamics of the spinon-gauge system, Nuclear Physics B 422 (1994) 617–633.
- [259] R. Shankar, Renormalization-group approach to interacting fermions, Reviews of Modern Physics 66 (1994) 129–192.
- [260] J. Polchinski, Effective Field Theory and the Fermi Surface, in: J. A. Harvey, J. Polchinski (Eds.), Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles, World-Scientific Publishing Co., Singapore, 1993, pp. 235–276, proceedings of TASI 1992, Boulder (Colorado, USA), June 1–26 (1992).
- [261] C. Nayak, F. Wilczek, Renormalization group approach to low temperature properties of a non-Fermi liquid metal, Nuclear Physics B 430 (1994) 534–562.
- [262] C. Nayak, F. Wilczek, Non-Fermi liquid fixed point in 2 + 1 dimensions, Nuclear Physics B 417 (1994) 359–373.
- [263] J. M. Maldacena, Large N limit of superconformal field theories and supergravity, Advances in Theoretical and Mathematical Physics 2 (1998) 231.
- [264] T. Faulkner, N. Iqbal, H. Liu, J. McGreevy, D. Vegh, Strange Metal Transport Realized by Gauge/Gravity Duality, Science 329 (2010) 1043–1047.
- [265] S. Fubini, Vertex Operators and the Quantum Hall Effect, Modern Physics Letters A 6 (1991) 347–358.
- [266] G. Moore, N. Read, Non-Abelions in the Fractional Quantum Hall Effect, Nuclear Physics B 360 (1991) 362.
- [267] G. Moore, N. Seiberg, Taming the Conformal Zoo, Phys. Lett. B 220 (1989) 422.
- [268] G. Moore, N. Seiberg, Classical and Quantum Conformal Field Theory, Commun. Math. Phys. 123 (1989) 177.
- [269] M. H. Freedman, A. Kitaev, Z. Wang, Simulation of topological field theories by quantum computers, Commun. Math. Phys. 227 (2002) 587.
- [270] M. H. Freedman, A. Kitaev, M. J. Larsen, Z. Wang, Topological Quantum Computation, Comm. Math. Phys. 227 (2002) 605.
- [271] J. Preskill, [Topological Quantum Computation](#), Lecture Notes for Physics 219: Quantum Computation, chapter 9; unpublished; Caltech (2004).
URL <http://www.theory.caltech.edu/~preskill/ph219/topological.pdf>
- [272] R. Willett, J. P. Eisenstein, H. L. Störmer, D. C. Tsui, A. C. Gossard, J. H. English, Observation of an even-denominator quantum number in the fractional quantum Hall effect, Physical Review Letters 59 (1987) 1776.
- [273] J. P. Eisenstein, R. Willett, H. L. Störmer, D. C. Tsui, A. C. Gossard, J. H. English, Collapse of the Even-Denominator Fractional Quantum Hall Effect in Tilted Fields, Physical Review Letters 61 (1988) 997.
- [274] W. Pan, J. S. Xia, H. L. Störmer, D. C. Tsui, C. Vicente, E. D. Adams, N. S. Sullivan, L. N. Pfeiffer, K. W. Baldwin, K. W. West, Experimental studies of the fractional quantum Hall effect in the first excited Landau level, Phys. Rev. B 77 (2008) 075307.
- [275] M. Greiter, X.-G. Wen, F. Wilczek, Paired hall State at Half Filling, Physical Review Letters 66 (1991) 3205.
- [276] M. Greiter, X. G. Wen, F. Wilczek, Paired Hall States, Nuclear Physics B 374 (1992) 567–614.
- [277] X. G. Wen, Non-Abelian statistics in the Fractional quantum Hall States, Physical Review Letters 66 (1991) 802.
- [278] W. Kohn, J. M. Luttinger, New Mechanism for Superconductivity, Physical Review Letters 15 (1965) 524.
- [279] A. V. Chubukov, Kohn-Luttinger effect and the instability of a two-dimensional repulsive Fermi liquid at $T = 0$, Physical Review B 48 (1993) 1097.
- [280] S. Raghu, S. A. Kivelson, Superconductivity from repulsive interactions in the two-dimensional electron gas, Physical Review B 83 (2011) 094518.
- [281] K. Park, V. Melik-Alaverdian, N. E. Bonesteel, J. K. Jain, Possibility of p -wave pairing of composite fermions at $\nu = 1/2$, Physical Review B 58 (1998) R10167.
- [282] B. I. Halperin, Theory of the quantized Hall conductance, hevetica Physica Acta 56 (1983) 75–102.
- [283] R. H. Morf, Transition from quantum Hall to compressible states in the second Landau level: New light on the $\nu = 5/2$ enigma, Physical Review Letters 80 (1998) 1505.
- [284] C. Nayak, F. Wilczek, $2n$ -quasihole states realize 2^{n-1} -dimensional spinor braiding statistics in paired quantum Hall states, Nuclear Physics B 479 (1996) 529–553.
- [285] N. Read, D. Green, Paired states of fermions in two dimensions with breaking of parity and time-reversal symmetries and the fractional quantum Hall effect, Physical Review B 61 (2000) 10267.
- [286] G. E. Volovik, An analog of the quantum Hall effect in a superfluid ^3He film, Soviet Physics JETP 67 (1988) 1804.
- [287] R. Jackiw, P. Rossi, Zero modes of the vortex-fermion system, Nuclear Physics B 190 (1981) 681.
- [288] E. J. Weinberg, Index calculations for the fermion-vortex system, Phys. Rev. D 24 (1981) 2669.
- [289] Y. Nishida, L. Santos, C. Chamon, Topological superconductors as nonrelativistic limits of Jackiw-Rossi and Jackiw-Rebbi models, Phys. Rev. B 82 (2010) 144513.
- [290] D. A. Ivanov, Non-abelian statistics of half-quantum vortices in p -wave superconductors, Phys. Rev. Lett. 86 (2001) 268.
- [291] M. Stone, S.-B. Chung, Fusion rules and vortices in $p_x + ip_y$ superconductors, Phys. Rev. B 73 (2006) 014505.
- [292] N. Read, E. Rezayi, Beyond paired quantum Hall states: Parafermions and incompressible states in the first excited Landau level, Physical Review B 59 (1999) 8084.
- [293] A. B. Z. V. A. Fateev, Nonlocal(parafermion) currents in two-dimensional conformal quantum field theory and self-dual critical points in \mathbb{Z}_N -symmetric statistical systems, Soviet Physics JETP 62 (1985) 215.
- [294] F. C. Alcaraz, R. Köberle, Hidden parafermions in $Z(N)$ theories, Physical Review D 24 (1981) 1652.
- [295] W. Pan, H. L. Störmer, D. C. Tsui, L. N. Pfeiffer, K. W. Baldwin, K. W. West, Fractional Quantum Hall Effect of Composite Fermions, Physical Review B 90 (2003) 016801.
- [296] J. S. Xia, W. Pan, C. L. Vicente, E. D. Adams, N. S. Sullivan, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer, K. W. Baldwin, K. W. West, Electron Correlation in the Second Landau Level: A Competition Between Many Nearly Degenerate Quantum Phases, Physical Review Letters 93 (2004) 176809.
- [297] E. H. Rezayi, N. Read, Non-Abelian quantized Hall states of electrons at filling factors $12/5$ and $13/5$ in the first excited Landau level, Physical Review B 79 (2009) 075306.
- [298] E. Fradkin, C. Nayak, A. Tsvelik, F. Wilczek, A Chern-Simons effective field theory for the Pfaffian quantum Hall state, Nuclear Physics B 516 (1998) 704.
- [299] E. Fradkin, C. Nayak, K. Schoutens, Landau-Ginzburg theories for non-abelian quantum Hall states, Nuclear

- Physics B 546 (1999) 711.
- [300] M. Barkeshli, X.-G. Wen, Effective field theory and projective construction for \mathbb{Z}_k parafermion fractional quantum Hall states, Physical Review B 81 (2010) 155302.
- [301] H. Goldman, R. Sohal, E. Fradkin, Landau-Ginzburg theories of non-Abelian quantum Hall states from non-Abelian bosonization, Physical Review B 100 (2019) 115111.
- [302] H. Goldman, R. Sohal, E. Fradkin, Non-Abelian fermionization and the landscape of quantum Hall phases, Phys. Rev. B 102 (2020) 195151.
- [303] H. Goldman, R. Sohal, E. Fradkin, Composite particle construction of the Fibonacci fractional quantum Hall state, Phys. Rev. B 103 (2021) 235118.
- [304] I. Affleck, Field theory methods and quantum critical phenomena, in: Strings, Fields and Critical Phenomena, Proceedings of the Les Houches Summer School 1988, Session XLIX, E. Brézin and J. Zinn-Justin editors, North-Holland, Amsterdam, the Netherlands, 1990, p. 563.
- [305] A. A. Belavin, A. M. Polyakov, A. B. Zamolodchikov, Infinite Conformal Symmetry in two-Dimensional Quantum Field Theory, Nuclear Physics B 241 (1984) 333.
- [306] I. Affleck, Universal Term in the Free Energy at a Critical Point and the Conformal Anomaly, Phys. Rev. Lett. 56 (1986) 746.
- [307] H. Blöte, J. L. Cardy, M. P. Nightingale, Conformal Invariance, the Central Charge, and Universal Finite-Size Amplitudes at Criticality, Phys. Rev. Lett. 56 (1986) 742.
- [308] G. Granger, J. P. Eisenstein, J. L. Reno, Observation of Chiral Heat Transport in the Quantum Hall Regime, Physical Review Letters 102 (2009) 086803.
- [309] A. Bid, N. Ofek, H. Inoue, M. Heiblum, C. L. Kane, V. Umansky, D. Mahalu, Observation of neutral modes in the fractional quantum Hall regime, Nature 466 (2010) 585–590.
- [310] C. L. Kane, M. Fisher, Transmission through barriers and resonant tunneling in an interacting one-dimensional electron gas, Phys. Rev. B 46 (1992) 15233.
- [311] C. Chamon, E. Fradkin, Distinct universal conductances in tunneling to quantum Hall states: The role of contacts, Phys. Rev. B 56 (1997) 2012.
- [312] A. M. Chang, L. N. Pfeiffer, K. W. West, Observation of Chiral Luttinger Behavior in Electron Tunneling into Fractional Quantum Hall Edges, Physical Review Letters 77 (1996) 2538.
- [313] A. M. Chang, Chiral Luttinger liquids at the fractional quantum Hall edge, Reviews of Modern Physics 75 (2004) 1449.
- [314] L. A. Cohen, N. L. Samuelson, T. Wang, T. Taniguchi, K. Watanabe, M. P. Zaletel, A. F. Young, Universal chiral Luttinger liquid behavior in a graphene fractional quantum Hall point contact, unpublished (2022). [arXiv:\protect\url{http://arxiv.org/abs/2212.01374}](http://arxiv.org/abs/2212.01374).
- [315] M. Hashisaka, T. Jonckheere, T. Akiho, S. Sasaki, J. Rech, T. Martin, K. Muraki, Andreev reflection of fractional quantum Hall quasiparticles, Nature Communications 12 (2021) 2794.
- [316] N. P. Sandler, C. Chamon, E. Fradkin, Noise measurements and fractional charge in fractional quantum Hall liquids, Phys. Rev. B 59 (1999) 12521.
- [317] F. P. Milliken, C. P. Umbach, R. A. Webb, Indications of a Luttinger liquid in the fractional quantum Hall regime, Solid State Communications 97 (1996) 309–313.
- [318] S. Roddaro, V. Pellegrini, F. Beltram, G. Giorgio Biasiol, L. Sorba, Interedge Strong-to-Weak Scattering Evolution at a Constriction in the Fractional Quantum Hall Regime, Physical Review Letters 93 (2004) 046801.
- [319] S. Roddaro, V. Pellegrini, F. Beltram, Quasi-particle tunneling at a constriction in a fractional quantum Hall state, Solid State Communications 131 (2004) 565.
- [320] P. W. Anderson, G. Yuval, D. R. Hammann, Exact results in the Kondo problem. II. Scaling theory, qualitatively correct solution, and some new results on one-dimensional classical statistical models, Phys. Rev. B 1 (1970) 4464.
- [321] N. Read, D. M. Newns, On the solution of the Coqblin-Schreiffer Hamiltonian by the large- N expansion technique, Journal of Physics C: Solid State Physics 16 (1983) 3273.
- [322] P. Fendley, A. W. W. Ludwig, H. Saleur, Exact Conductance through Point Contacts in the $\nu = 1/3$ Fractional Quantum Hall Effect, Phys. Rev. Lett. 74 (1995) 3005.
- [323] J. B. Miller, I. P. Radu, D. Z. Zumbühl, E. Levenson-Falk, M. A. Kastner, C. M. Marcus, L. N. Pfeiffer, K. W. West, Fractional quantum Hall effect in a quantum point contact at filling fraction 5/2, Nature Physics 3 (2007) 561.
- [324] I. Radu, J. B. Miller, C. M. Marcus, M. A. Kastner, L. N. Pfeiffer, K. W. Baldwin, K. W. West, Quasi-Particle Properties from Tunneling in the $\nu = 5/2$ Fractional Quantum Hall State, Science 320 (2008) 899.
- [325] C. L. Kane, M. Fisher, Nonequilibrium noise and fractional charge in the quantum Hall effect, Phys. Rev. Lett. 72 (1994) 724.
- [326] C. Chamon, D. E. Freed, X. G. Wen, Nonequilibrium quantum noise in chiral Luttinger liquids, Phys. Rev. B 53 (1996) 4033.
- [327] P. Fendley, A. Ludwig, H. Saleur, Exact Nonequilibrium dc Shot Noise in Luttinger Liquids and Fractional Quantum Hall Devices, Phys. Rev. Lett. 75 (1995) 2196.
- [328] M. Dolev, M. Heiblum, V. Umansky, A. Stern, D. Mahalu, Observation of a quarter of an electron charge at the $\nu = 5/2$ quantum Hall state, Nature 452 (2008) 829.
- [329] C. Chamon, D. E. Freed, S. A. Kivelson, S. L. Sondhi, X. G. Wen, Two point-contact interferometer for quantum Hall systems, Physical Review B 55 (1997) 2331.
- [330] F. E. Camino, W. Zhou, V. J. Goldman, Realization of a Laughlin quasiparticle interferometer: Observation of fractional statistics, Physical Review B 72 (2005) 075342.
- [331] B. I. Halperin, A. Stern, I. Neder, B. Rosenow, Theory of the Fabry-Pérot quantum Hall interferometer, Physical Review B 83 (2011) 155440.
- [332] P. Bonderson, K. Shtengel, J. K. Slingerland, Probing non-abelian statistics with quasiparticle interferometry, Physical Review Letters 97 (2006) 016401.
- [333] A. Stern, B. I. Halperin, Proposed Experiments to Probe the Non-Abelian $\nu = 5/2$ Quantum Hall State, Phys. Rev. Lett. 96 (2006) 016802.
- [334] R. L. Willett, L. N. Pfeiffer, K. W. West, Measurement of filling factor 5/2 quasiparticle interference with observation of charge $e/4$ and $e/2$ period oscillations, Proceedings of the National Academy of Sciences 106 (2009)

- [335] R. L. Willett, C. Nayak, K. Shtengel, L. N. Pfeiffer, K. W. West, Magnetic-Field-Tuned Aharonov-Bohm Oscillations and Evidence for Non-Abelian Anyons at $\nu = 5/2$, *Phys. Rev. Lett.* 111 (2013) 186401.
- [336] R. L. Willett, K. Shtengel, C. Nayak, L. N. Pfeiffer, Y. J. Chung, M. L. Peabody, K. W. Baldwin, K. W. W. West, Interference measurements of non-Abelian e/4 & Abelian e/2 quasiparticle braiding, (unpublished) (2019). [arXiv:\protect\vrule{width0pt}\protect\href{http://arxiv.org/abs/1905.10248}{arXiv:1905.10248}](http://arxiv.org/abs/1905.10248).
- [337] D. R. Nelson, J. M. Kosterlitz, Universal Jump in the Superfluid Density of Two-Dimensional Superfluids, *Phys. Rev. Lett.* 39 (1977) 1201–1205.
- [338] P. R. Thomas, M. Stone, Nature of the phase transition in a non-linear $O(2)_3$ model, *Nuclear Physics B* 144 (2) (1978) 513–524.
- [339] M. E. Peskin, Mandelstam-'t Hooft duality in abelian lattice models, *Annals of Physics* 113 (1978) 122–152.
- [340] C. Dasgupta, B. I. Halperin, Phase Transition in a Lattice Model of Superconductivity, *Phys. Rev. Lett.* 47 (1981) 1556–1560.
- [341] B. I. Halperin, T. C. Lubensky, S. keng Ma, First-Order Phase Transitions in Superconductors and Smectic-A Liquid Crystals, *Physical Review Letters* 32 (1974) 292.
- [342] M. Moshe, J. Zinn-Justin, Quantum field theory in the large N limit: a review, *Physics Reports* 385 (2003) 69–228.
- [343] H. Goldman, E. Fradkin, Loop Models, Modular Invariance, and Three Dimensional Bosonization, *Phys. Rev. B* 97 (2018) 195112.
- [344] E. Fradkin, F. A. Schaposnik, The Fermion-Boson Mapping in Three Dimensional Quantum Field Theory, *Phys. Lett. B* 338 (1994) 253.
- [345] C. P. Burgess, F. Quevedo, Bosonization as duality, *Nucl. Phys. B* 421 (1994) 373.
- [346] A. Chan, T. L. Hughes, S. Ryu, E. Fradkin, Effective field theories for topological insulators by functional bosonization, *Phys. Rev. B* 87 (2013) 085132.
- [347] A. P. O. Chan, T. Kvorning, S. Ryu, E. Fradkin, Effective hydrodynamic field theory and condensation picture of topological insulators, *Phys. Rev. B* 93 (2016) 155122.
- [348] S. Giombi, S. Minwalla, S. Prakash, S. P. Trivedi, S. R. Wadia, X. Yin, Chern-Simons theory with vector fermion matter, *Eur. Phys. J. C* 72 (2012) 1–65.
- [349] S. Jain, S. Minwalla, S. Yokoyama, Chern Simons duality with a fundamental boson and fermion, *Journal of High Energy Physics* 2013 (2013) 37.
- [350] O. Aharony, G. Gur-Ari, R. Yacoby, $d = 3$ bosonic vector models coupled to Chern-Simons gauge theories, *J. High Energy Phys.* 03 (2012) 37.
- [351] O. Aharony, G. Gur-Ari, R. Yacoby, Correlation functions of large N Chern-Simons-Matter theories and bosonization in three dimensions, *J. High Energy Phys.* 2012 (2012) 28.
- [352] O. Aharony, Baryons, monopoles and dualities in Chern-Simons-matter theories, *J. High Energy Phys.* 02 (2016) 93.
- [353] N. Seiberg, E. Witten, Gapped boundary phases of topological insulators via weak coupling, *Progress of Theoretical and Experimental Physics* 2016 (2016) ptw083.
- [354] A. Karch, D. Tong, Particle-Vortex Duality from 3D Bosonization, *Phys. Rev. X* 6 (2016) 031043.
- [355] D. T. Son, Is the composite fermion a Dirac particle?, *Phys. Rev. X* 5 (2015) 031027.
- [356] M. A. Metlitski, A. Vishwanath, Particle-vortex duality of two-dimensional dirac fermion from electric-magnetic duality of three-dimensional topological insulators, *Phys. Rev. B* 93 (2016) 245151.
- [357] C. Wang, T. Senthil, Dual Dirac liquid on the surface of the electron topological insulator, *Phys. Rev. X* 5 (2015) 041031.
- [358] J.-Y. Chen, J. H. Son, C. Wang, S. Raghu, Exact Boson-Fermion Duality on a 3D Euclidean Lattice, *Phys. Rev. Lett.* 120 (2018) 016602.
- [359] D. F. Mross, J. Alicea, O. I. Motrunich, Explicit Derivation of Duality between a Free Dirac Cone and Quantum Electrodynamics in (2 + 1) Dimensions, *Physical Review Letters* 117 (2016) 016802.
- [360] D. Simmons-Duffin, The Conformal Bootstrap, in: J. Polchinski, P. Vieira, O. DeWolfe (Eds.), *TASI 2015: Proceedings of the 2015 Theoretical Advanced Study Institute in Elementary Particle Physics, New Frontiers in Fields and Strings*, World Scientific, Singapore, 2017, pp. 1–74.
- [361] H. Goldman, E. Fradkin, Dirac composite fermions and emergent reflection symmetry about even-denominator filling fractions, *Physical Review B* 98 (2018) 165137.
- [362] D. Bohm, D. Pines, A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas, *Phys. Rev.* 92 (1953) 609–625.
- [363] F. D. M. Haldane, Luttinger's Theorem and Bosonization of the Fermi Surface, in: R. A. Broglia, J. R. Schrieffer (Eds.), *Proceedings of the International School of Physics "Enrico Fermi"*, Course CXXI "Perspectives in Many-Particle Physics", International School of Physics "Enrico Fermi", North-Holland, Amsterdam, The Netherlands, 1994, pp. 5–29, [arXiv:cond-mat/0505529](http://arxiv.org/abs/cond-mat/0505529).
- [364] A. Houghton, J. B. Marston, Bosonization and fermion liquids in dimensions greater than one, *Phys. Rev. B* 48 (1993) 7790–7808.
- [365] A. H. Castro Neto, E. Fradkin, Bosonization of the Low Energy Excitations of Fermi Liquids, *Phys. Rev. Lett.* 72 (1994) 1393.
- [366] A. H. Castro Neto, E. Fradkin, Bosonization of Fermi Liquids, *Phys. Rev. B* 49 (1994) 10877.
- [367] A. Houghton, H.-J. Kwon, J. B. Marston, Multidimensional bosonization, *Advances in Physics* 49 (2000) 141–228.
- [368] L. V. Delacrétaz, Y.-H. Du, U. Mehta, D. T. Son, Nonlinear bosonization of Fermi surfaces: The method of coadjoint orbits, *Phys. Rev. Research* 4 (2022) 033131.
- [369] Z. D. Shi, H. Goldman, D. V. Else, T. Senthil, Gifts from anomalies: Exact results for Landau phase transitions in metals, *SciPost Phys.* 13 (2022) 102.
- [370] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, M. P. A. Fisher, Deconfined Quantum Critical Points, *Science* 303 (2004) 1490.
- [371] A. L. Fitzpatrick, S. Kachru, J. Kaplan, S. Raghu, Non-Fermi-liquid fixed point in a Wilsonian theory of quantum critical metals, *Phys. Rev. B* 88 (2013) 125116.
- [372] M. A. Metlitski, D. F. Mross, S. Sachdev, T. Senthil, Cooper pairing in non-Fermi liquids, *Phys. Rev. B* 91 (2015) 115111.

- [373] V. Oganesyan, S. A. Kivelson, E. Fradkin, Quantum theory of the nematic Fermi liquid, *Physical Review B* 64 (2001) 195109.
- [374] M. J. Lawler, D. G. Barci, V. Fernández, E. Fradkin, L. Oxman, Non-perturbative behavior of the quantum phase transition to a nematic Fermi liquid, *Physical Review B* 73 (2006) 085101.
- [375] M. J. Lawler, E. Fradkin, Local quantum criticality at the nematic quantum phase transition, *Physical Review B* 75 (2007) 033304.
- [376] S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, Enhancement of Superconductivity near a Nematic Quantum Critical Point, *Phys. Rev. Lett.* 114 (2015) 097001.
- [377] S. A. Kivelson, E. Fradkin, V. J. Emery, Electronic liquid-crystal phases of a doped Mott insulator, *Nature* 393 (1998) 550–553.
- [378] E. Fradkin, S. A. Kivelson, M. J. Lawler, J. P. E. and Andrew P. Mackenzie, Nematic Fermi Fluids in Condensed Matter Physics, *Annual Review of Condensed Matter Physics* 1 (2010) 153–178.
- [379] E. Fradkin, S. A. Kivelson, J. M. Tranquada, Colloquium: Theory of intertwined orders in high temperature superconductors, *Reviews of Modern Physics* 87 (2015) 457–482.
- [380] R. M. Fernandes, P. P. Orth, J. Schmalian, Intertwined Vestigial Order in Quantum Materials: Nematicity and Beyond, *Annual Review of Condensed Matter Physics* 10 (2019) 133–154.