Field Theories of Condensed Matter Systems

Eduardo Fradkin
Physics Department
University of Illinois, Urbana-Champaign

ADDISON-WESLEY PUBLISHING COMPANY, 1991
The Advanced Book Program
Redwood City, California • Menlo Park, California • Reading, Massachusetts
New York • Don Mills, Ontario • Workingham, United Kingdom • Amsterdam
Bonn • Sydney • Singapore • Tokyo • Madrid • San Juan

their little brother, who helped by keeping everybody happy.

Finally, I must acknowledge the support of the Department of Physics and the Center for the Advanced Study of the University of Illinois. The help and understanding of the staff at Addison Wesley is also gratefully acknowleged.

Eduardo Fradkin
Urbana, Illinois
January 1991

Contents

CHAPTER 1

Introduction

1.1 Field Theory and Condensed Matter Physics

Condensed Matter Physics is a very rich and diverse field. If we are to define it as being "whatever gets published in the Condensed Matter section of a physics journal", we would conclude that it ranges from problems typical of material science to subjects which are as fundamental as particle physics or cosmology. Because of its diversity, it is sometimes hard to figure out where the field is going, particularly if you do not work in this field. Unfortunately, this is the case for people who have to make decisions about funding, grants, tenure and other unpleasant aspects in the life of a physicist. They have a hard time figuring-out where to put this subject which is neither applied science nor dealing with the smallest length scales or the highest energies. However, the richness of the field comes precisely from its diversity.

The past two decades have witnessed the development of two areas of Condensed Matter Physics which best illustrate the strengths of this field: Critical Phenomena and the Quantum Hall Effect. In both cases, it was the ability to produce extremely pure samples which allowed the discovery and the experimental study of the phenomenon. Its physical explanation required the use of concepts and the development of new theoretical tools, such as the renormalization group, conformal invariance and fractional statistics.

While the concept of conformal invariance was well known in field theory before Critical Phenomena was recognized as a field, its importance to the complete structure of the field theory was not understood. The situation changed with the development of the Renormalization Group (RG). For Condensed Matter Physics, the RG is the main tool for the interpretation of the experimental data, the conceptual framework and the computational algorithm which has allowed the theory to make powerful predictions. In Particle Physics, the RG is also a tool for the interpretation of the data. But, more importantly, the concept of infrared unstable fixed point has become the *definition* of the field theory itself.

Similarly, the Chern-Simons theories, which are field theories which describe systems exhibiting fractional statistics were known before the Quantum

Hall Effect (QHE) was discovered (actually they were discovered at about the same time) but were regarded as a curiosity of field theories below four dimensions: in other words, a beautiful piece of Mathemetical Physics but without relevance to "the world". We have come to recognize that Chern-Simons theories are the natural theoretical framework to describe the Quantum Hall Effect.

Another case to this point is superconductivity. Viable mechanisms for superconductivity have been known for the thirty-some years that have passed since the theory of Bardeen, Cooper and Schrieffer (BCS). This theory has successfully explained superconductivity, and a variety of related phenomena, in very diverse areas of Physics. This theory has been applied to such diverse areas of physics ranging from superconductivity in metals and superfluidity of liquid He_3 in condensed matter physics, to neutron stars and nuclear matter in nuclear physics, to dynamical symmetry breaking and grand unification mechanisms (such as technicolor) in elementary particle physics.

The origin of this constant interplay between Field Theory and Condensed Matter (or Statistical) Physics is that, despite their superficial differences, both fields deal with problems which involve a large number (macroscopic) of degrees of freedom which interact with each other. Thus, it should be of no surprise that the *same techniques* be used in both fields. The traditional trend was that field theory provided the tools (and the "sexy" terms) which were later adapted to a condensed matter problem. In turn, condensed matter models were used as "toy models" in which to try new techniques. For the most part, this is still the case. However, as the examples of the RG and the QHE show, the "toy models" can provide a framework for the understanding of much more general phenomenon. The *experimental accessibility* of condensed matter systems is just as important. The MOSFETS and heterojunctions, in which the Quantum Hall Effect is studied, have given us the surprisingly exact quantization of the Hall conductance whose understanding has required the use of Topology and Fiber Bundles.

The importance of Condensed Matter Physics to Field Theory, and viceversa, has been recognized at least since the 1950's. Landau and Feynman are perhaps the two theorists who best understood this deep connection. They worked in both fields and used their ideas and experience from one field in the other and then back.

1.2 What Has Been Included In This Book

This volume is an outgrowth of the course "Physics of Strongly Correlated Systems" which I taught at the University of Illinois at Urbana-Champaign during the Fall of 1989. Much of the material covered here has been the subject of intense research by a lot of people during the past four years. Most of what I discuss here has never been presented in a book, with the possible exception of some reprint volumes. While the *choice* of the material is motivated by

current work on High Temperature Superconductors, the methods and ideas have a wide range of applicability.

This book is not a textbook. Many of the problems, ideas and methods which are discussed here have become essential to our current understanding of Condensed Matter Physics. I have made a considerable effort to make the material largely self-contained. Many powerful methods, which are necessary for the study of condensed matter systems in the strong fluctuation limit, are discussed and explained in some detail within the context of the applications. Thus, although the theoretical apparatus is not developed systematically and in its full glory, this material may be useful to many graduate students, to learn both the subject and the methods. For the most part I have refrained myself from just quoting results without explaining where they come from. So, if a particular method happens to be appropriate to the study of a particular subject, I present a more or less detailed description of the method itself. Thus, a number of essential theoretical tools are discussed and explained. Unfortunately, I only was able to cover part of the material that I wanted to include. Perhaps the biggest omission is a description of Conformal Field Theory. This will have to wait for a second edition, if and when I ever get crazy enough to come back to this nightmare.

The material covered in this book deals with the theories of the three most fundamental problems in contemporary Condensed Matter Physics: Quantum Antiferromagnetism, the Quantum Hall Effect and High Temperature Superconducivity. The reader will find a detailed presntation of the modern theories of quantum magnetism and of the Quantum Hall effect but not an explicit treatment of the theories of High Temperature Superconductors.

In chapter 2 the symmetries of the Hubbard model are studied. The relation between the Hubbard model and quantum magnetism is also discussed. In chapter 3 I develop the theory of the magnetic instability of Fermi systems . As in the rest of the book, I move back and forth between the path integral and the Hamiltonian appoaches. The non-linear sigma model for the antiferromagnetic state is derived directly from the Hubbard model. In chapter 4 I give a detailed discussion of the physics of one-dimensional antiferromagnets. In chapter 5 I give a derivation of the path integral method for spin systems. The main applications here are quantum ferromagnets and antiferromagnets. The non-linear sigma model is derived within the semiclassical limit. I also present a detailed diuscussion of the role of topological excitations and of topological terms in the effective action for both one and two dimensional systems. In chapter 6 the reader will find the (so far) most current theories of disordered spin states. It includes the Resonating Valence Bond theories, flux phases, dimer or column states and the chiral spin liquid. The language of gauge theories is used throughout this chapter.

Theories of Anyons with applications in chiral spin liquids and Anyon Superconductivity, are presnted in chapters 7 and 8. The reader will find here a pedagogical presentation of the Chern-Simons theory of Fractional Statistics. Special attention is given to the connections between this problem and the topological theory of knots. A two dimensional version of the Jordan-Wigner

transformation is also derived.

Chapters 9 and 10 are devoted to the study of the Quantum Hall effect. In chapter 9 I discuss the topological theory of the quantization of the Hall conductance and its relation to fiber bundles. In chapter 10 the Fractiona Quantum Hall Effect is studied. The theory of the Laughlin wave function is presented in some detail as is the construction of the elementary excitations, their charge and statistics. The physics of the edge states is studied for a simple system. The methods of field theory are used to derive many of the results of Laughlin's theory but from a different persepctive. A careful discussion of the Landau-Ginzburg approach to the FQHE is given, including the issue of the existence of off-diagonal long range order. The chapter ends with an application of the Chern-Simons theory to the problem of the FQHE. The incompressibility is demonstrated by a direct computation of the spectrum of collective excitations. The fractional statistics of the elemntary excitations is also made explicit.

1.3 What Has Been Left Out

The course that I taught had for a subtitle "High Temperature Superconductors and Quantum Antiferromagnets". As the reader will soon find out, in the material that I have covered there is plenty of Quantum Antiferromagnetism but little superconductivity. This is not an oversight on my part. Rather, it is a reflection of what we understand today on this subject which is still a wide open field. Thus I chose not to include *the very latest fashion* on the subject but only what appears to be rather well established. This is a field that has produced a large number of very exciting ideas. However the *gedanken theories* still dominate. To an extent, this book reflects my own efforts in transforming several fascinating *gedanken theories* into something more or less concrete.

Still, the tantalizing properties of the High Temperature Superconductors seem to demand from us novel mechanisms such as Phil Anderson's RVB. But, of course, this is far from being universally accepted. After all, with theories like BCS being around, with so many successes in its bag, it seems strange that anybody would look for any other mechanism to explain the superconductivity of a set of rather complex materials. After all, who would believe that in order to understand the superconductivity produced by stuff made with copper and oxygen, mixed and cooked just right, would require the development of fundamentally new ideas. Right? Well, maybe yes, maybe not.

CHAPTER 2

The Hubbard Model

2.1 Introduction

All theories of strongly correlated electron systems begin with the Hubbard Model because of its simplicity. This is a model in which *band electrons* interact via a two-body *repulsive* Coulomb interaction. No phonons are present and in general no explicitly attractive interactions are included. For this reason, the Hubbard model has traditionally been associated with magnetism. Superconductivity, on the other hand, has traditionally (i.e. after BCS) been interpreted as an instability of the ground state resulting from *effectively attractive* interactions (say, electron-phonon as in BCS or other). A novel situation has arisen with Anderson's suggestion [Anderson 87] that the superconductivity of the new high T_c materials may arise from purely repulsive interactions. This suggestion was motivated by the fact that the superconductivity seems to originate from doping (i.e. extracting or adding charges) an otherwise insulating state.

The Hubbard model is a very simple model in which one imagines that, out of the many different bands which may exist in a solid, only very few states per unit cell do contribute significantly to the ground state properties. Thus, if a Bloch state of energy ϵ_p, momentum $\vec{p}$, and index α has a wavefunction $\Psi_{\vec{p},\alpha}$, one can construct Wannier states

$$\Psi_\alpha(\vec{r}_i) = \frac{1}{\sqrt{N}} \sum_{\vec{p}\epsilon B.Z.} e^{i\vec{p}\cdot\vec{r}_i} \Psi_{\vec{p},\alpha}(\vec{r}_i) \tag{2.1.1}$$

where $\vec{r}_i$ is the location of the i-th atom. The assumption here will be that only one (or a few) band indices matter, so I will drop the index α. The Coulomb interaction matrix elements are

$$U_{ij,i'j'} = \int d^3r_1 d^3r_2 \Psi_i^*(\vec{r}_1)\Psi_j^*(\vec{r}_2)\tilde{V}(\vec{r}_1 - \vec{r}_2)\Psi_{i'}(\vec{r}_1)\Psi_{j'}(\vec{r}_2) \tag{2.1.2}$$

(in three dimensions) where $\tilde{V}$ is the (screened) Coulomb interaction. Since $\tilde{V}$ is expected to decay as the separation increases, the largest term will be the "on site" term: $U_{ii,ii} \equiv U$. Next will come nearest neighbors, etc. Moreover,

since the Wannier functions have exponentially decreasing overlaps, $U_{ij,i'j'}$ is expected to decrease rather rapidly with the separation $|i-j|$.

The second quantized Hamiltonian (in the Wannier functions basis) is

$$H = -\sum_{\substack{\vec{r}_i,\vec{r}_j \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}_i) t_{ij} c_\sigma(\vec{r}_j) + c_\sigma^\dagger(\vec{r}_j) t_{ij} c_\sigma(\vec{r}_i) \right) + \\ + \frac{1}{2} \sum_{\substack{i,j,i',j' \\ \sigma=\uparrow,\downarrow}} U_{ij,i'j'} c_\sigma^\dagger(\vec{r}_i)\, c_\sigma^\dagger(\vec{r}_j) c_\sigma(\vec{r}_{j'}) c_\sigma(\vec{r}_{i'}), \tag{2.1.3}$$

where $c_\sigma^\dagger(\vec{r}\,)$ creates an electron at site $\vec{r}$ with spin σ (or more precisely, at the unit cell $\vec{r}$ in the band responsible for the Fermi surface) and satisfies

$$\{c_\sigma(\vec{r}\,), c_{\sigma'}^\dagger(\vec{r}\,')\} = \delta_{\sigma,\sigma'}\delta_{\vec{r},\vec{r}'}, \\ \{c_\sigma(\vec{r}\,), c_{\sigma'}(\vec{r}\,')\} = 0. \tag{2.1.4}$$

The Hubbard model is an approximation to the more general Hamiltonian, Eq. (2.1.3), in which the hopping is restricted to nearest neighboring sites:

$$t_{ij} = \begin{cases} t & \text{if } i,\ j \text{ are nearest neighbors,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.1.5}$$

and the Coulomb interaction is assumed to be screened. If only the "on-site" term is kept

$$U_{ij,i'j'} = U\delta_{ij}\delta_{i'j'}\delta_{ii'}, \tag{2.1.6}$$

the resulting model Hamiltonian

$$H = -t \sum_{\substack{\langle \vec{r},\vec{r}' \rangle \\ \sigma=\uparrow,\downarrow}} (c_\sigma^\dagger(\vec{r}\,) c_\sigma(\vec{r}\,') + \text{h.c.}) + U \sum_{\vec{r}} n_\uparrow(\vec{r}\,) n_\downarrow(\vec{r}\,) \tag{2.1.7}$$

is known as the one band Hubbard model. In Eq. (2.1.7), we have dropped the lattice sites labelling and $\langle , \rangle$ means nearest neighbor sites. This is the tight-binding approximation and represents the one-band Hubbard model. We have introduced

$$n_\sigma(\vec{r}\,) = c_\sigma^\dagger(\vec{r}\,) c_\sigma(\vec{r}\,) \tag{2.1.8}$$

From the Pauli principle we get $n_\sigma = 0, 1$ or $n_\sigma^2 = n_\sigma$ at every sites.

The Hilbert space of this system is the tensor product of only *four* states per site, representing $|0\rangle$ as nothing, $|\uparrow\rangle$ as an electron with spin up, $|\downarrow\rangle$ as an electron with spin down and $|\uparrow\downarrow\rangle$ as an up-down pair. The states $|0\rangle$ and $|\uparrow\downarrow\rangle$ are spin singlets (i.e. $s = 0$).

It is convenient to define the following operators. The spin operator $\vec{S}(\vec{r}\,)$ is defined by (summation convention is assumed)

$$\vec{S}(\vec{r}) = \frac{\hbar}{2} c_\sigma^\dagger(\vec{r}\,) \vec{\tau}_{\sigma\sigma'} c_{\sigma'}(\vec{r}\,), \tag{2.1.9}$$

where $\vec{\tau}$ are the (three) Pauli matrices

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.1.10}$$

The particle number operator at site $\vec{r}$ (or *charge*) is

$$n(\vec{r}\,) = \sum_\sigma n_\sigma(\vec{r}\,) = \sum_\sigma c_\sigma^\dagger(\vec{r}\,) c_\sigma(\vec{r}\,) \equiv c_\sigma^\dagger(\vec{r}\,) 1_{\sigma\sigma'} c_{\sigma'}(\vec{r}\,) \tag{2.1.11}$$

and the associated total charge Q

$$Q = e \sum_{\vec{r}} n(\vec{r}\,) \equiv e N_e. \tag{2.1.12}$$

2.2 Symmetries of the Hubbard model

2.2.1 SU(2) Spin

Suppose we rotate the local spin basis (i.e. the quantization axis)

$$c'_\sigma(\vec{r}\,) = U_{\sigma\sigma'} c_{\sigma'}(\vec{r}\,) \tag{2.2.1}$$

where U is an SU(2) matrix. Under such a unitary transformation, the spin $\vec{S}$ transforms as follows

$$\begin{aligned} S'^a(\vec{r}\,) &= R^{ab} S^b(\vec{r}\,) \\ &= \frac{\hbar}{2} c'^\dagger(\vec{r}\,) \tau^a c'(\vec{r}\,) \\ &= \frac{\hbar}{2} c^\dagger(\vec{r}\,) \left(U^{-1} \tau^a U\right) c(\vec{r}\,), \end{aligned} \tag{2.2.2}$$

where R^{ab} is a rotation matrix induced by the SU(2) transformation of the fermions:

$$U^{-1} \tau^a U = R^{ab} \tau^b. \tag{2.2.3}$$

In other words, we have a rotation of the quantization axis.

The axis of quantization can be chosen arbitrarily. Thus, the Hubbard Model Hamiltonian should not change its form under a rotation of the spin quantization axis. This is not apparent in the standard form of the interaction

$$H_1 = U \sum_{\vec{r}} n_\uparrow(\vec{r}\,) n_\downarrow(\vec{r}\,). \tag{2.2.4}$$

But we can write this in a somewhat different form in which the SU(2) symmetry becomes explicit. Consider the operator

$$\sum_{\vec{r}} \left(\vec{S}(\vec{r}\,)\right)^2 = \sum_{\substack{\vec{r} \\ a=1,2,3}} S^a(\vec{r}\,) S^a(\vec{r}\,). \tag{2.2.5}$$

By expanding the components and making use of the SU(2) identity

$$\sum_{a=1,2,3} \tau^a_{\alpha\beta}\tau^a_{\gamma\delta} = 2\delta_{\alpha\delta}\delta_{\beta\gamma} - \delta_{\alpha\beta}\delta_{\gamma\delta} \tag{2.2.6}$$

one gets

$$\sum_{\vec{r}} \left(\vec{S}(\vec{r}\,)\right)^2 = \sum_{\vec{r}} \left(\frac{1}{4}n(\vec{r}\,) - \frac{3}{2}n_\uparrow(\vec{r}\,)n_\downarrow(\vec{r}\,)\right). \tag{2.2.7}$$

Thus, we can write

$$H_1 = U\sum_{\vec{r}} n_\uparrow(\vec{r}\,)n_\downarrow(\vec{r}\,) = -\frac{2U}{3}\vec{S}^2(\vec{r}\,) + \frac{N_eU}{6}. \tag{2.2.8}$$

The last term is a constant which can be dropped. The Hamiltonian now has the form

$$H = -t\sum_{\substack{\langle \vec{r},\vec{r}'\rangle \\ \sigma=\uparrow,\downarrow}} \left(c^\dagger_\sigma(\vec{r}\,)c_\sigma(\vec{r}\,') + \text{h.c.}\right) - \frac{2U}{3}\sum_{\vec{r}} \left(\vec{S}(\vec{r}\,)\right)^2 + \frac{N_eU}{6} \tag{2.2.9}$$

which is manifestly SU(2) invariant.

For $U > 0$, the intearction energy is lowered if *the total spin at each site is maximized.* Thus, one should expect some sort of magnetic ground state, at least if each site has one particle (on the average). This state requires that the system somehow should pick a global (i.e. the same for all sites) quantization axis. In other words, the global SU(2) spin symmetry may be spontaneously broken. This has important consequences which we will discuss later.

2.2.2 U(1) (Charge)

We are free to change the phase of the one-particle wavefunction

$$c'_\sigma(\vec{r}\,) = e^{i\theta}c_\sigma(\vec{r}\,). \tag{2.2.10}$$

Here, $e^{i\theta}$ is an element of the group U(1), and group elements satisfy

$$e^{i\theta}e^{i\theta'} = e^{i(\theta+\theta')}. \tag{2.2.11}$$

The Hamiltonian is invariant under this U(1) transformation. This is nothing but charge conservation. For example, if we had terms which would not conserve charge, like

$$c^\dagger_\uparrow(\vec{r}\,)c^\dagger_\downarrow(\vec{r}\,') \to e^{i2\theta}c^\dagger_\uparrow(\vec{r}\,)c^\dagger_\downarrow(\vec{r}\,'), \tag{2.2.12}$$

we would not have this invariance.

Suppose now that we couple this system to the electromagnetic field $(A_0, \vec{A})$. We expect three effects.

1) A *Zeeman coupling* given by

$$H_{\text{Zeeman}} = g\sum_{\vec{r}} \vec{S}(\vec{r}\,)\cdot\vec{B}(\vec{r}\,) \tag{2.2.13}$$

which couples the spin $\vec{S}(\vec{r}\,)$ with the local magnetic field $\vec{B}(\vec{r}\,)$ so as to align it along the $\vec{B}(\vec{r}\,)$ direction.

2) An *orbital coupling* for electrons in a crystal with one-particle Hamiltonian

$$H(\vec{p}) = \frac{1}{2m_e}\left(\vec{p} - \frac{e}{c}\vec{A}\right)^2 + \text{periodic potential.} \tag{2.2.14}$$

In the tight-binding approximation, we must therefore modify the kinetic energy term according to

$$\begin{aligned} H_0 &\equiv -t \sum_{\substack{\langle \vec{r},\vec{r}\,'\rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,)c_\sigma(\vec{r}\,') + \text{h.c.}\right) \\ &\to -t \sum_{\substack{\langle \vec{r},\vec{r}\,'\rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,)e^{\frac{ie}{\hbar c}\int_{\vec{r}}^{\vec{r}\,'} d\vec{x}\cdot\vec{A}(\vec{x})}c_\sigma(\vec{r}\,') + c_\sigma^\dagger(\vec{r}\,')e^{-\frac{ie}{\hbar c}\int_{\vec{r}}^{\vec{r}\,'} d\vec{x}\cdot\vec{A}(\vec{x})}c_\sigma(\vec{r}\,)\right). \end{aligned} \tag{2.2.15}$$

We should now check the gauge invariance under the transformation

$$\vec{A}' = \vec{A} + \vec{\nabla}\Lambda \tag{2.2.16}$$

where Λ is an arbitrary function. We get the change

$$\begin{aligned} A'(\vec{r},\vec{r}\,') &\equiv \int_{\vec{r}}^{\vec{r}\,'} d\vec{x}\cdot\vec{A}'(\vec{x}) \\ &= A(\vec{r},\vec{r}\,') + \Lambda(\vec{r}\,') - \Lambda(\vec{r}\,). \end{aligned} \tag{2.2.17}$$

Thus the kinetic energy term is gauge invariant

$$\begin{aligned} H_0' &\equiv -t \sum_{\substack{\langle \vec{r},\vec{r}\,'\rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma'^\dagger(\vec{r}\,)e^{\frac{ie}{\hbar c}A'(\vec{r},\vec{r}\,')}c_\sigma'(\vec{r}\,') + \text{h.c.}\right) \\ &= -t \sum_{\substack{\langle \vec{r},\vec{r}\,'\rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,)e^{-i\theta(\vec{r}\,)}e^{\frac{ie}{\hbar c}\left(A(\vec{r},\vec{r}\,')+\Lambda(\vec{r}\,')-\Lambda(\vec{r}\,)\right)}e^{+i\theta(\vec{r}\,')}c_\sigma(\vec{r}\,') + \text{h.c.}\right), \end{aligned} \tag{2.2.18}$$

provided that the *local* change of phase be given by

$$\theta(\vec{r}\,) \equiv -\frac{e}{\hbar c}\Lambda(\vec{r}\,). \tag{2.2.19}$$

3) An *electrostatic coupling* given by

$$H_{\text{electrostatic}} = \sum_{\vec{r},\sigma} eA_0(\vec{r}\,)c_\sigma^\dagger(\vec{r}\,)c_\sigma(\vec{r}\,) \tag{2.2.20}$$

which couples the particle density to $A_0(\vec{r}\,)$.

2.2.3 Particle-Hole Transformations

In the case of a *bipartite lattice* (i.e. a lattice which is the union of two interpenetrating sublattices A and B) we get additional symmetries.

i) First, the *sign of* t can be changed. Consider the transformation

$$\begin{aligned} c_\sigma(\vec{r}\,) &\to +c_\sigma(\vec{r}\,) \qquad \text{if} \quad \vec{r}\epsilon A, \\ c_\sigma(\vec{r}\,) &\to -c_\sigma(\vec{r}\,) \qquad \text{if} \quad \vec{r}\epsilon B, \end{aligned} \tag{2.2.21}$$

under which the kinetic energy changes sign:

$$tc_\sigma^\dagger(\vec{r}\,)c_\sigma(\vec{r}\,') \to -tc_\sigma^\dagger(\vec{r}\,)c_\sigma(\vec{r}\,'), \quad \vec{r}\epsilon A \quad \vec{r}\,'\epsilon B, \tag{2.2.22}$$

while the potential energy is left unchanged. This transformation leaves the canonical commutation relations unchanged and therefore leaves the *spectrum unchanged.*

ii) Now consider the *particle-hole* transformation

$$\begin{aligned} c_\uparrow(\vec{r}\,) &= d_\uparrow(\vec{r}\,), \\ c_\downarrow(\vec{r}\,) &= \begin{cases} +d_\downarrow^\dagger(\vec{r}\,), & \vec{r}\,\epsilon\, A. \\ -d_\downarrow^\dagger(\vec{r}\,), & \vec{r}\,\epsilon\, B. \end{cases} \end{aligned} \tag{2.2.23}$$

The Hamiltonian $H(t,U)$, Eq. (2.1.7), changes into $H(t,-U)+UN_\uparrow$, where $N_\uparrow$ is the total number of up spin (which is conserved), since under this transformation we get

$$n_\uparrow + n_\downarrow = c_\uparrow^\dagger c_\uparrow + c_\downarrow^\dagger c_\downarrow = d_\uparrow^\dagger d_\uparrow + d_\downarrow d_\downarrow^\dagger = d_\uparrow^\dagger d_\uparrow - d_\downarrow^\dagger d_\downarrow + 1 \tag{2.2.24}$$

and

$$n_\uparrow - n_\downarrow = c_\uparrow^\dagger c_\uparrow - c_\downarrow^\dagger c_\downarrow = d_\uparrow^\dagger d_\uparrow - d_\downarrow d_\downarrow^\dagger = d_\uparrow^\dagger d_\uparrow + d_\downarrow^\dagger d_\downarrow - 1. \tag{2.2.25}$$

Similarly, the charge Q and the component S_z of the total spin transform as

$$Q \to S_z + 1, \qquad S_z \to Q - 1. \tag{2.2.26}$$

Thus the attractive and the repulsive cases map into each other and, at the same time, spin maps into charge and vice versa. Note that for *negative* U the Hamiltonian favors local singlets ($s = 0$), i.e. empty and doubly occupied sites.

2.3 The Strong Coupling Limit

We consider now the strong coupling limit of the Hubbard model, i.e. $U \to \infty$. The half-filled case is special but important. We will consider it first.

2.3.1 Half-Filled System ($U > 0$)

Recall that the interaction term

$$H_{\text{int}} \equiv -\frac{2}{3}U \sum_{\vec{r}} \left(\vec{S}(\vec{r}\,)\right)^2 \tag{2.3.1}$$

forces the spin $\vec{S}$ to be largest if U becomes infinitely large, i.e. doubly occupied sites are forbidden. Only $|\uparrow\rangle$ and $|\downarrow\rangle$ states are kept in this large U limit at half-filling. The interaction part of the Hamiltonian has infinitely many eigenstates. Any spin configuration is an eigenstate. In order to lift this massive degeneracy we will keep the effects of fluctuations induced by the kinetic energy term to leading order in an expansion in $\frac{t}{U}$. We have to solve a problem in degenerate perturbation theory.

Suppose we begin with any configuration which can be labeled by the local z-component of the spin $|\{\sigma(\vec{r})\}\rangle$ (see Fig. 2.1(a)). In an expansion in powers of $\frac{t}{U}$, we have intermediate states in which one site will become doubly occupied and, at the same time, another site becomes empty (see Fig.2.1(b)). This state has an energy U above that of the degenerate ground states. The matrix element (squared) is t^2. There is also a multiplicity factor of 2 since this process can occur in two different ways. Hence we expect that the relevant parameter of the effective Hamiltonian should be $\frac{2t^2}{U}$. Also, the final state has to be either the same one as the initial state or it can differ at most by a spin exchange (see Fig.2.1(c)). The natural candidate for the effective Hamiltonian is the Quantum Heisenberg Antiferromagnet.

We can obtain this result by carrying out this expansion explicitly [Emery 79]. Let H_0 and H_1 denote the kinetic and interaction terms of the Hubbard Hamiltonian H, Eq. (2.1.7),

$$\begin{aligned} H_0 &= -t \sum_{\substack{\langle \vec{r},\vec{r}' \rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}) c_\sigma(\vec{r}\,') + \text{h.c.} \right), \\ H_1 &= U \sum_{\vec{r}} n_\uparrow(\vec{r}) n_\downarrow(\vec{r}). \end{aligned} \tag{2.3.2}$$

Let $|\alpha\rangle$ be any of the 2^N states with every sites occupied by a spin either up or down. Here, $|\alpha\rangle$ is an eigenstate of H_1 with eigenvalue $E_1 = 0$.

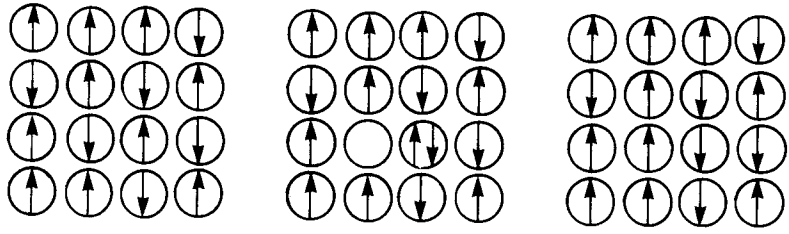


Figure 2.1 (a) and (c) are two configurations of spins corresponding to orthogonal groundstates of H_0. They differ by the exchange of two neighboring spins. Configuration (b) corresponds to a virtual state. The circles represent the unit cells of the cristal and the arrows are the components of the spin along the quantization axis.

We will use Brillouin-Wigner perturbation theory [Baym 74]. Consider the Schrödinger's Equation

$$H|\Psi\rangle = E|\Psi\rangle \tag{2.3.3}$$

where $|\Psi\rangle$ is any eigenstate. We can write

$$(E - H_1)\,|\Psi\rangle = H_0|\Psi\rangle. \tag{2.3.4}$$

Formally, we get

$$\begin{aligned} |\Psi\rangle &= \frac{1}{E - H_1} H_0|\Psi\rangle \\ &= \frac{\hat{P}}{E - H_1} H_0|\Psi\rangle + \sum_\alpha |\alpha\rangle \frac{\langle\alpha|H_0|\Psi\rangle}{E - E_1}, \end{aligned} \tag{2.3.5}$$

where

$$H_1|\alpha\rangle = E_1|\alpha\rangle \tag{2.3.6}$$

and

$$\hat{P} = 1 - \sum_\alpha |\alpha\rangle\langle\alpha| \tag{2.3.7}$$

projects out of the unperturbed states. Clearly $\hat{P}$ commutes with H_1. Define $|\Psi_\alpha\rangle$ as the solution of the equation

$$|\Psi_\alpha\rangle = |\alpha\rangle + \frac{\hat{P}}{E - H_1} H_0|\Psi_\alpha\rangle. \tag{2.3.8}$$

Let a_α be given by

$$a_\alpha = \frac{\langle\alpha|H_0|\Psi\rangle}{E - E_1}. \tag{2.3.9}$$

Then we can write

$$|\Psi\rangle = \sum_\alpha a_\alpha |\Psi_\alpha\rangle. \tag{2.3.10}$$

If we iterate Eq. (2.3.8) to first order in powers of $\frac{\hat{P}}{E-H_1}H_0$, we find

$$|\Psi_\alpha\rangle \approx |\alpha\rangle + \frac{\hat{P}}{E - H_1} H_0|\alpha\rangle \approx |\alpha\rangle - \frac{1}{U} H_0|\alpha\rangle \tag{2.3.11}$$

since $\langle\beta|H_0|\alpha\rangle = 0$ at half-filling. Thus, if we insert Eq. (2.3.11) into Eq. (2.3.10) and in turn insert this into Eq. (2.3.9), we get

$$(E - E_1)a_\alpha = \frac{1}{|U|} \sum_{\alpha'} \langle\alpha|H_0^2|\alpha'\rangle a_{\alpha'}. \tag{2.3.12}$$

This is the same as the Schrödinger equation for the Hamiltonian $H_0' = \frac{H_0^2}{|U|}$ where H_0', at half filling, is given by

$$H_0' = \frac{2t^2}{|U|} \sum_{\langle\vec{r},\vec{r}'\rangle} \vec{S}(\vec{r}\,) \cdot \vec{S}(\vec{r}\,'). \tag{2.3.13}$$

In other words we find the spin one-half Quantum Heisenberg Antiferromagnet with the exchange coupling $J = \frac{2t^2}{|U|}$. This result is valid for the half-filled system in any dimension and lattice.

2.3.2 Away from half-filling

Clearly other processes are now allowed. If $U \gg t$, doubly occupied sites are energetically very expensive. Thus the restricted Hilbert space now consists of configurations made of empty sites (holes), and up and down spins.

The kinetic energy term will allow for charge motion since empty sites (holes) will be able to move. These holes carry electric charge but they have no spin. The effective Hamiltonian now has the form of the *"t-J" model.*

$$H = -t \sum_{\substack{\langle \vec{r},\vec{r}' \rangle \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,) c_\sigma(\vec{r}\,') + \text{h.c.} \right) + J \sum_{\langle \vec{r},\vec{r}' \rangle} \vec{S}(\vec{r}\,) \cdot \vec{S}(\vec{r}\,') \tag{2.3.14}$$

where $J = \frac{2t^2}{|U|}$ with the *constraint*

$$\sum_{\sigma=\uparrow,\downarrow} c_\sigma^\dagger(\vec{r}\,) c_\sigma(\vec{r}\,) = n(\vec{r}\,) = 0, 1 \tag{2.3.15}$$

which eliminates doubly occupied sites. Now we have *two* separately conserved quantities: the charge Q which equals the number of holes and the spin component $S_z = \sum_{\vec{r}} S_z(\vec{r}\,)$. It is clear that as the holes move they can induce spin-flip processes. The spin configurations get disrupted by the motion of holes and the long-range order (antiferromagnetic) may be destroyed. Presumably it should take a finite density of holes to destroy the long-range order. We will discuss this strong coupling limit at great length. Let us first consider the opposite case.

2.4 Weak Coupling Limit

In the weak-coupling limit $U \ll t$ we may think of the interaction as a weak perturbation. One therefore expects that the states of a *weakly coupled* electron gas may be qualitatively similar to the states of a free-electron gas. This picture is usually called a *Fermi liquid* [Pines 66]. The main *assumption* is that there is a one-to-one correspondence between the states of a free fermion system and those in a weakly interacting one.

For free fermions, the Hamiltonian reduces to the kinetic energy term. For the Hubbard model we have

$$H_0 = -\sum_{\substack{\vec{r},\vec{r}' \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,) t(\vec{r} - \vec{r}\,') c_\sigma(\vec{r}\,') + c_\sigma^\dagger(\vec{r}\,') t(\vec{r} - \vec{r}\,') c_\sigma(\vec{r}\,) \right). \tag{2.4.1}$$

It is convenient to go to Fourier space (momentum). Assume that we are in d space dimensions and that the lattice has N^d sites with N even (for

simplicity). With $V \equiv N^d$, we define

$$c_\sigma(\vec{r}\,) = \frac{1}{V}\sum_{\vec{k}} e^{i\vec{k}\vec{r}} c_\sigma(\vec{k}) \tag{2.4.2}$$

where

$$\vec{k} = \frac{2\pi}{N}(n_1, \ldots, n_d) - (\pi, \ldots, \pi) \tag{2.4.3}$$

and $1 \le n_i \le N$. Thus the momenta k_i vary over the range $\frac{2\pi}{N} - \pi \le k_i \le \pi$. In the thermodynamic limit $N \to \infty, \frac{2\pi}{N} \to 0$ and the $k's$ become uniformly distributed in the interval $-\pi \le k_i \le \pi$ (the Brillouin zone).

Remember the following properties of Fourier transforms. Let $k = \frac{2\pi}{N}n - \pi, n = 1, \ldots, N$ and $f(k)$ some function of k. We have the Riemann sums

$$\begin{aligned}\sum_k f(k) &= \frac{1}{\Delta k}\sum_k \Delta k f(k) \\ &\to \frac{N}{2\pi}\int_{-\pi}^{\pi} dk f(k) \qquad \text{as} \quad N \to \infty\end{aligned} \tag{2.4.4}$$

where $\Delta k = \frac{2\pi}{N}$. The extension to the d-dimensinal case is

$$\sum_{\vec{k}} f(\vec{k}) \to N^d \int_{-\pi}^{\pi} \frac{d^d k}{(2\pi)^d} f(\vec{k}) \quad \text{as} \quad N \to \infty. \tag{2.4.5}$$

In particular, as $N \to \infty$

$$\begin{aligned}&\frac{1}{N^d}\sum_{\vec{k}} e^{i\vec{k}\cdot(\vec{r}-\vec{r}\,')} = \delta_{\vec{r},\vec{r}\,'} \to \int_{-\pi}^{\pi} \frac{d^d k}{(2\pi)^d} e^{i\vec{k}\cdot(\vec{r}-\vec{r}\,')} = \delta_{\vec{r},\vec{r}\,'}, \\ &\sum_{\vec{r}} e^{i(\vec{k}-\vec{k}\,')\cdot\vec{r}} = N^d \delta_{\vec{k},\vec{k}\,'} \to (2\pi)^d \delta^{(d)}(\vec{k}-\vec{k}\,'), \\ &c_\sigma(\vec{r}\,) \to \int_{-\pi}^{\pi} \frac{d^d k}{(2\pi)^d} e^{i\vec{k}\cdot\vec{r}} c_\sigma(\vec{k}).\end{aligned} \tag{2.4.6}$$

The canonical commutation relations

$$\{c_\sigma^\dagger(\vec{r}\,), c_{\sigma'}(\vec{r}\,')\} = \delta_{\sigma,\sigma'}\delta_{\vec{r},\vec{r}\,'} \tag{2.1.4}$$

become in the same limit

$$\{c_\sigma^\dagger(\vec{k}), c_{\sigma'}(\vec{k}')\} = \delta_{\sigma,\sigma'}\delta_{\vec{k},\vec{k}'} \to (2\pi)^d \delta_{\sigma,\sigma'}\delta^{(d)}(\vec{k}-\vec{k}'). \tag{2.4.7}$$

The kinetic energy then takes the form

$$\begin{aligned}H_0 &= -\sum_{\substack{\vec{r},\vec{r}\,' \\ \sigma=\uparrow,\downarrow}} \left(c_\sigma^\dagger(\vec{r}\,) t(\vec{r}-\vec{r}\,') c_\sigma(\vec{r}\,') + c_\sigma^\dagger(\vec{r}\,') t(\vec{r}-\vec{r}\,') c_\sigma(\vec{r}\,)\right) \\ &= -\sum_{\substack{\vec{r},\vec{r}\,' \\ \sigma=\uparrow,\downarrow}} t(\vec{r}-\vec{r}\,') \int \frac{d^d k}{(2\pi)^d} \int \frac{d^d k'}{(2\pi)^d} \left(e^{-i\vec{k}\cdot\vec{r}+i\vec{k}'\cdot\vec{r}\,'} c_\sigma^\dagger(\vec{k}) c_\sigma(\vec{k}') + \text{h.c.}\right).\end{aligned} \tag{2.4.8}$$

If by $t(\vec{k})$ we denote the Fourier transform of $t(\vec{l})$:

$$t(\vec{k}) = \sum_{\vec{l}} t(\vec{l}) e^{-i\vec{k}\cdot\vec{l}}, \tag{2.4.9}$$

we can write

$$\sum_{\vec{r},\vec{r}'} t(\vec{r} - \vec{r}\,') e^{-i\vec{k}\cdot\vec{r} + i\vec{k}'\cdot\vec{r}'} = t(\vec{k})(2\pi)^d \delta^{(d)}\left(\vec{k} - \vec{k}'\right). \tag{2.4.10}$$

For the case

$$t(\vec{r} - \vec{r}\,') \equiv t(\vec{l}) = \begin{cases} \mathrm{t} & \text{for nearest neighbors,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.4.11}$$

we get

$$t(\vec{k}) = 2t \sum_{j=1}^{d} \cos k_j \tag{2.4.12}$$

and a free Hamiltonian of the form

$$H_0 = \sum_{\sigma=\uparrow,\downarrow} \int \frac{d^d k}{(2\pi)^d} \epsilon(\vec{k}) c_\sigma^\dagger(\vec{k}) c_\sigma(\vec{k}) \tag{2.4.13}$$

with

$$\epsilon(\vec{k}) = -2t(\vec{k}) = -4\mathrm{t} \sum_{j=1}^{d} \cos k_j. \tag{2.4.14}$$

The ground state is found by filling up the Fermi sea. Thus, if we have N particles, the total number of momentum states with energy smaller than E is (assuming that $\epsilon_{\vec{k}}$ has its minimum at $\vec{k} = (0 \ldots 0)$) determined by the constant energy curves $\epsilon(\vec{k}) \equiv \epsilon$ (see Fig. 2.2 and Fig. 2.3). For instance, in the one-dimensional case we find

$$\epsilon(k) = -4t \cos k. \tag{2.4.15}$$

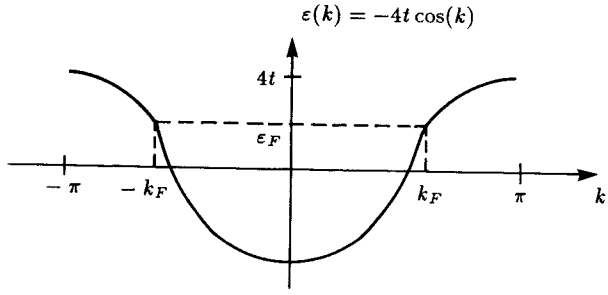


Figure 2.2 One particle spectrum of H_0.

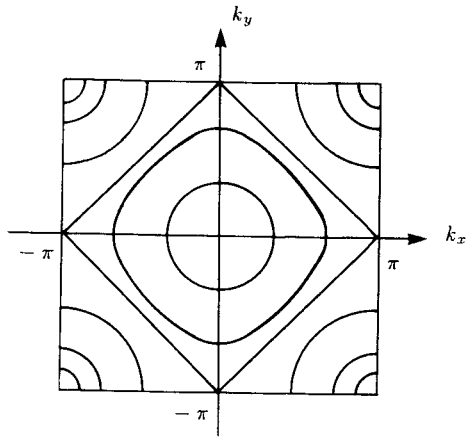


Figure 2.3 Constant energy curves for H_0 on a square lattice.

If $\mathcal{N}$ is the number of particles and N the number of sites, we get

$$\mathcal{N} = 2N \int_{-k_F}^{k_F} \frac{dk}{2\pi} = \frac{2Nk_F}{\pi} \tag{2.4.16}$$

and

$$k_F = \frac{\pi \mathcal{N}}{2N} \equiv \frac{\pi}{2}\rho \tag{2.4.17}$$

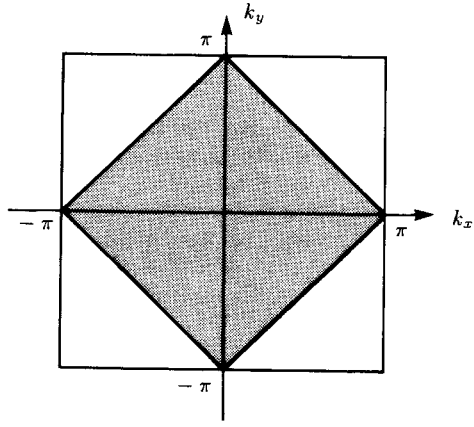


Figure 2.4 Fermi sea for free electrons on a square lattice at half filling (the lattice spacing is unity).

where ρ is the linear density. At half filling $k_F = \frac{\pi}{2}$ and $\epsilon(k_F)$ vanishes. In higher dimensions we determine the constant energy curves in the same way. For a half-filled system we just fill up the negative energy states to obtain the Fermi sea. This is so because this band has $E \leftrightarrow -E$ symmetry ("particle-hole") and there are as many states with positive energy as there are with negative energy. The *Fermi surface* is defined by $\epsilon_F = 0$ and for a square lattice is rectangular (square) (see Fig. 2.4).

The occupation number

$$n_{\vec{k}} = \sum_{\sigma=\uparrow,\downarrow} c_\sigma^\dagger(\vec{k}) c_\sigma(\vec{k}) \tag{2.4.18}$$

has a jump at the Fermi surface in both the free and interacting cases (see Fig.2.5).

2.5 Correlation Functions

The fermion Green's function (or propagator) plays an important role in the theory. We can define it in terms of field operators in the Heisenberg representation

$$c_\sigma(\vec{r},t) = e^{iHt} c_\sigma(\vec{r}\,) e^{-iHt}. \tag{2.5.1}$$

The fermion propagator is defined by

$$G_{\sigma\sigma'}(\vec{r},t;\vec{r}\,',t') = -i\langle \mathrm{Gnd}|T c_\sigma(\vec{r},t) c_{\sigma'}^\dagger(\vec{r}\,',t')|\mathrm{Gnd}\rangle \tag{2.5.2}$$

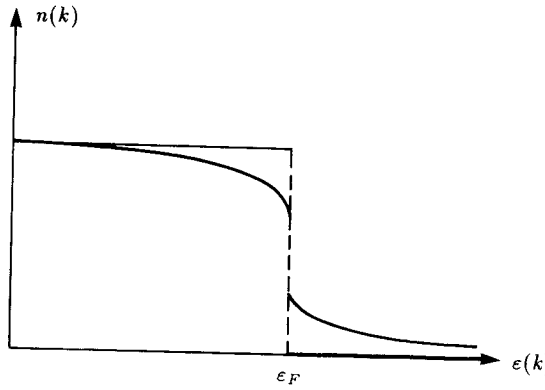


Figure 2.5 Occupancy of energy levels of the non-interacting case (straight line) and interacting (curved line) cases.

where $|\mathrm{Gnd}\rangle$ stands for the ground state of the system and T means a time ordered product of operators

$$TA(t)B(t') = A(t)B(t')\theta(t-t') \pm B(t')A(t)\theta(t'-t) \tag{2.5.3}$$

with a +(−) sign for bosons (fermions) and

$$\theta(t) = \begin{cases} 1 & \text{if } t > 0. \\ \\ 0 & \text{if } t < 0. \end{cases} \tag{2.5.4}$$

For a translationally invariant and time independent system, we can write $G_{\sigma\sigma'}(\vec{r},t;\vec{r}\,',t')$ in terms of its Fourier transform [Abrikosov 63], [Fetter 71], [Doniach 74]

$$G_{\sigma\sigma'}(\vec{r},t;\vec{r}\,',t') = \int \frac{d^dk}{(2\pi)^d} \int \frac{d\omega}{2\pi} e^{i(\vec{k}\cdot(\vec{r}-\vec{r}\,')-\omega(t-t'))} G_{\sigma\sigma'}(\vec{k},\omega). \tag{2.5.5}$$

In principle $G_{\sigma\sigma'}(\vec{k},\omega)$ is a 2x2 spin matrix. In the case of the non-interacting system (and for any spin isotropic ground states, for this matter), $G^{(0)}_{\sigma\sigma'}(\vec{k},\omega)$ is very simple to compute [Fetter 71]. The result is

$$G_{\sigma\sigma'}(\vec{k},\omega) = \delta_{\sigma\sigma'} \lim_{\nu\to 0^+} \left(\frac{\theta\left(\epsilon(\vec{k}) - \epsilon_F\right)}{\omega - \epsilon(\vec{k}) + i\nu} + \frac{\theta\left(\epsilon_F - \epsilon(\vec{k})\right)}{\omega - \epsilon(\vec{k}) - i\nu} \right). \tag{2.5.6}$$

The poles of $G^0_{\alpha\beta}(\vec{k},\omega)$ exhibit the physical one-particle excitation spectrum

$$\omega = \epsilon(\vec{k}). \tag{2.5.7}$$

A *weakly interacting system* (a *Fermi liquid*) resembles a non-interacting one in the sense that the physical low-energy excitations look like weakly interacting fermions. Thus the fermion propagator retains its pole structure albeit with a renormalized dispersion relation $\omega = \epsilon_{\mathrm{ren}}(\vec{k})$ and a non-trivial residue for energies close to the Fermi energy. In other words, for $\omega \sim \epsilon_F$ the propagator should look like

$$\lim_{\substack{\omega\to\epsilon_F \\ \nu\to 0^+}} G(\vec{k},\omega) \approx \lim_{\substack{\omega\to\epsilon_F \\ \nu\to 0^+}} \left(\frac{Z}{\omega - \epsilon_{\mathrm{ren}}(\vec{k}) + i\nu} + G_{\mathrm{reg}} \right) \tag{2.5.8}$$

where Z is the residue and G_{reg} does not have any singularities close to ϵ_F. The *wave function renormalization* Z measures the strength of the jump of the fermion occupation number $n_{\vec{k}}$ at the Fermi surface. These excitations are the fermion quasiparticles or *electrons* of the Fermi liquid. These quasiparticles are assumed to be stable in the sense that the poles lie on the real energy axis. Any imaginary part would imply a decay amplitude ("damping").

In addition to one-particle states, a Fermi liquid has a large class of many-particle excitations. These include particle-hole excitations (i.e. density fluctuations), spin excitations, paramagnons and others. These *collective modes* are *bound states* which exist only in an interacting system. Many of these modes are damped. Others are not. We can study the collective modes by means of

the many-particle Green functions. Several *correlation functions* are going to be important to us. They are:

1) Density correlation functions

$$K_{00}(\vec{r},t;\vec{r}\,',t') = \langle \mathrm{Gnd}|T\hat{n}(\vec{r},t)\hat{n}(\vec{r}\,',t')|\mathrm{Gnd}\rangle, \tag{2.5.9}$$

where $\hat{n}(\vec{r},t)$ is the local normal-ordered density operator

$$\hat{n}(\vec{r},t) = \sum_{\sigma=\uparrow,\downarrow} c_\sigma^\dagger(\vec{r},t)c_\sigma(\vec{r},t) - \rho \tag{2.5.10}$$

and ρ is the average density.

2) Current correlation functions

$$K_{ii'}(\vec{r},t;\vec{r}\,',t') = \langle \mathrm{Gnd}|TJ_i(\vec{r},t)J_{i'}(\vec{r}\,',t')|\mathrm{Gnd}\rangle, \tag{2.5.11}$$

where $J_i(\vec{r},t)$ is the current operator, which, in the case of the Hubbard model in the absence of external electromagnetic fields, is

$$J_i(\vec{r},t) = -it\sum_{\sigma=\uparrow,\downarrow}\left(c_\sigma^\dagger(\vec{r},t)c_\sigma(\vec{r}+\vec{e}_i,t) + \mathrm{h.c.}\right). \tag{2.5.12}$$

There is no need for normal-ordering here since the ground state is not expected to spontaneously carry a non-zero current.

3) Spin correlation functions

$$K^{aa'}(\vec{r},t;\vec{r}\,',t') = \langle \mathrm{Gnd}|TS^a(\vec{r},t)S^{a'}(\vec{r}\,',t')|\mathrm{Gnd}\rangle. \tag{2.5.13}$$

We will see below that if the spin symmetry is broken spontaneously (i.e. magnetism), $K^{aa'}$ can be non-zero as $(\vec{r},t)$ and $(\vec{r}\,',t')$ are infinitely separated from each other. In fact, the limit (at equal time!)

$$\lim_{|\vec{r}-\vec{r}\,'|\to\infty}\frac{1}{3}\sum_{a=1}^{3}K^{aa}(\vec{r},t;\vec{r}\,',t) = M^2 \tag{2.5.14}$$

represents the amplitude of the order parameter M. In the case of Neel states the limit has to be taken on a given sublattice. A non-zero M is a signal of a spontaneously broken symmetry.

All of these two-particle operators have the common feature that they conserve particle numbers locally and the excitations are electrically neutral. In the Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity local particle number conservation is lost locally (but not globally) since one could break a Cooper pair at one location and form it again somewhere else. In BCS theory [Schrieffer 64], a Cooper pair is a bound state of an *electron* with momentum $\vec{k}$ and spin up and another *electron* with momentum $-\vec{k}$ and spin down, with $\vec{k}$ on the Fermi surface. This state has charge $2e$ and is a spin singlet. Other bound states are also possible, as in ^3He and in heavy fermion systems. We can now define a Cooper pair correlation function.

4) Cooper pair correlation function

$$C(\vec{k},t;\vec{k}',t') = \langle \mathrm{Gnd}|Tc_\uparrow^\dagger(\vec{k},t)c_\downarrow^\dagger(-\vec{k},t)c_\uparrow(\vec{k}',t')c_\downarrow(-\vec{k}',t')|\mathrm{Gnd}\rangle \tag{2.5.15}$$

for the s-wave case, and its generalizations for the p-wave and other cases.

From an experimental point of view, what one can measure are *susceptibilities*. In other words, one couples the system to a weak external field. From Linear Response Theory [Fetter 71]we know how to relate the *response functions* (i.e. causal propagators) to the time-ordered functions. The susceptibilities are the Fourier transformed causal propagators. For instance, the magnetic susceptibility $X^{aa'}(\vec{k},\omega)$ is defined as follows

$$\chi^{aa'}(\vec{k},\omega) = \sum_{\vec{r}} \int dt e^{i(\vec{k}\cdot\vec{r}-\omega t)} \chi^{aa'}(0,0;\vec{r},t), \tag{2.5.16}$$

where $\chi^{aa'}(\vec{r},t;\vec{r}\,',t')$ is the *causal (or retarded) propagator*

$$\chi^{aa'}(\vec{r},t;\vec{r}\,',t') = \theta(t'-t)\langle \mathrm{Gnd}|S^a(\vec{r},t)S^{a'}(\vec{r}\,',t')|\mathrm{Gnd}\rangle. \tag{2.5.17}$$

In particular the *static susceptibility* $\chi^{aa'}(\vec{k},0)$ measures the response of the system to a weak external magnetic field. In the uniform limit $\vec{k} \to 0$, we are measuring the response to a weak uniform magnetic field. Thus this is the static ferromagnetic susceptibility. If we want to probe a Neel state we must couple to the *staggered magnetization* and hence use a staggered field. This is difficult to achieve. However, using neutron scattering we can measure $\chi(\vec{k},0)$ for all $\vec{k}$, in particular the case $\vec{k} = (\pi,\pi,\pi)$ which is the staggered or Neel susceptibility.

References

[Abrikosov 63] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinski. *Methods of Quantum Field Theory in Statistical Physics*, Dover, 1963.

[Anderson 87] P.W. Anderson. Science 235, 1196 (1987).

[Baym 74] G. Baym. *Lectures on Quantum Mechanics*, Benjamin, 1974.

[Doniach 74] S. Doniach and E.H. Sondheimer. *Green's Functions for Solid State Physicists*, Benjamin, 1974.

[Emery 79] V.J. Emery. In *Highly Conducting One-Dimensional Solids*. Edited by J.T. Devreese, R.P. Evrard and V.E. van Doren, Plenum, 1979.

[Fetter 71] A.L. Fetter and J.D. Walecka. *Quantum Theory of Many-Particle Systems*, McGraw-Hill, 1971.

[Pines 66] D. Pines and P. Nozieres. *The Theory of Quantum Liquids*, Benjamin, 1966.

[Schrieffer 64] J.R. Schrieffer. *Theory of Superconductivity*, Addison-Wesley, 1964.

CHAPTER 3

The Magnetic Instability of the Fermi System

The Hubbard model was originally introduced as the simplest system which may exhibit an insulating (Mott) state. This state is the result of strong electron-electron interactions. In this chapter we consider the Hubbard model at half-filling. The main goal here is the study of the magnetic properties of its ground state. Apart from an exact solution in one-dimension, no exact results are available for this problem. This leads to the use of several approximations. The most popular one, and the oldest, is the Mean-Field-Theory (MFT). In the MFT one has the *bias* that the ground state does have some sort of magnetic order (i.e. ferromagnet, Néel antiferromagnet, etc.). The problem is then usually solved by means of a variational ansatz. However, one is usually interested in more than just the ground state energy which, after all, is not directly measurable and depends very sensitively on the properties at short distances. Most often we wish to know what are the long distance, low frequency, properties of the correlation and response functions of this theory. Moreover, in some cases, such as in one-dimension, the fluctuations overwhelm the MFT predictions.

In this chapter we will consider the standard MFT (i.e. Hartree-Fock) which is expected to become accurate at *weak* coupling. We will consider both ferromagnetic and antiferromagnetic states. We will also rederive these results using path-integrals. As a byproduct, we will also have a theory of the fluctuations: the non-linear sigma model.

3.1 Mean-Field Theory

Let us consider now the effects of interactions on the unperturbed ground state. It is convenient to consider the Fourier transform of the interaction

term of the Hubbard Hamiltonian, Eq. (2.1.6),

$$H_1 = U \sum_{\vec{r}} n_\uparrow(\vec{r}\,) n_\downarrow(\vec{r}\,)$$
$$= U \int_{\vec{k}_1 \ldots \vec{k}_4} (2\pi)^d \delta^d(-\vec{k}_1 + \vec{k}_2 - \vec{k}_3 + \vec{k}_4) c_\uparrow^\dagger(\vec{k}_1) c_\uparrow(\vec{k}_2) c_\downarrow^\dagger(\vec{k}_3) c_\downarrow(\vec{k}_4) \tag{3.1.1}$$

where $\int_{\vec{k}_i}$ is a shorthand for $\int_{BZ} \frac{d^d k_i}{(2\pi)^d}$. Notice that on a lattice momentum is conserved, modulo a reciprocal lattice vector $\vec{G}$.

Let us discuss first the simpler one-dimensional case. There are two Fermi points (at $\pm k_F$). Thus we can classify the excitations as left or right moving particles and holes with either spin orientation. We have the following scattering processes (see Fig. 3.1): (a) Forward scattering, (b) Backward scattering and (c) Umklapp scattering.

In case (a) two particles scatter with a small momentum transfer and do not change the direction of their individual motion. In case (b) a right mover becomes a left mover and vice-versa. In case (c) two right movers become left movers. This process violates momentum conservation but if the total momentum violation equals a reciprocal lattice vector, the (Umklapp) process is allowed. This occurs for $k_F = \frac{\pi}{2}$, which is the half-filled case (see Fig. 3.1c).

Case (b), backward scattering, implies a scattering process involving two degenerate states: exchanging a right mover with a left mover and vice-versa (see Fig. 3.1b). Since the energy denominator is zero we may have an instability of the perturbation theory. This is an antiferromagnetic instability since the momentum transfer is π. Conversely, instability in the forward scattering channel is a symptom of ferromagnetism.

In dimensions higher than one, the situation is more complex due to the intricacies of the Fermi surface. For instance, in the case of a half-filled square lattice the Fermi surface is a square (see Fig. 2.5). A scattering process of

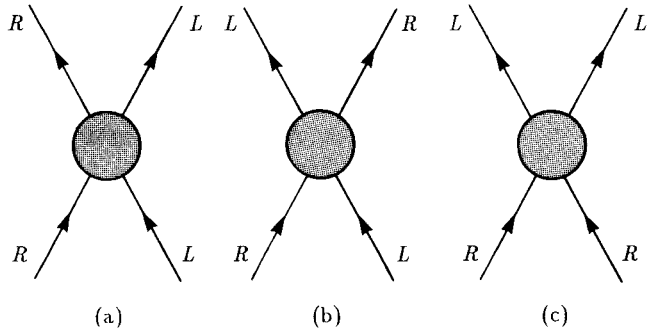


Figure 3.1 Forward (a), Backward (b) and Umklapp (c) scattering of right (R) and left (L) movers.

the type shown in Fig. 3.2 may induce an antiferromagnetic instability. That is, we *exchange* particles with opposite spin from opposite sides of the Fermi surface. Once again this involves a momentum exchange of (π, π) or $(\pi, -\pi)$, depending the case.

However, for a nearly empty band, only quasiforward scattering should matter, and the relevant momentum exchange should be zero (ferromagnetism). Other cases, involving other momentum exchanges, are possible. These instabilities generally give rise to a *spin-density-wave* of wave vector $\vec{k}$. The ferromagnetic state occurs when $\vec{k} = 0$ and the Néel antiferromagnetic state occurs when $\vec{k} = (\pi, \pi)$.

We want to develop a theory of these instabilities. As we see, we need to find bound states of a certain wave vector $\vec{k}$ and the ground state will have to be rebuilt in the form of a coherent superposition of these bound states. Since we do not know how to solve this problem exactly, some sort of Mean-Field theory is necesary. There are several ways of achieving this goal: (a) Hartree-Fock and random phase approximation (RPA), (b) variational wavefunctions and (c) $\frac{1}{N}$ expansions.

These three approaches are, to some extent, physically equivalent. While (a) and (b) (and mostly (a)) are commonly discussed in textbooks, the $\frac{1}{N}$ expansion is a rather novel technique and, for that reason, is not usually available to students (although it has become pervasive throughout the current

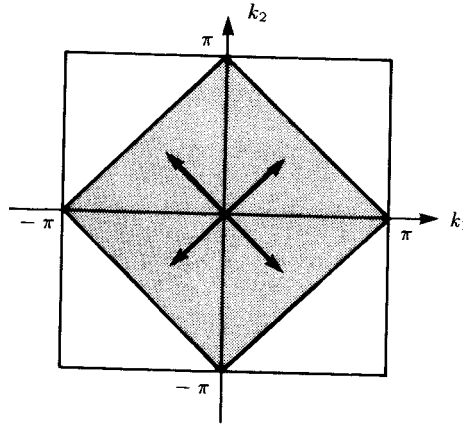


Figure 3.2 Nesting vectors for the Brillouin zone of a square lattice at half filling. This nesting property is responsible for an antiferromagnetic instability.

literature).

The Hartree-Fock-RPA approach involves choosing a particular set of the Feynman diagrams which one can argue gives the "most important" contributions. While in one dimension it is possible to select diagrams according to their degree of divergence in the infrared, the situation is far less clear in two or more dimensions. Typically, one has to choose a particular process and sum all the leading contributions which contribute to the process and, at the same time, do not violate any conservation laws, a "conserving approximation", in the terminology of Baym and Kadanoff [Kadanoff 62]. Such is the spirit of Hartree-Fock-RPA theories. Similarly in the variational wavefunction approach, one chooses variational states which are essentially inspired by RPA-like calculations.

Let us first discuss a simple form of Mean-Field theory. We start from the Hubbard Hamiltonian

$$H = -t \sum_{\langle \vec{r},\vec{r}' \rangle} c^\dagger_\sigma(\vec{r}\,) c_\sigma(\vec{r}\,') + \text{ h.c. } - \frac{2}{3} U \sum_{\vec{r}} \left(\vec{S}(\vec{r}\,) \right)^2 \tag{3.1.2}$$

which is quartic in the fermionic operators, since

$$\vec{S}(\vec{r}\,) = \frac{1}{2} c^\dagger_\sigma(\vec{r}\,) \vec{\tau}_{\sigma\sigma'} c_{\sigma'}(\vec{r}\,). \tag{3.1.3}$$

(From now on, I will be using the summation convention on repeated spin indices.)

The interaction term of the Hubbard Hamiltonian, Eq. (3.1.2), is quartic in fermion operators. In general a non-linear problem of this sort is not solvable except in some very special cases, such as one-dimensional systems. A standard approach is the Mean-Field-Approximation (or *Hartree-Fock approximation*) in which the quartic term is factorized in terms of a fermion *bilinear* times a Bose field, which is usually treated classically. In other words one simply ignores the dynamics of the Bose field. Consider, for instance, the Hamiltonian H'

$$H' = -t \sum_{\langle \vec{r},\vec{r}\,' \rangle} c^\dagger_\sigma(\vec{r}\,) c_\sigma(\vec{r}\,') + \text{ h.c. } + \frac{3}{8U} \sum_{\vec{r}} \vec{M}^2(\vec{r}\,) + \sum_{\vec{r}} \vec{M}(\vec{r}\,) \cdot \vec{S}(\vec{r}\,). \tag{3.1.4}$$

which can be regarded as a *linearized* version of H in terms of a Bose field $\vec{M}(\vec{r}\,)$ which, as we will see below, represents the local magnetization.

However, there is something in this expression that isn't quite right, since the field $\vec{M}(\vec{r}\,)$ does not have any dynamics. It looks like a variational parameter. Indeed, in the *Hartree-Fock* approximation, one assumes that a certain operator, say $\vec{S}(\vec{r}\,)$, picks up an expectation value. One then has to shift the operator by its expectation value and neglect fluctuations (this is the Mean-Field-Approximation). Therefore, one writes

$$\vec{S}(\vec{r}\,) = \langle \vec{S}(\vec{r}\,) \rangle + \left(\vec{S}(\vec{r}\,) - \langle \vec{S}(\vec{r}\,) \rangle \right). \tag{3.1.5}$$

The term in brackets clearly represents fluctuations. Thus

$$\vec{S}^2 = \langle\vec{S}\rangle^2 + \left(\vec{S} - \langle\vec{S}\rangle\right)^2 + 2\langle\vec{S}\rangle \cdot \left(\vec{S} - \langle\vec{S}\rangle\right). \tag{3.1.6}$$

To neglect fluctuations means to drop the second term. Thus we write

$$H = H_{MF} + H_{fl} \tag{3.1.7}$$

where

$$H_{MF} = -t \sum_{\langle\vec{r},\vec{r}\,'\rangle} c^\dagger_\sigma(\vec{r}\,)c_\sigma(\vec{r}\,') + \text{ h.c. } - \frac{2}{3}U\sum_{\vec{r}}\langle\vec{S}(\vec{r}\,)\rangle^2 + \\ - \frac{4}{3}U\sum_{\vec{r}}\langle\vec{S}(\vec{r}\,)\rangle \cdot \left(\vec{S}(\vec{r}\,) - \langle\vec{S}(\vec{r}\,)\rangle\right) \tag{3.1.8}$$

and the fluctuation part H_{fl} is the rest. We can also write

$$H_{MF} = -t \sum_{\langle\vec{r},\vec{r}\,'\rangle} c^\dagger_\sigma(\vec{r}\,)c_\sigma(\vec{r}\,') + \text{ h.c. } + \frac{2U}{3}\sum_{\vec{r}}\langle\vec{S}(\vec{r}\,)\rangle^2 - \frac{4U}{3}\sum_{\vec{r}}\langle\vec{S}(\vec{r}\,)\rangle \cdot \vec{S}(\vec{r}\,) \tag{3.1.9}$$

which is just H', Eq. (3.1.4), if we identify $\vec{M}(\vec{r}\,) \equiv -\frac{4U}{3}\langle\vec{S}(\vec{r}\,)\rangle$.

We can give dynamics to $\vec{M}\,(\vec{r})$ by using the following device. Consider first the simple classical oscillator problem with a degree of freedom $\vec{M}$ and Lagrangian

$$L = \frac{1}{2}m\dot{\vec{M}}^2 - \frac{g}{2}\vec{M}^2 - \vec{M}\cdot\vec{S}. \tag{3.1.10}$$

The equations of motion of this oscillator are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\vec{M}}} = \frac{\partial L}{\partial \vec{M}} \tag{3.1.11}$$

which imply

$$m\ddot{\vec{M}} = -g\vec{M} - \vec{S}. \tag{3.1.12}$$

At the quantum level these equations become the equation of motion of the *operator* $\vec{M}(t)$ in the Heisenberg representation.

Consider now the limit $m \to 0$. The *only* smooth trajectories, i.e. with $\ddot{\vec{M}}$ finite, which are possible in this limit satisfy

$$g\vec{M} + \vec{S} = 0. \tag{3.1.13}$$

The Hamiltonian H' has to be regarded in precisely the same way. We have to add a kinetic energy term at each site of the form $\sum_{\vec{r}} \frac{\vec{P}^2(\vec{r}\,)}{2m}$, where $\vec{P}$ and $\vec{M}$ obey canonical commutation relations, and consider the limit $m \to 0$. One should not panic at the apparent divergence in the kinetic energy term: the equations of motion are taking care of it. We are going to come back to this later on, when we discuss the path-integral form. There everything is simpler.

Thus we see that the Lagrange multiplier field $\vec{M}$ is dynamical in the sense that it follows the configurations of fermions in detail. In Mean-Field theory

(i.e. the Hartree-Fock approximation), one replaces $\vec{M}$ by some *static* configuration (time-independent) which, in turn, is determined by the condition that the ground state energy should be lowest.

Let us now look at $H_{\rm MF}$ in Fourier space. The Lagrange multiplier field $\vec{M}(\vec{r}\,)$ has the Fourier transform

$$\vec{M}(\vec{r}\,) = \frac{1}{N^d}\sum_{\vec{k}} e^{i\vec{k}\cdot\vec{r}}\vec{M}(\vec{k}) \to \int \frac{d^dk}{(2\pi)^d} e^{i\vec{k}\cdot\vec{r}}\vec{M}(\vec{k}) \quad \text{for} \quad N\to\infty. \tag{3.1.14}$$

We can now write $H_{\rm MF}$ in the form

$$H_{\rm MF} = N^d \int_{\vec{k}} \left(\sum_\sigma \epsilon(\vec{k}) n_\sigma(\vec{k}) + \frac{3}{8U}|\vec{M}(\vec{k})|^2 + \vec{M}^*(\vec{k})\cdot\vec{S}(\vec{k}) \right) \tag{3.1.15}$$

since

$$\vec{M}(\vec{k}) = \vec{M}^*(-\vec{k}) \tag{3.1.16}$$

and where

$$\vec{S}(\vec{k}) = \int_{\vec{k}'} c^\dagger_\alpha(\vec{k}')\frac{\vec{\tau}_{\alpha\beta}}{2}c_\beta(\vec{k}'+\vec{k}). \tag{3.1.17}$$

while $\epsilon(\vec{k}) = -4t\sum_{j=1}^d \cos k_j$.

The second term on the right hand side of Eq. (3.1.15) implies that a configuration with the Fourier component $\vec{M}(\vec{k})$ induces scattering processes which mix one-particle states differing by $\vec{k}$.

3.1.1 Ferromagnetic State

Let us consider first the *ferromagnetic* solution in which $\vec{M}(\vec{r}\,)$ is a constant $\vec{M}_0$. In Fourier space, we have

$$\vec{M}(\vec{k}) = \vec{M}_0(2\pi)^d\delta^d(\vec{k}). \tag{3.1.18}$$

Then $H_{\rm MF}$ is (with the volume $V = N^d$)

$$H_{\rm MF} = \frac{3}{8U}V\vec{M}_0^2 + V\int_{\vec{k}} \left(\epsilon(\vec{k})c^\dagger_\sigma(\vec{k})c_\sigma(\vec{k}) + c^\dagger_\sigma(\vec{k})\frac{\vec{\tau}_{\sigma\sigma'}}{2}c_{\sigma'}(\vec{k})\cdot\vec{M}_0 \right). \tag{3.1.19}$$

Since the direction of $\vec{M}_0$ is arbitrary, one can choose the z-axis (i.e. the quantization axis) to be parallel to $\vec{M}_0$ without any loss of generality. One then finds

$$H_{\rm MF} = \frac{3}{8U}V\vec{M}_0^2 + V\int_{\vec{k}} \left(\left(\epsilon(\vec{k}) + \frac{1}{2}|\vec{M}_0| \right) n_\uparrow(\vec{k}) + \left(\epsilon(\vec{k}) - \frac{1}{2}|\vec{M}_0| \right) n_\downarrow(\vec{k}) \right). \tag{3.1.20}$$

The result is that if $|\vec{M}_0|$ is non-zero, we can lower the *electronic* energy by filling up a number of down spin states and, at the same time, emptying the same number of up spin states. Since the first term penalizes a non-zero value of $|\vec{M}_0|$ we must search for a balance. We also need to keep track of the fact that there is a *total* of N electrons (both with up and down spins). As usual, this is taken care of by shifting $H_{\rm MF}$ by $\mu\hat{N} = \mu\sum_{\vec{r}} c^\dagger_\sigma(\vec{r}\,)c_\sigma(\vec{r}\,)$.

Consider now a state with $N_\uparrow$ electrons with spin up and $N_\downarrow$ with spin down. Let $\epsilon_\uparrow(\epsilon_\downarrow)$ be the one-particle energy of the top of the filled up(down) states. The total energy of such a state E is a function of $|\vec{M}_0|, \mu, \epsilon_\uparrow$ and $\epsilon_\downarrow$ (or, equivalently $N_\uparrow$ and $N_\downarrow$).

The energy is $\left(|\vec{M}_0| \equiv M_0\right)$

$$\begin{aligned} E_0(M_0,\mu,\epsilon_\uparrow,\epsilon_\downarrow) =& \frac{3}{8U} V M_0^2 + V \int_{\vec{k}} \left(\epsilon(\vec{k}) + \frac{1}{2}M_0\right) \theta\left(\epsilon_\uparrow - \epsilon(\vec{k})\right) + \\ &+ V \int_{\vec{k}} \left(\epsilon(\vec{k}) - \frac{1}{2}M_0\right) \theta\left(\epsilon_\downarrow - \epsilon(\vec{k})\right) + \\ &+ \mu V \int_{\vec{k}} \left(\theta\left(\epsilon_\uparrow - \epsilon(\vec{k})\right) + \theta\left(\epsilon_\downarrow - \epsilon(\vec{k})\right)\right). \end{aligned} \tag{3.1.21}$$

By introducing the one-particle *band* density of state (DOS) (i.e. the DOS of the unperturbed system without spin), $\rho(\epsilon)$, we get

$$\begin{aligned} \mathcal{E} \equiv& \frac{E_0}{V} \\ =& \frac{3}{8U} M_0^2 + \int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon (\epsilon + \frac{1}{2}M_0)\rho(\epsilon) + \int_{\epsilon_0}^{\epsilon_\downarrow} d\epsilon (\epsilon - \frac{1}{2}M_0)\rho(\epsilon) + \\ &+ \mu \left(\int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon \rho(\epsilon) + \int_{\epsilon_0}^{\epsilon_\downarrow} d\epsilon \rho(\epsilon)\right) \end{aligned} \tag{3.1.22}$$

where ϵ_0 is the energy of the bottom of the band.

Since the ground state energy must be an extremum (actually a minimum) we have to find the values of $\mu, M_0, \epsilon_\uparrow$ and $\epsilon_\downarrow$ which make the energy density have a minimum *at fixed density*. That is

$$\frac{\partial \mathcal{E}}{\partial \mu} = \frac{N}{V}, \qquad \frac{\partial \mathcal{E}}{\partial |\vec{M}_0|} = 0, \qquad \frac{\partial \mathcal{E}}{\partial \epsilon_\uparrow} = 0, \qquad \frac{\partial \mathcal{E}}{\partial \epsilon_\downarrow} = 0. \tag{3.1.23}$$

An explicit calculation gives

$$\begin{aligned} \frac{N}{V} &= \int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon \rho(\epsilon) + \int_{\epsilon_0}^{\epsilon_\downarrow} d\epsilon \rho(\epsilon), \\ 0 &= \frac{3}{4U} M_0 + \frac{1}{2} \int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon \rho(\epsilon) - \frac{1}{2} \int_{\epsilon_0}^{\epsilon_\downarrow} d\epsilon \rho(\epsilon), \\ 0 &= \left(\epsilon_\uparrow + \frac{1}{2}M_0\right) \rho(\epsilon_\uparrow) + \mu \rho(\epsilon_\uparrow), \\ 0 &= \left(\epsilon_\downarrow - \frac{1}{2}M_0\right) \rho(\epsilon_\downarrow) + \mu \rho(\epsilon_\downarrow). \end{aligned} \tag{3.1.24}$$

Provided that $\rho(\epsilon_{\uparrow,\downarrow}) \neq 0$, we see that the polarization M_0 is given by

$$M_0 = (\epsilon_\downarrow - \epsilon_\uparrow) \tag{3.1.25}$$

and the chemical potential μ is equal to

$$\mu = -\frac{1}{2}(\epsilon_\downarrow + \epsilon_\uparrow). \tag{3.1.26}$$

Clearly, since M_0 is positive, $\epsilon_\uparrow < \epsilon_\downarrow$ and there are more occupied down spin states than up spin states. We can also write

$$\epsilon_\downarrow - \epsilon_\uparrow = \frac{2U}{3} \int_{\epsilon_\uparrow}^{\epsilon_\downarrow} d\epsilon \rho(\epsilon) \tag{3.1.27}$$

and

$$\frac{N}{V} = 2 \int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon \rho(\epsilon) + \int_{\epsilon_\uparrow}^{\epsilon_\downarrow} d\epsilon \rho(\epsilon). \tag{3.1.28}$$

Equations (3.1.27) and (3.1.28) determine $\epsilon_\uparrow$ and $\epsilon_\downarrow$ and, thus, the solution to the problem. In general these equations need to be solved numerically.

Equation (3.1.27) has two solutions: $\epsilon_\uparrow = \epsilon_\downarrow$ (i.e. $M_0 = 0$, the *paramagnetic* state) and $\epsilon_\uparrow \neq \epsilon_\downarrow$ ($M_0 \neq 0$, the *ferromagnetic* state). The analysis of these equations follows closely the solution of the Curie-Weiss equation in the theory of phase transitions. We can write Eq. (3.1.27) in the form

$$x = \frac{2U}{3} \int_0^x d\epsilon \rho(\epsilon + \epsilon_\uparrow) \tag{3.1.29}$$

where $x = \epsilon_\downarrow - \epsilon_\uparrow$. Also, we get

$$\frac{N}{V} = 2 \int_{\epsilon_0}^{\epsilon_\uparrow} d\epsilon \rho(\epsilon) + \frac{3}{2U} x. \tag{3.1.30}$$

For $\epsilon_\uparrow$ given, the integral in Eq. (3.1.29) is a monotonically increasing function of x (see Fig. 3.3). For values of $U > U_c$ there are two solutions $x_0 = 0$ and $x_0 \neq 0$ whereas for $U < U_c$ there is only one solution, $x_0 = 0$. The critical Hubbard coupling U_c is determined by the condition

$$1 = \frac{2U_c}{3} \rho(\bar{\epsilon}_\uparrow) \tag{3.1.31}$$

where $\rho(\bar{\epsilon}_\uparrow)$ is determined by Eq. (3.1.30) at $x = 0$

$$\frac{N}{V} = 2 \int_{\epsilon_0}^{\bar{\epsilon}_\uparrow} d\epsilon \rho(\epsilon). \tag{3.1.32}$$

Equation (3.1.31) is known as the Stoner criterion. The statement is that for $U > U_c$ the *ferromagnetic* solution appears and has a lower energy than the *paramagnetic* state, $|\vec{M}_0| = 0$.

If the DOS is a *smooth* function near the Fermi energy of the paramagnetic state, we can find the solution close to U_c by using a power series expansion. The result, to leading order in $\frac{U_c - U}{U_c}$, is

$$x_0 = \epsilon_\downarrow - \bar{\epsilon}_\uparrow \approx \frac{2\bar{\rho}}{\bar{\rho}'} \left(\frac{U_c - U}{U_c} \right) + \ldots \tag{3.1.33}$$

and

$$\delta = \epsilon_\uparrow - \bar{\epsilon}_\uparrow \approx -\frac{3}{2U_c \bar{\rho}'} \left(\frac{U_c - U}{U_c} \right) + \ldots \tag{3.1.34}$$

where $\bar{\epsilon}_\uparrow$ satisfies Eq. (3.1.32) and $\bar{\rho}$ and $\bar{\rho}'$ are the DOS and its derivative at $\bar{\epsilon}_\uparrow$. There are a number of important cases in which the DOS $\rho(\epsilon)$ has

singularities at certain energies, the Van Hove singularities. This happens at half-filling for systems like a square lattice, for which the Fermi surface has the property of nesting.

3.1.2 Néel State

We now will look for solutions of the Mean-Field equations in which $\vec{M}(\vec{r}\,)$ is not a constant. Ultimately the problem boils down to a comparison of the energies for different solutions. However, for situations in which nesting takes place we can argue that a Néel state, or, more generally, a spin-density-wave (SDW) is the ground state.

Let us consider the Mean-Field Hamiltonian of Eq. (3.1.15) and assume that $M(\vec{r}\,)$ has the form

$$\vec{M}(\vec{r}\,) = \vec{M}_0 \cos(\vec{Q}\cdot\vec{r}) \tag{3.1.35}$$

where $\vec{Q} = (\pi, \pi)$. We saw before that, at half-filling, the Fermi surface has the shape shown in Fig. 2.4. The states across the Fermi surface differ by a wave vector $\vec{Q} = (\pi, \pi)$ which is at the Brillouin Zone (BZ) edge (this is the nesting property). Furthermore, for a square lattice, we have $\epsilon(\vec{k}) = -4t(\cos k_1 + \cos k_2)$. Thus we get

$$\epsilon(\vec{k}) = -\epsilon(\vec{k}+\vec{Q}). \tag{3.1.36}$$

The Mean-Field Hamiltonian can now be written in the form

$$\begin{aligned} H_{\mathrm{MF}} = &\frac{3}{8U}|\vec{M}_0|^2 V + V\int_{\vec{k}} \epsilon(\vec{k}) c^\dagger_\sigma(\vec{k}) c_\sigma(\vec{k}) + \\ &+ V\int_{\vec{k}} c^\dagger_\alpha(\vec{k}) \frac{1}{2}\vec{\tau}_{\alpha\beta} c_\beta(\vec{k}+\vec{Q})\cdot\frac{1}{2}\vec{M}_0 + \\ &+ V\int_{\vec{k}} c^\dagger_\alpha(\vec{k}) \frac{1}{2}\vec{\tau}_{\alpha\beta} c_\beta(\vec{k}-\vec{Q})\cdot\frac{1}{2}\vec{M}_0. \end{aligned} \tag{3.1.37}$$

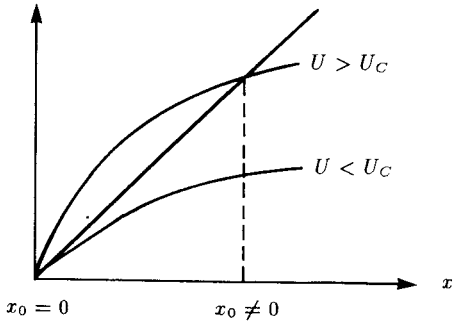


Figure 3.3 Solution of the Mean-Field equation.

Consider the spinor $\Psi_\sigma(\vec{k})$

$$\Psi_\sigma(\vec{k}) = \begin{pmatrix} c_\sigma(\vec{k}) \\ c_\sigma(\vec{k}+\vec{Q}) \end{pmatrix}. \tag{3.1.38}$$

If we restrict ourselves to the two dimensional case we can now write

$$H_{\rm MF} = \frac{3}{8U}|\vec{M}_0|^2 V + V \int_{\vec{k}}' \Psi^\dagger_{a\sigma}(\vec{k}) \mathcal{H}_{a\sigma,a'\sigma'}(\vec{k}) \Psi_{a'\sigma'}(\vec{k}) \tag{3.1.39}$$

where the integral now ranges over the upper half of the Brillouin Zone (it has been "folded") and $a = 1, 2, \sigma = \uparrow, \downarrow$, indicating the upper and lower components of Ψ with either spin. The one-particle Hamiltonian $\mathcal{H}$ is a 4×4 matrix which has the block form

$$\mathcal{H} = \begin{pmatrix} \epsilon(\vec{k}) & \frac{1}{2}\vec{\tau}\cdot\vec{M}_0 \\ \frac{1}{2}\vec{\tau}\cdot\vec{M}_0 & \epsilon(\vec{k}+\vec{Q}) \end{pmatrix} = \begin{pmatrix} \epsilon(\vec{k}) & \frac{1}{2}\vec{\tau}\cdot\vec{M}_0 \\ \frac{1}{2}\vec{\tau}\cdot\vec{M}_0 & -\epsilon(\vec{k}) \end{pmatrix}. \tag{3.1.40}$$

This matrix can be diagonalized very easily (see, for instance, the diagonalization of the Dirac Hamiltonian). It has two doubly-degenerate (spin) eigenvalues $E_\pm(\vec{k})$ with

$$E_\pm = \pm\sqrt{\epsilon^2(\vec{k}) + \frac{1}{4}|\vec{M}_0|^2}. \tag{3.1.41}$$

Thus, we see that the system has now acquired a gap Δ with

$$\Delta = |\vec{M}_0| \tag{3.1.42}$$

at the "Fermi surface". The energy-momentum relations, say along a main diagonal of the Brillouin Zone varies as shown in Fig. 3.4.

The ground state is obtained by filling up the negative energy modes with both spins up and down. This state has a vanishing z-component of the total spin: $S_z = 0$. The energy density is

$$\mathcal{E} = \frac{3}{8U}|\vec{M}_0|^2 - \int_{\vec{k}}' E_+(\vec{k}) \tag{3.1.43}$$

where the integral ranges over the upper half of the Brillouin Zone. We can now use the symmetries of $E(\vec{k}) \equiv E_+(\vec{k})$ to write this expression in terms of an integral over the upper right quadrant of the Brillouin Zone

$$\mathcal{E} = \frac{3}{8U}|\vec{M}_0|^2 - 2\int_{0\le k_i\le\pi} \frac{d^2k}{(2\pi)^2} E(\vec{k}). \tag{3.1.44}$$

We must now determine the value of M_0 for which $\mathcal{E}$ is lowest. The condition for an extremum is

$$\frac{\partial\mathcal{E}}{\partial M_0} = \frac{3}{4U}|\vec{M}_0| - 2\int_{0\le k_i\le\pi} \frac{d^2k}{(2\pi)^2} \frac{\partial E(\vec{k})}{\partial M_0} = 0. \tag{3.1.45}$$

Clearly we get

$$\frac{3}{4U}|\vec{M}_0| - \frac{1}{2}\int_{0\le k_i\le\pi} \frac{d^2k}{(2\pi)^2} \frac{|\vec{M}_0|}{E(\vec{k})} = 0. \tag{3.1.46}$$

One solution is $|\vec{M}_0| = 0$ (i.e. no long-range order). This is the paramagnetic state. The other solution obeys

$$\frac{3}{2U} = \int_{0 \le k_i \le \pi} \frac{d^2k}{(2\pi)^2} \frac{1}{\sqrt{\epsilon^2(\vec{k}) + \frac{1}{4}|\vec{M}_0|^2}}. \tag{3.1.47}$$

In the case of the square lattice (this result is true for any other case with nesting) the integral on the right hand side of Eq. (3.1.47) is *logarithmically divergent* (for $M_0 \to 0$). In fact, the integral is dominated by contributions with momenta around $k_1 + k_2 = \pi$. With this intuition in mind, we use

$$\cos k_1 + \cos k_2 = 2 \cos \frac{k_1 + k_2}{2} \cos \frac{k_1 - k_2}{2} \tag{3.1.48}$$

to write $E^2(\vec{k})$ in the form

$$E^2(\vec{k}) = 16t^2 \cos^2 \frac{(k_1 + k_2)}{2} \cos^2 \frac{(k_1 - k_2)}{2} + \frac{1}{4}|\vec{M}_0|^2. \tag{3.1.49}$$

Now we make the approximation of neglecting the dependence in $k_1 - k_2$ (i.e. setting $\cos \frac{(k_1 - k_2)}{2} \approx 1$) and approximating $\cos \frac{(k_1 + k_2)}{2} \approx \frac{\pi - k_1 - k_2}{2}$. One then finds

$$E^2(k) \simeq 4t^2(\pi - k_1 - k_2)^2 + \frac{|\vec{M}_0|^2}{4}. \tag{3.1.50}$$

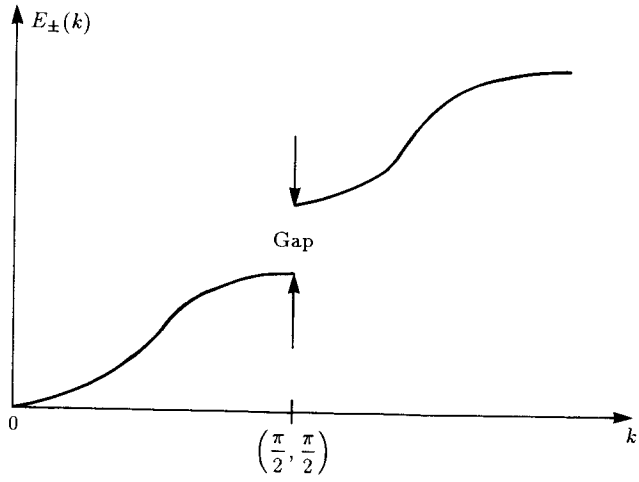


Figure 3.4 Energy-momentum relation $E_\pm(\vec{k})$ of the one-particle states around a Néel ground state. The horizontal axis scans the main diagonal of the Brillouin Zone for the square lattice.

Thus the integral in Eq. (3.1.47) can be evaluated. The result is

$$\frac{3}{2U} \simeq \frac{1}{2\pi}\left(\frac{1}{2t}\right)\sinh^{-1}\left(\frac{2tD}{|\vec{M}_0|}\right). \tag{3.1.51}$$

This expression is accurate in the weak coupling limit $\frac{U}{2t} \to 0$. Here D is a cutoff approximately equal to π. The solution is

$$|\vec{M}_0| \simeq \frac{(2t)D}{\sinh\left(3\pi(\frac{2t}{U})\right)} \approx (4tD)e^{-3\pi(\frac{2t}{U})}. \tag{3.1.52}$$

We conclude that the order parameter $\langle \vec{S}(\vec{r}\,)\rangle$ is also non-zero: $\langle \vec{S}(\vec{r}\,)\rangle = \frac{3}{8U}\vec{M}_0\cos(\vec{Q}\cdot\vec{r})$. Thus an antiferromagnetic solution is found even for arbitrarily weak Hubbard coupling U. Please keep in mind that the ferromagnetic solution requires a finite value of U to exist. It is also easy to see that this *Néel state* has less energy than the paramagnetic state $|\vec{M}_0| = 0$. Thus, at least at half-filling, the ground state appears to be a Néel antiferromagnet.

This solution is also remarkable for other reasons. First, the dependence of $|\vec{M}_0|$ on U is highly non-analytic: we get an essential singularity. This is exactly analogous to what one finds in BCS theory and in the case of the Peierls instability of one-dimensional electron-phonon systems. Secondly, the electronic spectrum has a *gap* Δ which is equal to $|\vec{M}_0|$. Thus the gap also has an essential singularity in the coupling constant. But, is the spectrum truly massive? Are there any gapless (or massless) excitations present. What this calculation says is that the one-particle spectrum is massive. What about the two-particle spectrum? We will see below that there are massless spin-waves in agreement with Goldstone's theorem.

3.2 Path Integral Representation of the Hubbard Model

So far we have discussed some features of these systems within a Mean-Field theory based on the canonical Hamiltonian formalism. It is possible to gain further insight by going to the path-integral form. Truly, both representations are equivalent and certainly whatever one can do in one form can be reproduced in the other picture. However, certain aspects of the problem can be dealt with in a more natural and concise way in the path-integral picture. The issue of the symmetry, and its breaking, effective theories for the low-lying modes, etc., are more answerable in path-integral form. And, of course, the semi-classical treatment, including non-perturbative features such as solitons and the like, are very simple to picture in terms of path-integrals.

Typically we are interested in studying both zero-temperature and finite-temperature properties of the system. At finite temperature, the equlibrium properties are determined by the partition function

$$Z = \mathrm{Tr}\; e^{-\beta H} \tag{3.2.1}$$

where $\beta = \frac{1}{T}$ (in units in which $k_B = 1$). Usually, one would also like to know the behavior of the correlation functions.

At zero temperature one is interested in the "vacuum persistence amplitude" [Coleman 85]

$$Z = \lim_{t \to \infty} \mathrm{Tr}\ e^{iHt} \tag{3.2.2}$$

which is just the trace of the evolution operator at long times. Feynman showed [Feynman 65] that Eq. (3.2.2) can be written as a sum over histories. Also, it is apparent that Eq. (3.2.1) is related to Eq. (3.2.2) by an analytic continuation procedure known as "Wick rotation"

$$it = -\tau \tag{3.2.3}$$

which amounts to going to imaginary time [Abrikosov 63]. I will use both forms more or less simultaneously.

I do not intend to give a thorough description of the path-integral method. Qualitatively, the path integral is derived as follows. Let H be the Hamiltonian of the system and $\{|\alpha\rangle\}$ be a set of states. In most cases we will demand that $\{|\alpha\rangle\}$ be a complete set of states. However, it will also be convenient to work with a system of *coherent states* which are overcomplete. In either case, what matters is the existence of an identity of the form ("resolution of unity")

$$1 = N \int d\alpha |\alpha\rangle\langle\alpha| \tag{3.2.4}$$

where N is a normalization constant and $d\alpha$ is an integration measure. It is worthwhile to comment that the states $\{|\alpha\rangle\}$ need not be position eigenstates. In the usual derivation ("sum over histories ") the position space (or coordinate) representation is used. On the other hand, in many problems, such as the quantization of spin systems, there isn't a natural separation between canonical coordinates and momenta. Thus the space of states $\{|\alpha\rangle\}$ can be quite general and abstract. In fact the coherent state representation is in some sense more primitive (or fundamental).

The standard strategy that one takes is the following. First one defines the states $\{|\alpha\rangle\}$. In the case of a many-body system such as the Hubbard model, these states should be antisymmetrized many-fermion states. I will work in the grand canonical ensemble and, hence, use second quantization. The need to antisymmetrize the states will bring some complications, which will be taken care of by using Grassmann variables.

The second step is to split up the time interval t into N_t segments of infinitesimal length Δ_t such that $N_t \Delta_t = t$. The same prescription applies to the imaginary time (Euclidean) formalism. The vacuum persistence amplitude is

$$Z = \mathrm{Tr}\ \hat{T} e^{i \int_{-\infty}^{+\infty} dt H(t)} \tag{3.2.5}$$

where $\hat{T}$ is the time ordering operator. For a time-independent Hamiltonian, Eq. (3.2.5) reduces to Eq. (3.2.2). For infinitesimal intervals $\Delta_t \to 0$ we can

write

$$Z = \mathrm{Tr}\ \hat{T} e^{i \sum_{j=1}^{N_t} \Delta_t H(t_j)} \approx \mathrm{Tr}\ \hat{T} \prod_{j=1}^{N_t} e^{i \Delta_t H(t_j)}. \tag{3.2.6}$$

Now we proceed by inserting the resolution of unity, Eq. (3.2.4), at each intermediate time t_j. Let $\{|\alpha_j\rangle\}$ be a set of states at each time t_j. We get

$$Z = \sum_{\{\alpha_j\}} \prod_{j=1}^{N_t} \langle \alpha_j | e^{-i \Delta t H(t_j)} | \alpha_{j+1} \rangle \tag{3.2.7}$$

with $|\alpha_{N+1}\rangle = |\alpha_1\rangle$ and where $|\alpha_j\rangle \equiv |\alpha(t_j)\rangle$ are the states at time t_j.

We can regard the $\alpha(t_j)$ as a set of parameters spanning a manifold defining the states $|\alpha_j\rangle$. Thus what we actually have is a sum over configurations $\{\alpha(t)\}$. Notice that this procedure is absolutely general. We are supposed to take the limit $\Delta_t \to 0, N_t \to \infty$ at the end. That this limit exists is a highly non-trivial issue and certainly not a formal matter. This procedure applies for both single-particle problems or states of a many-body second quantized system, i.e. a field theory.

Let us first review the particle-in-a-potential problem. In this case the Hamiltonian is

$$H = \frac{\hat{p}^2}{2m} + V(\hat{q}) \tag{3.2.8}$$

where $\hat{p}$ and $\hat{q}$ obey canonical commutation relations ($\hbar = 1$)

$$[\hat{q}, \hat{p}] = i. \tag{3.2.9}$$

Thus the states $|\alpha\rangle$ can be the complete set of position eigenstates $|q\rangle$. The resolution of the identity is

$$1 = \int dq |q\rangle\langle q|. \tag{3.2.10}$$

Conversely, we could also use momentum eigenstates. The momentum operator $\hat{p}$ is not diagonal in this basis. Thus the amplitude

$$\langle q(t_j) | e^{i \Delta_t H} | q(t_{j+1}) \rangle \tag{3.2.11}$$

can be written in the form

$$\langle q(t_j) | e^{i \Delta_t \frac{\hat{p}^2}{2m}} | q(t_{j+1}) \rangle\ e^{i \Delta_t V(q(t_{j+1}))} \tag{3.2.12}$$

where we have used the fact that $\hat{V}$ is diagonal in the coordinate representation. Now we use a complete set of momentum eigenstates $\{|p(t_j)\rangle\}$ and write

$$e^{i \Delta_t \frac{\hat{p}^2}{2m}} = \int_{-\infty}^{+\infty} \frac{dp}{2\pi}\ e^{i \Delta_t \frac{p^2(t_j)}{2m}} |p(t_j)\rangle\langle p(t_j)|. \tag{3.2.13}$$

Collecting these various contributions we get

$$Z = \int \mathcal{D}p \mathcal{D}q \prod_{j=1}^{N_t} e^{i \Delta_t \frac{p^2(t_j)}{2m} + i \Delta_t V(q(t_{j+1}))} \langle p(t_j) | q(t_{j+1}) \rangle \langle p(t_j) | q(t_j) \rangle^* \tag{3.2.14}$$

where I used the definition of the measure

$$\mathcal{D}p\mathcal{D}q = \prod_{j=1}^{N_t} \frac{dp(t_j)dq(t_j)}{2\pi}. \tag{3.2.15}$$

Thus, we can write Eq. (3.2.14) in the form

$$Z = \int \mathcal{D}p\mathcal{D}q \; e^{-i\int dt[p\dot{q}-H(p,q)]} \tag{3.2.16}$$

by making use of the fact that

$$\langle p|q\rangle = e^{-ip\cdot q}. \tag{3.2.17}$$

Equation (3.3.16) is nothing but a sum over the configurations in the *phase space* of the action S

$$S = \int dt(p\dot{q} - H) \tag{3.2.18}$$

of each configuration. Since we are computing a *trace* the *field* $q(t)$ obeys periodic boundary conditions in time. Note that $\hat{p}$ and $\hat{q}$ do not commute. The phase space integral is actually a coherent state path integral [Faddeev 75]. Equation (3.2.16) is generally valid even for Hamiltonians for which it is not possible to clearly separate coordinates and momenta. I will adopt the phase-space (or coherent state) path integral as the definition.

This procedure can be trivially generalized to second quantized systems. In the case of bosons we have second quantized field operators $\hat{\Psi}(\vec{r}\,)$ and $\hat{\Psi}^\dagger(\vec{r}\,)$ and a Hamiltonian H. The field operators obey the equal-time commutation relations

$$\left[\hat{\Psi}(\vec{r}\,), \hat{\Psi}^\dagger(\vec{r}\,')\right] = \delta(\vec{r} - \vec{r}\,'). \tag{3.2.19}$$

Consider the classical Lagrangian L

$$L = \sum_{\vec{r}} \Psi^* i\partial_t \Psi - H. \tag{3.2.20}$$

The commutation relations in Eq. (3.2.19) follow from canonically quantizing L. The canonical momentum $\hat{\Pi}(\vec{r}\,)$ is given by

$$\hat{\Pi}(\vec{r}\,) \equiv \frac{\delta L}{\delta \partial_t \hat{\Psi}(\vec{r}\,)} \equiv i\hat{\Psi}^\dagger(\vec{r}\,). \tag{3.2.21}$$

Thus the canonical commutation relations

$$\left[\hat{\Psi}(\vec{r}\,), \hat{\Pi}(\vec{r}\,')\right] = i\delta(\vec{r} - \vec{r}\,') \tag{3.2.22}$$

are equivalent to Eq. (3.2.19) after $\hat{\Pi}$ is identified with $i\hat{\Psi}^\dagger$.

A discussion analogous to what we did for the particle case yields a phase space path integral

$$Z = \int \mathcal{D}\Psi^*\mathcal{D}\Psi \; e^{i\int dt[\sum_{\vec{r}} \Psi^* i\partial_t \Psi - H(\Psi^*,\Psi)]} \tag{3.2.23}$$

where Ψ and Ψ^* are complex c-number fields which parametrize the coherent states [Faddeev 75]. Since we are dealing with bosons, the fields Ψ are c-numbers and commute. The boundary conditions in turn are periodic. The case of fermions can also be dealt with provided that one takes care of the anticommuting nature of fermion operators.

It is convenient to introduce coherent states for fermions. Let $\hat{\Psi}^\dagger$ and $\hat{\Psi}$ be Fermi creation and annihilation operators which satisfy

$$\{\hat{\Psi}, \hat{\Psi}^\dagger\} = 1 \tag{3.2.24}$$

and

$$\hat{\Psi}^2 = \hat{\Psi}^{\dagger 2} = 0. \tag{3.2.25}$$

In the occupation number representation we have two states, $|0\rangle$ and $|1\rangle$ with the properties:

$$\begin{aligned} &\hat{\Psi}|0\rangle = 0, \quad \hat{\Psi}^\dagger|0\rangle = |1\rangle, \quad \hat{\Psi}^\dagger\hat{\Psi}|0\rangle = 0, \\ &\hat{\Psi}^\dagger|1\rangle = 0, \quad \hat{\Psi}|1\rangle = |0\rangle, \quad \hat{\Psi}^\dagger\hat{\Psi}|1\rangle = |1\rangle. \end{aligned} \tag{3.2.26}$$

We introduce the two Grassmann numbers Ψ and $\bar{\Psi}$ that we will associate to the Fermi operators $\hat{\Psi}$ and $\hat{\Psi}^\dagger$. Their defining property is the (Grassmann) Algebra

$$\{\Psi, \Psi\} = \{\bar{\Psi}, \bar{\Psi}\} = \{\Psi, \bar{\Psi}\} = 0 \tag{3.2.27}$$

that they satisfy. It is natural to extend these anticommutation relations by imposing [Negele 88]

$$\{\Psi, \hat{\Psi}\} = 0 \qquad \text{and} \qquad (\Psi\hat{\Psi})^\dagger = \hat{\Psi}^\dagger\bar{\Psi}. \tag{3.2.28}$$

We define the coherent state $|\Psi\rangle$ in terms of the Grassmann number Ψ

$$|\Psi\rangle \equiv |0\rangle - \Psi|1\rangle = |0\rangle - \Psi\hat{\Psi}^\dagger|0\rangle. \tag{3.2.29}$$

With the help of Eqs. (3.2.27) and (3.2.28) we obtain

$$|\Psi\rangle = \left(1 - \Psi\hat{\Psi}^\dagger\right)|0\rangle \equiv e^{-\Psi\hat{\Psi}^\dagger}|0\rangle, \tag{3.2.30}$$

$$\hat{\Psi}|\Psi\rangle = \hat{\Psi}|0\rangle - \hat{\Psi}\Psi\hat{\Psi}^\dagger|0\rangle = \Psi|0\rangle = \Psi|\Psi\rangle \tag{3.2.31}$$

and

$$\hat{\Psi}^\dagger|\Psi\rangle = \hat{\Psi}^\dagger|0\rangle - \hat{\Psi}^\dagger\Psi\hat{\Psi}^\dagger|0\rangle = |1\rangle \equiv -\frac{\delta}{\delta\Psi}|\Psi\rangle. \tag{3.2.32}$$

As usual the exponential on the right hand side of Eq. (3.2.30) is defined as a power expansion in $\Psi\hat{\Psi}^\dagger$ and Eq. (3.2.32) *defines* operationaly the "left" Grassmann derivative.

The adjoint coherent state $\langle\Psi|$ is defined in terms of the Grassmann number Ψ

$$\langle\Psi| \equiv \langle 0| - \langle 1|\bar{\Psi} = \langle 0| - \langle 0|\hat{\Psi}\bar{\Psi}. \tag{3.2.33}$$

As before

$$\langle\Psi| = \langle 0|\left(1 - \hat{\Psi}\bar{\Psi}\right) \equiv \langle 0|e^{-\hat{\Psi}\bar{\Psi}}, \tag{3.2.34}$$

$$\langle\Psi|\hat{\Psi}^\dagger = \langle 0|\hat{\Psi}^\dagger - \langle 0|\hat{\Psi}\bar{\Psi}\hat{\Psi}^\dagger = \langle 0|\bar{\Psi} = \langle\Psi|\bar{\Psi} \tag{3.2.35}$$

and

$$\langle\Psi|\hat{\Psi} = \langle 0|\hat{\Psi} - \langle 0|\hat{\Psi}\bar{\Psi}\hat{\Psi} = \langle 1| \equiv -\langle\Psi|\frac{\delta}{\delta\bar{\Psi}}. \tag{3.2.36}$$

The right hand side of Eq. (3.2.36) defines the "right" Grassmann derivative in complete analogy to Eq. (3.2.32). It is natural to require that the "left" ("right") derivative $\frac{\delta}{\delta\Psi}$ anticommutes with Ψ so that $\frac{\delta}{\delta\Psi}\Psi = -\Psi\frac{\delta}{\delta\Psi}$.

From Eqs. (3.2.34), (3.2.30) and (3.2.28) the inner product $\langle\Psi|\Psi'\rangle$ is equal to

$$\langle\Psi|\Psi'\rangle = e^{\bar{\Psi}\Psi'}. \tag{3.2.37}$$

This together with Eqs. (3.2.31) and (3.2.35) gives for the matrix elements of a normal ordered operator $:\mathcal{U}\left(\hat{\Psi}^\dagger, \hat{\Psi}\right):$

$$\langle\Psi| :\mathcal{U}\left(\hat{\Psi}^\dagger, \hat{\Psi}\right): |\Psi'\rangle = \mathcal{U}(\bar{\Psi}, \Psi')e^{\bar{\Psi}\Psi'} \tag{3.2.38}$$

where $\mathcal{U}(\bar{\Psi}, \Psi')$ is obtained by replacing, inside the normal ordered operator $:\mathcal{U}:$, $\hat{\Psi} \to \Psi$ and $\hat{\Psi}^\dagger \to \bar{\Psi}$.

The resolution of unity in this representation is just

$$1 = \int d\bar{\Psi}d\Psi\ e^{-\bar{\Psi}\Psi}|\Psi\rangle\langle\Psi|. \tag{3.2.39}$$

This identity can be checked by computing the inner product $\langle\Psi'|\Psi''\rangle$ where Ψ' and Ψ'' are the Grassmann variables Ψ or $\bar{\Psi}$. The integrals in Eq. (3.2.39) are understood to be linear functionals on the space of functions of the Grassmann variables with

$$\int d\Psi\Psi \equiv 1, \qquad \int d\Psi \equiv 0. \tag{3.2.40}$$

We can now repeat the procedure outlined at the begining of this section for a second quantized system of fermions except that now we will use fermion coherent states $|\{\Psi_\sigma(\vec{r}\,)\}\rangle$ at each site and for each spin degrees of freedom

$$|\{\Psi_\sigma(\vec{r}\,)\}\rangle = \exp\left(-\sum_{\vec{r},\sigma}\Psi_\sigma(\vec{r}\,)\hat{\Psi}^\dagger_\sigma(\vec{r}\,)\right)|0\rangle \tag{3.2.41}$$

where $|0\rangle$ is the empty state (not quite the "vacuum" as we will see in chapter 4). Following our noses we find

$$Z = \lim_{\substack{N_t\to\infty \\ \Delta_t\to 0}}\left(\prod_{j=1}^{N_t}\int d\bar{\Psi}(t_j)d\Psi(t_j)\right)\prod_{j=1}^{N_t} e^{-\bar{\Psi}(t_j)\Psi(t_j)}\langle\Psi(t_j)|\left(1 - i\Delta_t H\right)|\Psi(t_{j+1})\rangle \tag{3.2.42}$$

where, for the sake of simplicity, I have dropped the space and spin labels. In the limit $N_t \to \infty$ and $\Delta_t \to 0$, one finds

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi\ e^{i\int dtL} \tag{3.2.43}$$

where L is given by

$$L = \sum_{\vec{r}} \bar{\Psi}_\sigma(\vec{r}\,) i\partial_t \Psi_\sigma(\vec{r}\,) - H(\bar{\Psi}_\sigma(\vec{r}\,), \Psi_\sigma(\vec{r}\,)). \tag{3.2.44}$$

For the case of the Hubbard model or any other one, for that matter, and in the presence of a non-zero chemical potential μ, we get (see Eq. (2.2.9))

$$L = \sum_{\vec{r}} \bar{\Psi}_\sigma(\vec{r}\,)(i\partial_t + \mu)\Psi_\sigma(\vec{r}\,) + 2t \sum_{\langle \vec{r}, \vec{r}' \rangle} \bar{\Psi}_\sigma(\vec{r}\,)\Psi_\sigma(\vec{r}\,') - H_{\rm int}(\bar{\Psi}, \Psi). \tag{3.2.45}$$

From Eq. (2.2.8) the interaction for the Hubbard model is

$$H_{\rm int} = -\frac{U}{6} \sum_{\vec{r}} c^\dagger_\alpha(\vec{r}\,) \tau^a_{\alpha\beta} c_\beta(\vec{r}\,) c^\dagger_\gamma(\vec{r}\,) \tau^a_{\gamma\delta} c_\delta(\vec{r}\,). \tag{3.2.46}$$

Normal ordering relative to the empty state $|0\rangle$ gives

$$: H_{\rm int} := -\frac{U}{6} \sum_{\vec{r}} c^\dagger_\alpha(\vec{r}\,) c^\dagger_\gamma(\vec{r}\,) c_\delta(\vec{r}\,) c_\beta(\vec{r}\,) \tau^a_{\alpha\beta} \tau^a_{\gamma\delta} - \frac{U}{2} \sum_{\vec{r}} c^\dagger_\alpha(\vec{r}\,) c_\alpha(\vec{r}\,). \tag{3.2.47}$$

Thus $H_{\rm int}(\bar{\Psi}, \Psi)$ is given by

$$H_{\rm int}(\bar{\Psi}, \Psi) = -\frac{U}{6} \sum_{\vec{r}} \bar{\Psi}_\alpha(\vec{r}\,) \bar{\Psi}_\gamma(\vec{r}\,) \Psi_\delta(\vec{r}\,) \Psi_\beta(\vec{r}\,) \tau^a_{\alpha\beta} \tau^a_{\gamma\delta} - \frac{U}{2} \sum_{\vec{r}} \bar{\Psi}_\alpha(\vec{r}\,) \Psi_\alpha(\vec{r}\,). \tag{3.2.48}$$

The last term on the right hand side can obviously be canceled by means of a shift of the chemical potential μ.

The final property of Grassmann integrals which will be useful for us is the integral for actions which are quadratic in the fields

$$S = \sum_{\vec{r}, \vec{r}\,'} \bar{\Psi}(\vec{r}\,) M(\vec{r}, \vec{r}\,') \Psi(\vec{r}\,') \tag{3.2.49}$$

where $M(\vec{r}, \vec{r}\,')$ is an antisymmetric matrix (operator). We get the gaussian integral

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \; e^{-\int \bar{\Psi} M \Psi} = {\rm Det} M. \tag{3.2.50}$$

This expression should be contrasted with the analogous result for bosonic fields ϕ

$$Z = \int \mathcal{D}\phi^* \mathcal{D}\phi \; e^{-\int \phi^* M \phi} = ({\rm Det} M)^{-1}. \tag{3.2.51}$$

Both results can be derived quite easily by expanding Ψ and $\bar{\Psi}$ in a basis of eigenstates of M [Faddeev 75] or [Negele 88].

3.3 Path-Integral Picture of the Mean-Field Theory of the Hubbard Model

We now turn to the Mean-Field theory for the Hubbard model in path-integral form. The advantage of this description is that we will be able to extract an effective field theory for the low-lying modes in the Néel state: spin waves.

The Lagrangian density for the Hubbard model in two dimensions, in real time and at zero temperature is, from Eqs. (3.2.45) and (3.2.48),

$$\begin{aligned}\mathcal{L} =& \bar{\Psi}_\alpha(\vec{r},t)(i\partial_t+\mu)\Psi_\alpha(\vec{r},t)+ \\ &+t\sum_{j=1,2}\left(\bar{\Psi}_\alpha(\vec{r},t)\Psi_\alpha(\vec{r}+\vec{e}_j,t)+\bar{\Psi}_\alpha(\vec{r},t)\Psi_\alpha(\vec{r}-\vec{e}_j,t)+\text{ c.c.}\right)+ \\ &+\frac{U}{6}(\bar{\Psi}_\alpha(\vec{r},t)\vec{\tau}_{\alpha\beta}\Psi_\beta(\vec{r},t))^2.\end{aligned} \tag{3.3.1}$$

The associated path-integral contains quartic terms, the interaction, and hence we do not know how to compute the partition function. The strategy is to write another theory which is quadratic in Grassmann fields and which is equivalent to Eq. (3.3.1). We will make extensive use of the gaussian identity for bosonic fields (or Hubbard-Stratonovich transformation)

$$\int d\vec{\phi}\; e^{-i\left(\frac{1}{2}\vec{\phi}^2+\lambda\vec{\phi}\cdot\bar{\Psi}\vec{\tau}\Psi\right)} = \text{const}\times e^{i\frac{1}{2}\lambda^2(\bar{\Psi}\vec{\tau}\Psi)^2}. \tag{3.3.2}$$

Thus at any point in space time $(\vec{r},t)$, we introduce a three-component real bose field $\vec{\phi}(\vec{r},t)$ coupled bilinearly to the fermions as in Eq. (3.3.2). If one chooses the coupling constant λ to be equal to $\sqrt{\frac{U}{3}}$, one finds the interaction term of Eq. (3.2.48). Thus the Lagrangian density $\mathcal{L}'$

$$\begin{aligned}\mathcal{L}' =& \bar{\Psi}_\alpha(\vec{r},t)(i\partial_t+\mu)\Psi_\alpha(\vec{r},t)+ \\ &+t\sum_{j=1,2}\left[\bar{\Psi}_\alpha(\vec{r},t)\Psi_\alpha(\vec{r}+\vec{e}_j,t)+\bar{\Psi}_\alpha(\vec{r},t)\Psi_\alpha(\vec{r}-\vec{e}_j,t)+\text{ c.c. }\right]+ \\ &-\sqrt{\frac{U}{3}}\vec{\phi}(\vec{r},t)\cdot\bar{\Psi}_\alpha(\vec{r},t)\vec{\tau}_{\alpha\beta}\Psi_\beta(\vec{r},t)-\frac{1}{2}\vec{\phi}^2(\vec{r},t)\end{aligned} \tag{3.3.3}$$

is equivalent to the Lagrangian of the Hubbard model. Equation (3.3.3) has the advantage of being bilinear in Fermi fields (compare Eqs. (3.3.3) and (3.1.3)). Thus, using Eq. (3.2.50) we can now integrate out the fermions. The result is an *effective action* for the bose fields $\vec{\phi}$. We will see that the $\vec{\phi}$ fields represent the collective modes associated with spin fluctuations. The result is

$$Z=\int \mathcal{D}\vec{\phi}\; e^{iS_{\text{eff}}(\vec{\phi})} \tag{3.3.4}$$

where the effective action $S_{\rm eff}(\vec{\phi})$ is given by

$$S_{\rm eff}(\vec{\phi}) = -\int dt \sum_{\vec{r}} \frac{1}{2}\vec{\phi}^{\,2}(\vec{r},t) - i \ln \mathrm{Det}\left(i\partial_t + \mu - M(\vec{\phi})\right). \tag{3.3.5}$$

The operator $M(\vec{\phi})$ in Eq.(3.3.5) has the matrix elements

$$\begin{aligned}\langle \vec{r}t\alpha|M(\vec{\phi})|\vec{r}\,'t'\beta\rangle = &-\delta_{\alpha\beta}\delta(t-t')2t\sum_{j=1,2}\left(\delta_{\vec{r}\,',\vec{r}+\vec{e}_j}+\delta_{\vec{r}\,',\vec{r}-\vec{e}_j}\right)+\\ &+\sqrt{\frac{U}{3}}\delta(t-t')\delta_{\vec{r},\vec{r}\,'}\vec{\phi}(\vec{r},t)\cdot\vec{\tau}_{\alpha\beta}.\end{aligned} \tag{3.3.6}$$

The Mean-Field theory for this problem is just the evaluation of the path-integral Eq. (3.3.4) by means of the saddle-point (or stationary phase) approximation. For this problem, this approximation is equivalent to a Hartree-Fock decoupling. The stationary condition is

$$0 = \frac{\delta S_{\rm eff}}{\delta\phi^a(\vec{r},t)} = -\phi^a(\vec{r},t) - i\frac{\delta}{\delta\phi^a(\vec{r},t)}\ln \mathrm{Det}(i\partial_t + \mu - M(\vec{\phi})). \tag{3.3.7}$$

Using the identity

$$\ln \mathrm{Det}A = \mathrm{Tr}\ \ln A$$

and Eq. (3.3.6), one finds

$$\begin{aligned}\phi^a(\vec{r},t) &= -i\frac{\delta}{\delta\phi^a(\vec{r},t)}\mathrm{Tr}\ \left(\ln[i\partial_t + \mu - M(\vec{\phi})]\right)\\ &= +i\mathrm{Tr}\left(\frac{1}{i\partial_t+\mu-M(\vec{\phi})}\frac{\delta M(\vec{\phi})}{\delta\phi^a(\vec{r},t)}\right)\\ &= i\sqrt{\frac{U}{3}}\langle\vec{r}t\alpha|\frac{1}{i\partial_t+\mu-M(\vec{\phi})}|\vec{r}t\beta\rangle\tau^a_{\beta\alpha}.\end{aligned} \tag{3.3.8}$$

The expression in angular brackets is just the (in space and time diagonal) matrix element of the fermion one-particle Green's function in a background field $\vec{\phi}(\vec{r},t)$

$$G_{\alpha\beta}(\vec{r}t;\vec{r}\,'t';\phi) \equiv -i\langle\vec{r}t\alpha|\frac{1}{i\partial_t+\mu-M(\phi)}|\vec{r}\,'t'\beta\rangle. \tag{3.3.9}$$

Hence we can write Eq. (3.3.9) in the form

$$\phi^a(\vec{r},t) = -\sqrt{\frac{U}{3}}G_{\alpha\beta}(\vec{r}t;\vec{r}t;\phi)\tau^a_{\beta\alpha}. \tag{3.3.10}$$

On the other hand, the local magnetic moment $\langle S^a(\vec{r}t)\rangle$ is equal to

$$\langle S^a(\vec{r}t)\rangle = +G_{\alpha\beta}(\vec{r}t;\vec{r}t;\phi)\frac{\tau^a_{\beta\alpha}}{2}. \tag{3.3.11}$$

Thus, the saddle-point approximation, Eq. (3.3.7), is the same as the Hartree-Fock condition

$$\phi^a(\vec{r},t) = -\sqrt{\frac{4U}{3}}\langle S^a(\vec{r},t)\rangle. \tag{3.3.12}$$

At the level of a one-band Hubbard model, there is no quantitative justification for the validity of this approach, since there is no small parameter other than $\hbar$ to control this expansion. Thus, this is essentially a semi-classical approximation.

We also know that an angular momentum degree of freedom, such as spin itself, becomes semi-classical if the angular momentum becomes large. The main assumption of the Mean-Field-Theory is that the order parameter thus obtained, in this case the staggered magnetization, is close to its saturation value and hence it is large.

We can formally introduce a small parameter to control this expansion by means of the following device. Let us imagine that the band electrons have an orbital degeneracy labelled by an index $a = 1, ..., N_b$, where N_b is the number of degenerate bands. The total band spin at a given site $\vec{r}$ is now given by $(i = 1, 2, 3)$

$$S^i(\vec{r}\,) = \sum_{\alpha,\beta=\uparrow,\downarrow} \sum_{a=1}^{N_b} \Psi^\dagger_{\alpha,a}(\vec{r}\,)\tau^i_{\alpha\beta}\Psi_{\beta,a}(\vec{r}\,). \tag{3.3.13}$$

The generalized Hubbard model is then given by the Hamiltonian

$$H = -\sum_{\substack{(\vec{r},\vec{r}\,')\\ \alpha,a}} \Psi^\dagger_{\alpha,a}(\vec{r}\,)t(\vec{r},\vec{r}\,')\Psi_{\alpha,a}(\vec{r}\,') + \text{ c.c. } - \frac{2}{3}U\sum_{\vec{r}} \left(\vec{S}(\vec{r}\,)\right)^2, \tag{3.3.14}$$

where $\vec{S}(\vec{r}\,)$ is the total band spin at $\vec{r}$. This system still has the global SU(2) invariance of spin rotations. For large values of U, i.e. $\frac{U}{t} \to \infty$, the local spin gets to be as large as possible. The equivalent Heisenberg model has a total spin quantum number s at each site equal to $s = \frac{N_b}{2}$ or equivalently, $N_b = 2s$. The limit $N_b \to \infty$ is then the same as the semi-classical limit $s \to \infty$. This limit is usually treated by spin-wave theory [Bloch 30], [Holstein 40], [Dyson 55], [Maleev 57] (for a review see the book of Mattis [Mattis 65]).

The path-integral approach is particularly well suited to deal with this limit. As a matter of fact, all the formulas derived above carry over to this case. The Hubbard-Stratonovich transformation works with the only change that $\vec{\phi}$ couples now to the total band spin. Since all N_b orbital species couple exactly in the same way to the Hubbard-Stratonovich field $\vec{\phi}$, the only change that occurs is that the fermion determinant *factorizes* and is given by the N_b-th power of the determinant of a single species. After a trivial rescaling of the field $\vec{\phi}$ by $\sqrt{N_b}$, the effective action $S^{N_b}_{\text{eff}}$ for the theory with orbital degeneracy is simply given by

$$S^{N_b}_{\text{eff}}(\vec{\phi}) = N_b S_{\text{eff}}(\vec{\phi}). \tag{3.3.15}$$

In the large N_b-limit (i.e. large s-limit), the saddle-point approximation becomes exact. For the rest of this section, we will carry on with this expansion assuming that it is valid. We should keep in mind, however, that the results will only become accurate in the $s \to \infty$ limit.

It is apparent that if $\phi^a(\vec{r},t)$ is a solution, any *uniform* rotation of it is also a solution

$$\phi'_a = R_{ab}\phi_b \tag{3.3.16}$$

where R_{ab} is a constant rotation matrix. This implies that the global spin symmetry has been preserved. We will see now that this implies the existence of Goldstone modes, spin-waves, if this symmetry is spontaneously broken.

Let us consider the half-filled case. Here, we expect an antiferromagnetic Néel state. The classical solution is (for the case of a square lattice)

$$\phi^a(\vec{r},t) = |\vec{\phi}|n^a(-1)^{x_1+x_2}. \tag{3.3.17}$$

This solution is (a) static, and (b) staggered (i.e. a Néel state). It really represents an infinite number of solutions parameterized by the unit vector $\vec{n}$ (in spin space). The *amplitude* $|\vec{\phi}|$ is determined by solving the saddle-point Eq. (3.3.12). In the notation of Eq. (3.1.9), we can write

$$\vec{M} = \sqrt{\frac{4U}{3}}\vec{\phi}. \tag{3.3.18}$$

In Section 3.1 we determined that the Néel state was energetically preferred to both a paramagnetic state and a ferromagnetic state. Notice that this argument does not rule out other solutions. However, the existent numerical evidence seems to indicate that a Néel state is the ground state at half-filling except for one-dimensional systems.

In Section 3.1 we showed that (a) the amplitude $|\vec{\phi}|$ is always non-zero at zero temperature and (b) the single-particle excitation spectrum has a gap $\Delta = |\vec{M}_0|$. This last result can be seen to follow by computing the one-particle Green function $G_{\alpha\beta}(\vec{r}t;\vec{r}\,'t';\phi)$ and writing it as a 4×4 matrix in spin and sublattice space.

Equation (3.3.9) is equivalent to the statement

$$\begin{aligned} -i\delta(t-t')\delta_{\vec{r},\vec{r}'}\delta_{\alpha\beta} =& (i\partial_t+\mu)G_{\alpha\beta}\left(\vec{r}t;\vec{r}\,'t';\vec{\phi}\right)+ \\ &+2t\sum_{j=1,2}[G_{\alpha\beta}(\vec{r}+\vec{e}_j t;\vec{r}\,'t';\vec{\phi})+G_{\alpha\beta}(\vec{r}-\vec{e}_j t;\vec{r}\,'t';\vec{\phi})]+ \\ &-\sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau}_{\alpha\gamma}(-1)^{x_1+x_2}G_{\gamma\beta}(\vec{r}t;\vec{r}\,'t';\vec{\phi}). \end{aligned} \tag{3.3.19}$$

If we Fourier transform Eq. (3.3.19) we find

$$-i\delta_{\alpha\beta} = (\omega-\epsilon(\vec{k}))G_{\alpha\beta}(\vec{k},\omega) - \sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau}_{\alpha\gamma}G_{\gamma\beta}(\vec{k}-\vec{Q},\omega) \tag{3.3.20}$$

where $\vec{Q}=(\pi,\pi)$ is the Brillouin corner vector. This equation can be solved in terms of the "spinor" operator $\mathcal{G}(\vec{k},\omega)$ defined componentwise by

$$\mathcal{G}_{\alpha\beta}(\vec{k},\omega) = \begin{pmatrix} G_{\alpha\beta}(\vec{k},\omega) \\ G_{\alpha\beta}(\vec{k}-\vec{Q},\omega) \end{pmatrix}. \tag{3.3.21}$$

Equation (3.3.20) now takes the matrix form

$$-i\delta_{\alpha\beta}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}(\omega-\epsilon(\vec{k}))\delta_{\alpha\gamma} & -\sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau}_{\alpha\gamma}\\ -\sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau}_{\alpha\gamma} & (\omega+\epsilon(\vec{k}))\delta_{\alpha\gamma}\end{pmatrix}\mathcal{G}_{\gamma\beta}(\vec{k},\omega). \quad (3.3.22)$$

The solution of this equation is

$$\mathcal{G}(\vec{k},\omega) = \frac{-i}{\omega^2-\left(\epsilon^2(\vec{k})+\frac{U}{3}|\vec{\phi}|^2\right)+i\nu}\begin{pmatrix}(\omega+\epsilon(\vec{k})) & \sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau}\\ \sqrt{\frac{U}{3}}|\vec{\phi}|\vec{n}\cdot\vec{\tau} & (\omega-\epsilon(\vec{k}))\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} \quad (3.3.23)$$

with $\nu \to 0^+$. This solution clearly shows the gap

$$\Delta = 2\sqrt{\epsilon^2(\vec{k})+\frac{U}{3}|\vec{\phi}|^2} = 2\sqrt{\epsilon^2(\vec{k})+\frac{|\vec{M}|^2}{4}} \quad (3.3.24)$$

in the excitation spectrum and we recover Eq. (3.1.42) as $\vec{k} \to \frac{\vec{Q}}{2}$.

3.4 Fluctuations Around the Néel State: The Non-Linear Sigma Model

In the previous section we obtained an effective action for the order-parameter field $\vec{\phi}$ and solved the saddle-point equations. We now wish to estimate the role and size of the quantum mechanical fluctuations about this classical Néel state.

When solving the saddle-point equations, we observed that if a non-trivial solution $\vec{\phi}_c$ with broken symmetry could be found, then any configuration obtained by means of a *rigid rotation* in spin space from $\vec{\phi}_c$ is also a solution. This reflects the fact that the spin sector has continuous symmetry group, in this case $O(3)$.

Imagine now not a solution of the saddle-point equation but a slowly varying configuration $\vec{\phi}(\vec{r},t)$ not far from a solution. The fluctuation part $\delta\vec{\phi}(\vec{r},t)$ is small and slowly varying. By slowly varying I mean slow on time scales being compared with $\tau = \frac{1}{\Delta}$ and on length scales being compared with $\xi = \frac{v_f}{\Delta}$ (where v_F is the Fermi velocity of the unperturbed system). This last length ξ is the (Mean-Field) correlation length of the system. It will turn out that ξ and τ determine the scales on which the *magnitude* $|\vec{\phi}|$ of the order parameter flcutuates, at least in Mean-Field theory.

The existence of an infinite number of solutions of the saddle-point equation indicates that there are configurations $\delta\vec{\phi}$ with arbitrarily low action. These are the Goldstone bosons of this problems and are spin-waves. We wish to find an effective theory for these spin waves.

Our first step will be to study the (gaussian) fluctuations around the Mean-Field solution. Thus, we will expand the effective action $S_{\text{eff}}(\phi)$ in powers of $\delta\vec{\phi}(\vec{r},t)$. Since $M(\vec{\phi})$ of Eq. (3.3.6) is linear in $\vec{\phi}$ we can write

$$M(\vec{\phi}) = M(\vec{\phi}_c) + M(\delta\vec{\phi}) \tag{3.4.1}$$

where the matrix elements of $M(\delta\vec{\phi})$ are equal to

$$\langle \vec{r}t\alpha|M(\delta\vec{\phi})|\vec{r}\,'t'\alpha'\rangle = \sqrt{\frac{U}{3}}\delta\vec{\phi}(\vec{r},t)\cdot\vec{\tau}_{\alpha\alpha'}\delta_{\vec{r},\vec{r}'}\delta(t-t'). \tag{3.4.2}$$

By expanding in powers of $M(\delta\vec{\phi})$, we find that the effective action

$$S_{\text{eff}}(\vec{\phi}) = -\int dt\sum_{\vec{r}}\frac{\vec{\phi}^2(\vec{r},t)}{2} - i\text{Tr}\ \ln\left(i\partial_t + \mu - M(\vec{\phi})\right) \tag{3.4.3}$$

can be written in the form

$$\begin{aligned} S_{\text{eff}}(\vec{\phi}) = &-\int dt\sum_{\vec{r}}\frac{\vec{\phi}^2(\vec{r},t)}{2} - i\text{Tr}\ \ln\left(i\partial_t + \mu - M(\vec{\phi}_c)\right) + \\ &- i\text{Tr}\ \ln\left(1 - iG(\vec{\phi}_c)M(\delta\vec{\phi})\right) \end{aligned} \tag{3.4.4}$$

with the *Mean-Field* Green function $G(\vec{\phi}_c)$ defined componentwise by

$$G_{\alpha\alpha'}(\vec{r}t;\vec{r}\,'t';\vec{\phi}_c) = -i\langle\vec{r}t\alpha|\frac{1}{i\partial_t + \mu - M(\vec{\phi}_c)}|\vec{r}\,'t'\alpha'\rangle. \tag{3.4.5}$$

By expanding the logarithm, we get

$$\begin{aligned} S_{\text{eff}}(\vec{\phi}) = &-\int dt\sum_{\vec{r}}\frac{\vec{\phi}^2(\vec{r},t)}{2} - i\text{Tr}\ \ln\left(i\partial_t + \mu - M(\vec{\phi}_c)\right) + \\ &+\sum_{n=1}^{\infty}\frac{i^{n+1}}{n}\text{Tr}\ \left(G(\vec{\phi}_c)M(\delta\vec{\phi})\right)^n. \end{aligned} \tag{3.4.6}$$

Equation (3.4.6) can be organized as follows

$$S_{\text{eff}}(\vec{\phi}) = \sum_{n=0}^{\infty} S^{(n)}(\vec{\phi}_c,\delta\vec{\phi}) \tag{3.4.7}$$

where the classical action (i.e. the action of the Mean-Field solution) $S^{(0)}(\vec{\phi}_c)$ is

$$S^{(0)}(\vec{\phi}_c) = -\int dt\sum_{\vec{r}}\frac{\vec{\phi}_c^2(\vec{r},t)}{2} - i\text{Tr}\ \ln[i\partial_t + \mu - M(\vec{\phi}_c)]. \tag{3.4.8}$$

Since the Mean-Field solution is *static* (i.e. time independent), we can write

$$S^{(0)}(\vec{\phi}_c) = TE^{(0)}_{\text{Gnd}}(\vec{\phi}_c) \tag{3.4.9}$$

where T is the time-span (not the temperature!) and $E^{(0)}_{\text{Gnd}}(\vec{\phi}_c)$ is the Mean-Field theory of ground state energy. The first order term in $\delta\vec{\phi}$ cancels out

since $\vec{\phi}_c$ is a solution of the saddle point equation

$$\frac{\delta S_{\text{eff}}}{\delta \vec{\phi}} = 0. \tag{3.4.10}$$

Thus we can write S_{eff} in the form

$$S_{\text{eff}}(\vec{\phi}) = TE^{(0)}_{\text{Gnd}}(\vec{\phi}_c) + \sum_{n=2}^{\infty} S^{(n)}(\vec{\phi}_c, \delta\vec{\phi}). \tag{3.4.11}$$

The gaussian theory, i.e. the quadratic terms in Eq. (3.4.11), has the action $S^{(2)}(\delta\vec{\phi})$ where

$$S^{(2)}(\delta\vec{\phi}) = -\int dt \sum_{\vec{r}} \frac{\delta\vec{\phi}^2(\vec{r},t)}{2} - \frac{i}{2}\text{Tr}\,\left(G(\vec{\phi}_c)M(\delta\vec{\phi})\right)^2. \tag{3.4.12}$$

This expression can be expanded out in components to yield

$$\begin{aligned} S^{(2)}(\delta\vec{\phi}) = &-\int dt \sum_{\vec{r}} \frac{\delta\vec{\phi}^2(\vec{r},t)}{2} + \\ &- i\frac{U}{6}\int dtdt' \sum_{\vec{r},\vec{r}'} \left(G_{\alpha,\alpha'}(\vec{r}t;\vec{r}\,'t';\vec{\phi}_c)\tau^{a'}_{\alpha'\beta}G_{\beta\beta'}(\vec{r}\,'t';\vec{r}t;\vec{\phi}_c)\tau^{a}_{\beta'\alpha}\right) \\ &\times \delta\phi^{a'}(\vec{r}\,',t')\delta\phi^{a}(\vec{r},t). \end{aligned} \tag{3.4.13}$$

We will see now that fluctuations $\delta\vec{\phi}(\vec{r},t)$ with wave-vector $\vec{p}$ *close* to the Brillouin Zone corner $\vec{Q} = (\pi,\pi)$ are gapless, i.e. have vanishingly small energy. These are the antiferromagnetic spin waves. Conversely, the excitations with $\vec{p}$ close to the zone center ($\vec{p} \approx 0$), which describe *uniform* ferromagnetic fluctuations, have *large* energies.

In Fourier components, Eq. (3.4.13) has the form

$$S^{(2)}(\delta\vec{\phi}) = \int_{\vec{p},\Omega} \frac{1}{2}\delta\phi^a(\vec{p},\Omega)\delta\phi^{b*}(\vec{p},\Omega)K^{ab}(\vec{p},\Omega) \tag{3.4.14}$$

where the kernel $K^{ab}(\vec{p},\Omega)$ is given by

$$K^{ab}(\vec{p},\Omega) = -\delta^{ab} - i\frac{U}{3}\int_{\vec{k},\omega} \text{Tr}\,\left(G(\vec{k},\omega)\tau^a G(\vec{k}+\vec{p},\omega+\Omega)\tau^b\right). \tag{3.4.15}$$

According to Eq. (3.3.23) and with $E^2(\vec{k}) = \epsilon^2(\vec{k}) + \frac{U}{3}|\vec{\phi}|^2$ the fermion Green function is

$$iG_{\alpha\alpha'}(\vec{k},\omega) = \frac{(\omega+\epsilon(\vec{k}))\delta_{\alpha\alpha'} + \sqrt{\frac{U}{3}}|\vec{\phi}|(\vec{n}\cdot\vec{\tau}_{\alpha\alpha'})}{\omega^2 - E^2(\vec{k}) + i\nu} \tag{3.4.16}$$

on the full Brillouin Zone. The trace identities for the Pauli matrices will be useful to compute the kernel

$$\begin{aligned} &\text{Tr } 1 = 2, \qquad \text{Tr } \tau^a = 0, \\ &\text{Tr } \tau^a\tau^b = 2\delta^{ab}, \qquad \text{Tr } \tau^a\tau^b\tau^c = 2i\epsilon^{abc}, \\ &\text{Tr } \tau^a\tau^b\tau^c\tau^d = 2\left(\delta^{ab}\delta^{cd} - \delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc}\right). \end{aligned} \tag{3.4.17}$$

We can now write the kernel $K^{ab}(\vec{p},\Omega)$ in the form

$$K^{ab}(\vec{p},\Omega) = K_0(\vec{p},\Omega)\delta^{ab} + K_1(\vec{p},\Omega)\epsilon^{abc}n_c + K_2(\vec{p},\Omega)n^a n^b \tag{3.4.18}$$

after inserting Eq. (3.4.16) into Eq. (3.4.15) and using the trace identities. One obtains for K_0, K_1 and K_2

$$\begin{aligned} K_0(\vec{p},\Omega) &= -1 + i\frac{2U}{3}\int_{\vec{k},\omega} \frac{\left((\omega+\epsilon(\vec{k}))(\omega+\Omega+\epsilon(\vec{k}+\vec{p})) - \frac{1}{4}\Delta^2\right)}{\left(\omega^2 - E^2(\vec{k}) + i\nu\right)\left((\omega+\Omega)^2 - E^2(\vec{k}+\vec{p}) + i\nu\right)}, \\ K_1(\vec{p},\Omega) &= +\frac{U}{3}\Delta\int_{\vec{k},\omega} \frac{\epsilon(\vec{k}) - \Omega - \epsilon(\vec{k}+\vec{p})}{\left(\omega^2 - E^2(\vec{k}) + i\nu\right)\left((\omega+\Omega)^2 - E^2(\vec{k}+\vec{p}) + i\nu\right)}, \\ K_2(\vec{p},\Omega) &= +i\frac{U}{3}\Delta^2\int_{\vec{k},\omega} \frac{1}{\left(\omega^2 - E^2(\vec{k}) + i\nu\right)\left((\omega+\Omega)^2 - E^2(\vec{k}+\vec{p}) + i\nu\right)}, \end{aligned} \tag{3.4.19}$$

where the integrals range over the full Brillouin Zone and $\Delta = 2\sqrt{\frac{U}{3}}|\vec{\phi}|$ (see Eq. (3.3.24)).

We are interested in studying the low-energy limit ($\Omega\Delta \ll 1$) of this system. It will be convenient to decompose the fluctuations $\delta\vec{\phi}(\vec{p},\Omega)$ into a longitudinal component $\sigma(\vec{p},\Omega)$ parallel to $\vec{n}$ and two transverse components $\pi_i(\vec{p},\Omega)$ perpendicular to $\vec{n}$. In terms of σ and $\vec{\pi}$, the gaussian action $S^{(2)}$, Eq. (3.4.14), has the form

$$\begin{aligned} S^{(2)}(\delta\vec{\phi}) = &\int_{\vec{p},\Omega} \left(K_0(\vec{p},\Omega) + K_2(\vec{p},\Omega)\right)|\sigma(\vec{p},\Omega)|^2 + \\ &+ \int_{\vec{p},\Omega} [K_0(\vec{p},\Omega)\vec{\pi}(\vec{p},\Omega)\cdot\vec{\pi}^*(\vec{p},\Omega) + K_1(\vec{p},\Omega)\epsilon^{abc}n_a\pi_b(\vec{p},\Omega)\pi_c^*(\vec{p},\Omega)]. \end{aligned} \tag{3.4.20}$$

I will now show that for wavevectors $\vec{p}$ close to the ordering wavevector $\vec{Q} = (\pi,\pi)$ the transverse components $\vec{\pi}$ become gapless. In contrast, the longitudinal component σ, remains massive (i.e. with a non-zero gap) over the entire Brillouin Zone. Thus, in a Néel state, the Hubbard Model has *two* gapless transverse spin waves $\vec{\pi}$ and a massive longitudinal amplitude mode σ.

We can check this statement by considering the limit $\Omega \to 0$ and $\vec{p} = \vec{Q} - \vec{q}$ where $|\vec{q}|$ is small, i.e. $|\vec{q}|\xi \ll 1$. Thus, the relevant limit is

$$\Omega\Delta \ll 1, \qquad |\vec{q}|\frac{v_F}{\Delta} \ll 1. \tag{3.4.21}$$

We wish to expand the kernels of Eq. (3.4.19) in powers of $\vec{q}$ and Ω, i.e. we are performing a gradient expansion.

Let us first compute K_0, K_1 and K_2 for $\Omega = 0$ *and* $\vec{p} = \vec{Q}$. The result, for $K_0(\vec{Q}, 0)$, is zero

$$K_0(\vec{Q}, 0) = -1 + i\frac{2U}{3}\int_{\vec{k},\omega} \frac{1}{\omega^2 - E^2(\vec{k}) + i\nu} \equiv 0. \tag{3.4.22}$$

This result follows from the gap equation. Similarly, the two other kernels K_1 and K_2 have the limit, after integrating over ω,

$$K_1(\vec{Q}, 0) = +\frac{1}{2}\frac{U}{3}\Delta \int_{\vec{k}} \frac{\epsilon(\vec{k})}{\left(\epsilon^2(\vec{k}) + \frac{1}{4}\Delta^2\right)^{3/2}} = 0 \tag{3.4.23}$$

which is a consequence of the symmetry of $\epsilon(\vec{k})$, and

$$K_2(\vec{Q}, 0) = +\frac{i}{4}\frac{U\Delta^2}{3}\int_{\vec{k}} \frac{1}{\left(\epsilon^2(\vec{k}) + \frac{1}{4}\Delta^2\right)^{3/2}} = +\frac{2\alpha}{3\pi^2} i\frac{U\Delta^2}{t^3} \tag{3.4.24}$$

where α is a numerical positive constant (of the order 1) which results from the evaluation of the integral.

These results imply the *presence* of a *mass* term for the *longitudinal* σ mode for $\vec{p} \approx \vec{Q}$ and $\Omega \to 0$ but the *absence* of such a *mass* term for both *transverse* $\vec{\pi}$ modes (see Fig. 3.5). The mass of the σ mode can be used to define a spin correlation length ξ_{spin}.

We can also estimate the spin-wave stiffness from these results. However, we need to go to leading order in both Ω and $\vec{q}$ to get these results. By expanding in powers of both Ω and $\vec{q}$ we find for $K_0(\vec{Q} - \vec{q}, \Omega)$

$$K_0(\vec{Q} - \vec{q}, \Omega) \approx a\Omega^2 - b\vec{q}^{\,2} + \text{h.o.t.} \tag{3.4.25}$$

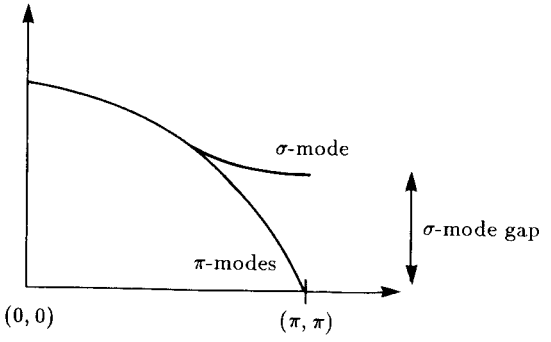


Figure 3.5 Spin-wave spectrum along the main diagonal of the Brillouin Zone

where

$$a = -i\frac{2U}{3}\int_{\vec{k},\omega} \frac{1}{\left(\omega^2 - E^2(\vec{k}) + i\nu\right)^2} \tag{3.4.26}$$

and

$$b = +i\frac{2U}{3}\int_{\vec{k},\omega} \frac{\frac{1}{4}\nabla^2\epsilon^2(\vec{k})}{\left(\omega^2 - E^2(\vec{k}) + i\nu\right)^2}. \tag{3.4.27}$$

Thus, the transverse $\vec{\pi}$ modes have the effective action (for $\vec{p} \approx \vec{Q}$ and low frequencies)

$$S^{(2)}_{\vec{\pi}} = \frac{1}{2}\int_{\vec{q},\Omega} |\vec{\pi}(\vec{q},\Omega)|^2 \rho(\Omega^2 - v_s^2 \vec{q}^{\,2}) \tag{3.4.28}$$

where we have defined ρ, the spin-wave stiffness, and v_s, the spin-wave velocity, to be given by

$$\rho = a \qquad \text{and} \qquad v_s^2 = \frac{b}{a}. \tag{3.4.29}$$

These results show that the transverse $\vec{\pi}$ modes are massless and have a linear dispersion relation

$$\Omega = v_s|\vec{q}|. \tag{3.4.30}$$

3.5 The Néel State and the Non-linear Sigma Model

I will now show that these results can be embodied in a very simple effective Lagrangian which includes all the relevant non-linear effects. The key to the argument is the observation that the Néel state breaks, on each sublattice, the spin rotation symmetry. This is a continuous symmetry. The transverse spin waves are gapless excitations because they only involve tilting the spins relative to the classical Mean-Field pattern. But they do not change the amplitude. Thus, it is natural to ask for the effective Lagrangian in which the amplitude fluctuations are frozen but the transverse ones are not. In mathematical terms this means that, in position space, the *staggered order parameter* $\vec{n}(\vec{r},t)$ will be slowly varying and its length will be constrained to be equal to a fixed value $|\vec{\phi}_c|$, the classical minimum. Thus, the fields $\sigma(\vec{k},\omega)$ and $\vec{\pi}(\vec{k},\omega)$ need to be scaled by a factor of $|\vec{\phi}_c|^{-1}$. The net effect is that the spin-wave stiffness ρ, Eq. (3.4.28), gets multiplied by $|\vec{\phi}_c|^2$, the solution of Eq. (3.3.7).

This calculation can in fact be carried out because the transverse spin waves remain massless to *all* orders in an expansion around the Mean-Field solution. This is guaranteed by a Ward Identitiy, which is a consequence of the symmetry. We shall derive this identity below.

Let us consider the Hubbard model coupled to an external magnetic field $\vec{H}(\vec{r},t)$

$$H_{\text{Zeeman}} = \sum_{\vec{r}} \vec{H}(\vec{r},t) \cdot c^{\dagger}_{\alpha}(\vec{r},t)\vec{\tau}_{\alpha\beta}c_{\beta}(\vec{r},t). \tag{3.5.1}$$

If we retrace the steps that led to the path-integral for the Hubbard-Stratonovich field $\vec{\phi}(\vec{r},t)$, we find a new effective action of the form

$$S_{\text{eff}}\left(\vec{\phi},\vec{H}\right) = -\int dt \sum_{\vec{r}} \frac{\vec{\phi}^2(\vec{r},t)}{2} - i\text{Tr } \ln\left(i\partial_t + \mu - M(\vec{\phi}) - \vec{H}(\vec{r},t)\cdot\vec{\tau}\right) \tag{3.5.2}$$

where the magnetic field $\vec{H}(\vec{r},t)$ is a c-number operator with matrix elements

$$\langle \vec{r}t\alpha|\vec{H}(\vec{r},t)|\vec{r}\,'t'\alpha'\rangle = \delta_{\alpha,\alpha'}\delta_{\vec{r},\vec{r}'}\delta(t-t')\vec{H}(\vec{r},t). \tag{3.5.3}$$

In principle, we will want to study the Néel state. Hence we will choose $\vec{H}(\vec{r},t)$ to be staggered and time-independent, i.e.

$$\vec{H}_s(\vec{r},t) = \vec{H}_s\ e^{i\vec{Q}\cdot\vec{r}} \tag{3.5.4}$$

with

$$\vec{Q} = (\pi,\pi). \tag{3.5.5}$$

We can prove the existence of gapless excitations by deriving a Ward Identity. This identity can be derived by standard methods [Amit 78]. Let us first shift the $\vec{\phi}$ field

$$\vec{\phi}'(\vec{r},t) \equiv \vec{\phi}(\vec{r},t) + \sqrt{\frac{3}{U}}\vec{H}(\vec{r},t). \tag{3.5.6}$$

We can write

$$S_{\text{eff}}\left(\vec{\phi},\vec{H}\right) = S_{\text{eff}}(\vec{\phi}',0) + \sqrt{\frac{3}{U}}\int dt \sum_{\vec{r}} \vec{\phi}'(\vec{r},t)\cdot\vec{H}(\vec{r},t) - \frac{3}{U}\int dt \sum_{\vec{r}} \frac{\vec{H}^2(\vec{r},t)}{2}. \tag{3.5.7}$$

Next, we can make use of the invariance of the integration measure $\mathcal{D}\vec{\phi}$ under the rotation $\phi''_a = R_{ab}\phi'_b$ (where R_{ab} is a rotation matrix) to shift the coordinates of the functional integral. The rotation matrix R_{ab} can be written in terms of Euler angles θ^c and rotation generators L^c in the form

$$R_{ab} = \left(e^{-iL^c\theta^c}\right)_{ab} \tag{3.5.8}$$

where

$$(L^c)_{ab} = -i\epsilon_{abc}. \tag{3.5.9}$$

Thus, for an infinitesimal rotation ($\theta^c \ll 1$)

$$\phi''_a = \phi'_a - \epsilon_{abc}\phi'_b\theta_c. \tag{3.5.10}$$

Using the invariance of the measure, we can now write for the vacuum persistence amplitude in the presence of the "source" $\vec{H}$

$$\begin{aligned} Z[\vec{H}] &\equiv e^{iF[\vec{H}]-i\frac{3}{U}\int dt \sum_{\vec{r}} \frac{\vec{H}^2}{2}} \\ &= \int \mathcal{D}\vec{\phi}' \; e^{i(S_{\rm eff}(\vec{\phi}',0) + \sqrt{\frac{3}{U}}\int dt \sum_{\vec{r}} \vec{\phi}'\cdot\vec{H} - \frac{3}{U}\int dt \sum_{\vec{r}} \frac{\vec{H}^2}{2})} \\ &= \int \mathcal{D}\vec{\phi}'' \; e^{i(S_{\rm eff}(\vec{\phi}'',0) + \sqrt{\frac{3}{U}}\int dt \sum_{\vec{r}} \vec{\phi}''\cdot\vec{H} - \frac{3}{U}\int dt \sum_{\vec{r}} \frac{\vec{H}^2}{2})} \\ &= \int \mathcal{D}\vec{\phi}' \; e^{i(S_{\rm eff}(\vec{\phi}',0) + \sqrt{\frac{3}{U}}\int dt \sum_{\vec{r}} \vec{\phi}'\cdot\vec{H} - \frac{3}{U}\int dt \sum_{\vec{r}} \frac{\vec{H}^2}{2}) + i\delta S_{\rm eff}(\vec{\phi}',\vec{H},\vec{\theta})} \end{aligned} \tag{3.5.11}$$

where

$$\delta S_{\rm eff}(\vec{\phi}', \vec{H}, \vec{\theta}) = -\sqrt{\frac{3}{U}} \int dt \sum_{\vec{r}} \epsilon^{abc} H^a(\vec{r},t)\phi'^b(\vec{r},t)\theta^c. \tag{3.5.12}$$

By expanding Eq. (3.5.11) in powers of θ we obtain to leading order

$$Z[\vec{H}] = Z[\vec{H}]\,(1 + \delta\bar{S}_{\rm eff} + ...)\,, \tag{3.5.13}$$

and thus since the θ^c are arbitrary

$$0 = \int dt \sum_{\vec{r}} \epsilon^{abc} H^a(\vec{r},t)\bar{\phi}'^b(\vec{r},t). \tag{3.5.14}$$

Here $\bar{\phi}'^b$ is the exact expectation value of ϕ'^b in the presence of $\vec{H}$.

Define now the generating functional of vertex functions $\Gamma[\bar{\phi}']$ by means of the Legendre transform

$$\Gamma[\bar{\phi}'] = \sqrt{\frac{3}{U}} \int dt \sum_{\vec{r}} \bar{\phi}'_a(\vec{r},t) H_a(\vec{r},t) - F[\vec{H}] \tag{3.5.15}$$

with

$$\frac{\delta F}{\delta H^b(\vec{r},t)} = \sqrt{\frac{3}{U}} \bar{\phi}'^b(\vec{r},t). \tag{3.5.16}$$

It follows [Amit 78] that

$$\frac{\delta \Gamma}{\delta \bar{\phi}'^a(\vec{r},t)} = \sqrt{\frac{3}{U}} H^a(\vec{r},t). \tag{3.5.17}$$

The one-particle irreducible vertex functions can be defined in terms of functional derivatives of Γ relative to $\bar{\phi}'_a$. For instance, the two point-function

$$\Gamma^{(2)}_{ab}(\vec{r},t,\vec{r}\,',t') \equiv \frac{\delta^2 \Gamma}{\delta\bar{\phi}'_a(\vec{r},t)\delta\bar{\phi}'_b(\vec{r}\,',t')} \tag{3.5.18}$$

is the inverse of the $\vec{\phi}\,'$ two-point Green function

$$\sum_{\vec{r}\,',b} \frac{\delta^2\Gamma}{\delta\bar{\phi}'^a(\vec{r},t)\delta\bar{\phi}'^b(\vec{r}\,',t')} \frac{\delta^2 F}{\delta H^b(\vec{r}\,',t')\delta H^c(\vec{r}\,'',t'')} = \delta_{ac}\delta_{\vec{r},\vec{r}''}\delta(t-t''). \tag{3.5.19}$$

With those definitions, Eq. (3.5.14) can be brought to the form

$$0 = \int dt \sum_{\vec{r}} \left(\epsilon^{abc} \frac{\delta \Gamma}{\delta \bar{\phi}'^a(\vec{r},t)} \bar{\phi}'^b(\vec{r},t) \right). \tag{3.5.20}$$

To avoid cumbersome notations we denote from now on $(\vec{r},t) \equiv (x_1,x_2,x_0) \equiv x$ and $\delta_{\vec{r},\vec{r}'}\delta(t-t') \equiv \delta(x-x')$. By taking a further derivative with respect to $\bar{\phi}'^d(x')$, one gets the Ward identity

$$0 = \int dt \sum_{\vec{r}} \epsilon^{abc} \left(\frac{\delta^2 \Gamma}{\delta \bar{\phi}'^d(x') \delta \bar{\phi}'^a(x)} \bar{\phi}'^b(x) + \frac{\delta \Gamma}{\delta \bar{\phi}'^a(x)} \delta^{bd} \delta(x-x') \right). \tag{3.5.21}$$

In particular, for the Néel state

$$\bar{\phi}'^b(x) = |\vec{\phi}| n^b (-1)^{x_1+x_2} \tag{3.5.22}$$

and the corresponding staggered field of Eq. (3.5.4), we get the Ward Identity

$$\epsilon^{adc} \sqrt{\frac{3}{U}} H_s^a(x') = -\epsilon^{abc} |\vec{\phi}| n^b \int dt \sum_{\vec{r}} \frac{\delta^2 \Gamma}{\delta \bar{\phi}'^d(x') \delta \bar{\phi}'^a(x)} (-1)^{x_1+x_2}. \tag{3.5.23}$$

In momentum space, Eq. (3.5.23) simply becomes

$$\epsilon^{acd} \sqrt{\frac{3}{U}} H_s^a(\vec{Q}) = \lim_{\substack{\omega \to 0 \\ \vec{p} \to 0}} \epsilon^{abc} |\vec{\phi}| n^b \Gamma^{(2)}_{da}(\vec{Q}+\vec{p},\omega). \tag{3.5.24}$$

The spontaneous breaking of symmetry means that as the external field is switched off, the order parameter remains finite. In this case, for the right hand side of Eq. (3.5.24) to vanish, the contraction of the vertex two-point function with the Levi-Cevita tensor and with $\vec{n}$ must vanish in this limit. Thus if $|\vec{\phi}| \neq 0$, the *transverse* components of $\vec{\phi}$ must have a *pole* at $\omega = 0$ and $\vec{p} = \vec{Q}$ in their correlation function (see Eq. (3.5.18)). We found before that this was indeed the case in the leading order of an expansion around Mean-Field theory.

Since now we know that this leading-order-pole must persist to *all orders*, we can look for an effective Lagrangian with the following properties: (a) the amplitude fluctuations are suppressed, (b) a massless pole for each transverse component and (c) a minimum number of gradients.

The simplest expression satisfying these properties is the non-linear σ-model with a Lagrangian density given by

$$\mathcal{L}_{\text{eff}} = \frac{\rho}{2} \left((\partial_t \vec{n})^2 - v_s^2 (\nabla \vec{n})^2 \right) + \ldots \tag{3.5.25}$$

where $\vec{n}$ satisfies the constraint

$$|\vec{n}| = 1. \tag{3.5.26}$$

Here $\vec{n}$ represents the slow fluctuations of the order parameter field. The spin-wave stiffeness ρ and velocity v_s appearing in Eq. (3.5.25) are not generally identical to the values we calculated above (see Eq. (3.4.29)). The reason for that is that the non-linear σ-model is the effective theory at low frequencies and for wavevectors close to $\vec{Q}$. It is the result of integrating out all fluctuations

at high energies. This process significantly renormalizes the values of ρ and v_s. In addition, I have ignored the possible existence of topological terms. We will see in chapters 5 and 6 that in one dimension these terms are generated but not in two or higher dimensions.

It is a simple matter to see that if one solves for the constraint $|\vec{n}| = 1$ in terms of a σ and a $\vec{\pi}$ field

$$\vec{n} = \begin{pmatrix} \sigma \\ \vec{\pi} \end{pmatrix} \tag{3.5.27}$$

and expands in powers of $\vec{\pi}$, one finds to the leading quadratic order in $\vec{\pi}$, the effective action Eq. (3.5.2).

We conclude that the quantum fluctuations around a Néel state are described by a non-linear σ-model. One key assumption that we have made here is that ρ was assumed to be large. Otherwise this expansion does not make sense. In fact for ρ sufficiently small, the non-linear σ-model has wild fluctuations which destroy the Néel state. The system becomes a (quantum) paramagnet and the spin symmetry is unbroken. We will see below that *frustrating* interactions will generally produce this effect. On the other hand, given the large renormalizations in ρ and v_s, it is not possible to be sure whether the half-filled Hubbard model is in a Néel state (i.e. an antiferromagnet) or in a disordered state. In practice only numerical calculations (i.e. fermion Monte Carlo or finite size exact diagonalizations) can yield more reliable answers for a specific Hubbard model. So far the evidence strongly favors a Néel state.

Finally we should consider the connection between the collective excitations $\vec{\phi}(\vec{x}, t)$ and the susceptibilities of the Hubbard model. From Eq. (3.5.1) we see that the field $\vec{H}(x)$ couples to the local moment $c^\dagger_\alpha(x)\vec{\tau}_{\alpha\beta}c_\beta(x)$. Thus by functionally differentiating with respect to $\vec{H}(x)$ we should be able to compute expectation values related to the spin degrees of freedom. Indeed the spin correlation function

$$K^{aa'}(x, x') = \langle \mathrm{Gnd}|\hat{T}S^a(x')S^{a'}(x')|\mathrm{Gnd}\rangle$$

is equivalent to

$$K^{aa'}(x, x') = -g(x, x')\frac{\delta^2 F}{\delta H_s^a(x)\delta H_s^{a'}(x')} \equiv G^{(2)}_{aa'}(x, x') \tag{3.5.28}$$

where $g(x, x')$ is a sign

$$g(x, x') = (-1)^{x_1+x_2+x_1'+x_2'}. \tag{3.5.29}$$

Since $G^{(2)}_{aa'}(x, x')$ is the inverse of $\Gamma^{(2)}_{aa'}(x, x')$, we conclude that a *zero* in $\Gamma^{(2)}_{aa'}(\vec{p}, \omega)$ at $(\vec{Q}, 0)$ implies a divergence of $G^{(2)}_{aa'}(x, x')$ also at $(\vec{Q}, 0)$. Thus the staggered static susceptibility $\chi_\perp(\vec{Q}, 0)$, which is the Fourier transform of $K_{aa'}(x, x')$ must have a delta-function peak at $(\vec{Q}, 0)$ if the ground state is a Néel State $(\vec{Q}, 0)$. This peak only appears in the transverse components of $K_{aa'}$ since the longitudinal components are connected to excitations with a non-zero gap.

References

[Abrikosov 63] A.Abrikosov, L.Gor'kov and I.Dzyaloshinsky. *Methods of Quantum Field Theory in Statistical Physics* , Dover, 1963.

[Amit 78] D.J. Amit. *Field Theory, the Renormalization Group, and critical Phenomena* , World Scientific, 1978.

[Bloch 30] F. Bloch. Z. Physik 61, 206 (1930).

[Coleman 85] S. Coleman. *Aspects of Symmetry* , Cambridge University Press, 1985.

[Dyson 56] F.J. Dyson. Phys. Rev. 102, 1217 (1956) and Phys. Rev. 102, 1230 (1956).

[Faddeev 75] L. Faddeev . In *Methods of Field Theory*, Les Houches Summer School 1975, Session XXVIII, edited by R. Stora and J. Zinn-Justin, North-Holland, 1976.

[Feynman 65] R.P. Feynman and A.R. Hibbs. *Path Integrals and Quantum Mechanics* , McGraw-Hill, 1965.

[Holstein 40] T. Holstein and H. Primakoff. Phys. Rev. 58, 1098 (1940).

[Kadanoff 62] L .P. Kadanoff and G. Baym. *Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and non-Equilibrium Problems*, Benjamin, 1962.

[Maleev 57] S.V. Maleev. J. exptl. Theoret. Phys. 33, 1010 (1957) and JETP 6, 776 (1958).

[Mattis 65] D.C. Mattis. *The Theory of Magnetism* , Harper-Row, 1965.

[Negele 88] J.W. Negele and H. Orland. *Quantum Many-Particle Systems*, Addison-Wesley, 1988.

CHAPTER 4

One-Dimensional Quantum Antiferromagnets

In this chapter we will discuss the physical behavior of one-dimensional quantum antiferromagnets. Although these systems are not connected in a direct way with the physics of problems of current interest, such as high-temperature superconductors, it is worthwhile to study them for several reasons: (a) in many cases we have exact solutions (which are lacking in higher dimensions), (b) they exhibit a wealth of ground states, including disordered phases, and (c) they are a natural testing ground for methods and approximations. We shall first discuss the spin one-half Heisenberg chain and later discuss its generalization to (a) higher spin-S and (b) other symmetry groups.

4.1 The Spin-One-Half Heisenberg Chain

Consider the Heisenberg model on a one-dimensional chain of N sites. The Hamiltonian is

$$H = J \sum_{n=1}^{N} \vec{S}(n) \cdot \vec{S}(n+1) \tag{4.1.1}$$

where $J > 0$. I will asumme that N is an even integer and that we have periodic boundary conditions. Much of what we know about this system comes from (a) Bethe-Ansatz solution for the ground state [Bethe 31] and excitation spectrum [Yang 69], (b) mapping to the Sine-Gordon theory [Luther 75], (c) non-Abelian bosonization [Affleck 85], and (d) mapping to the sigma model [Haldane 83].

The exact solution via Bethe-Ansatz is very peculiar to one-dimensional integrable systems and hence is not generalizable. The other methods are also very specific to one dimension but they are more generally applicable, and higher dimensional versions of them are currently being developed. Thus we shall concentrate mainly on them. The mapping to the Sine-Gordon system is based on the Abelian bosonization transformation ([Bloch 34], [Lieb 65],

[Luther 75], [Coleman 75], and [Mandelstam 75]). In a deep sense, it is a particular case of the non-Abelian bosonization developed by Witten [Witten 84], Polyakov and Wiegmann [Polyakov 84], and applied to spin systems by Affleck [Affleck 85]. The main advantage of all these approaches is that they are non-perturbative; they yield the exact behavior of the ground state properties at long distances, and that, in principle, one can find the low energy spectrum. One important feature is the existence, in addition to spin-waves, of soliton states. These states are highly extended configurations of spins which cannot be created locally and that comprise the lowest portion of the spectrum of these systems.

4.1.1 The Bethe-Ansatz Solution

I will not attempt to give a detailed description of the Bethe-Ansatz solution, which is fairly technical. A good, reasonably recent summary can be found in the Les Houches Lectures of 1981, in particular the articles by Faddeev and Lowenstein [Faddeev 81] [Lowenstein 81].

Here I will review very quickly the method as given by Lowenstein. The main idea is to consider the wave function for a pure state of N spins one-half, each labeled by an index $s(n) = \pm\frac{1}{2}$, $(n = 1, \ldots, N)$. The total spin of the system is $\vec{S} = \sum_{n=1}^{N} \vec{S}(n)$. We will consider states $\Psi(s(1), \ldots, s(n))$ in which $(N - M)$ spins are up $(+\frac{1}{2})$ and M are down $(-\frac{1}{2})$. Thus, the total z-component of the spin is

$$S_z \Psi(s(1), \ldots, s(N)) = \left(\sum_{n=1}^{N} s(n) \right) \Psi(s(1), \ldots, s(N)) \tag{4.1.2}$$

with

$$\sum_{n=1}^{N} s(n) = \frac{N}{2} - M. \tag{4.1.3}$$

We shall denote with

$$\Psi(s(1), \ldots, s(N)) \equiv \phi(x_1, \ldots, x_M) \tag{4.1.4}$$

a state with the j-th down spin located at the site x_j $(1 \leq x_1 \leq \ldots \leq x_j \leq \ldots \leq x_M \leq N)$. Thus, if Ψ_0 is the *ferromagnetic* state

$$\Psi_0 = |\uparrow \ldots \uparrow\rangle \tag{4.1.5}$$

the most general state with M spins down has the form

$$\Psi = \sum_{\{x_j\}} \phi(x_1, \ldots, x_M) S^-(x_1) \ldots S^-(x_M) \Psi_0 \tag{4.1.6}$$

where $S^-(n)$ is the lowering operator at site n.

The Heisenberg model is translationally invariant and on a chain with periodic boundary conditions has the translation symmetry in which the n-th site is identified with the $(N + n)$-th site. Thus we can look for a basis in

which not only $\vec{S}^2$ and S_z are diagonal but the cyclic permutation operator $\tilde{P}$ where

$$\tilde{P}\Psi(s(1),\ldots,s(N)) = \Psi(s(N),s(1),\ldots,s(N-1)), \tag{4.1.7}$$

is also diagonal.

4.1.2 The Basis Functions

Bethe's method begins by first writing the Hamiltonian in terms of a spin-exchange operator $P_{n,m}$ where

$$\begin{aligned} P_{n,m}\Psi(s(1),\ldots,s(n),\ldots,s(m),\ldots,s(N)) = \\ \Psi(s(1),\ldots,s(m),\ldots,s(n),\ldots,s(N)), \end{aligned} \tag{4.1.8}$$

in the form

$$H = J\sum_{n=1}^{N}(P_{n,n+1}-1) \tag{4.1.9}$$

with periodic boundary conditions. Consider first a state with one-spin down

$$\Psi(s_1,\ldots,s_N) = \sum_{p=1}^{N}\phi(p)|\uparrow\ldots\downarrow\ldots\uparrow\rangle \tag{4.1.10}$$

where the spin at site p is down.

By using the cyclic translation operator $\tilde{P}$ we see that its main effect is just to shift the location of the down spin by one. Thus an eigenstate of $\tilde{P}$ with eigenvalue μ should satisfy

$$\tilde{P}\phi(p) = \phi(p+1) = \mu\phi(p) \quad \text{for} \quad p = 1,\ldots,N-1 \tag{4.1.11}$$

and

$$\tilde{P}\phi(N) = \phi(1) = \mu\phi(N). \tag{4.1.12}$$

Hence

$$\phi(p) = \mu^{p-1}\phi(1) \tag{4.1.13}$$

and if we set $\phi(1) = 1$, we get

$$\phi(p) = \mu^{p-1} \tag{4.1.14}$$

and in particular μ must satisfy

$$1 = \phi(1) = \phi(N+1) = \mu^N \tag{4.1.15}$$

i.e., it is an N-th root of unity.

Now, a state with one spin down can either be a member of the multiplet with highest total spin-$\frac{N}{2}$ or of the multiplet with total spin-$(\frac{N}{2}-1)$. In this last case the state is the highest weight state in the multiplet and satisfies

$$S^+\Psi = 0 \tag{4.1.16}$$

where

$$S^+ = \sum_{n=1}^{N}S^+(n). \tag{4.1.17}$$

Thus we get

$$\begin{aligned} S^+\Psi &= \sum_{p=1}^{N} \phi(p) S^+ |\uparrow \ldots \downarrow \ldots \uparrow\rangle \\ &= \sum_{p=1}^{N} \phi(p) S^+(p) |\uparrow \ldots \downarrow \ldots \uparrow\rangle \\ &= \left(\sum_{p=1}^{N} \phi(p) \right) \Psi_0 . \end{aligned} \tag{4.1.18}$$

Using Eqs. (4.1.16) and (4.1.14), we obtain

$$0 = \sum_{p=1}^{N} \phi(p) = \sum_{p=1}^{N} \mu^{p-1} = \frac{1-\mu^N}{1-\mu}. \tag{4.1.19}$$

Thus $\mu^N = 1$ and $\mu \neq 1$. Hence we found $N-1$ spin $(\frac{N}{2} - 1)$ *multiplets* (the other members of the multiplet can be found by applying S^-). The case $\mu = 1$ represents the state $S = \frac{N}{2}$ which belongs to the multiplet of the ferromagnetic state.

If we now consider the case of M spins down we can still find states with $S = S_z$. They satisfy

$$\tilde{P}\phi(p_1, \ldots, p_M) = \phi(p_1 + 1, \ldots, p_M + 1) = \lambda \phi(p_1, \ldots, p_M) \qquad (p_M < N) \tag{4.1.20}$$

and

$$\tilde{P}\phi(p_1, \ldots, p_{M-1}, N) = \phi(1, p_1 + 1, \ldots, p_{M-1} + 1) = \lambda \phi(p_1, \ldots, p_{M-1}, N). \tag{4.1.21}$$

We look for wave functions which are *products* of *single* down-spin wave functions

$$\phi(p_1, \ldots, p_M) = \mu_1^{p_1 - 1} \ldots \mu_M^{p_M - 1}. \tag{4.1.22}$$

By choosing

$$\lambda = \prod_{j=1}^{M} \mu_j \tag{4.1.23}$$

we can satisfy Eq. (4.1.20). As it stands, this ansatz does *not* satisfy Eq. (4.1.21), but if we permute the *order* of the parameters $\mu_1, \ldots, \mu_M$ we can find a solution. Thus, Bethe introduced the *Bethe-Ansatz*

$$\phi(p_1, \ldots, p_M) = \sum_{P \epsilon \mathcal{S}_M} A_P \mu_{P1}^{n_1} \ldots \mu_{PM}^{n_M} \tag{4.1.24}$$

where P belongs to the permutation group $\mathcal{S}_M$ (i.e. $(P1, \ldots, PM)$ is a permutation of $(1, \ldots, M)$). Now everything is consistent provided that the identity

$$A_{P\tilde{Q}^{-1}} \equiv A_{PM, P1, \ldots, P(M-1)} = A_P \mu_{PM}^N \tag{4.1.25}$$

where $(\tilde{Q}1, \tilde{Q}2, \tilde{Q}3, \ldots, \tilde{Q}M) = (2, 3, 4, \ldots, M, 1)$ holds. The demand that ϕ be a highest weight state with $S = \frac{N}{2} - M \equiv S_z$ yields the constraint [Lowenstein 81]

$$\frac{A_{P'}}{A_P} = -\frac{2\mu_{P(k+1)} - \mu_{Pk}\mu_{P(k+1)} - 1}{2\mu_{Pk} - \mu_{Pk}\mu_{P(k+1)} - 1} \tag{4.1.26}$$

for all k and for all pairs of permutations P and P' such that $(P'1 \ldots P'kP'(k+1) \ldots P'M) \equiv (P1 \ldots P(k+1)Pk \ldots PM)$.

Define χ_j

$$\mu_j = \frac{\chi_j + i\frac{\pi}{2}}{\chi_j - i\frac{\pi}{2}}. \tag{4.1.27}$$

Then one finds

$$\frac{A_{P'}}{A_P} = \frac{\chi_{P'j} - \chi_{Pj} + i\pi}{\chi_{P'j} - \chi_{Pj} - i\pi}. \tag{4.1.28}$$

By combining Eqs. (4.1.27) and (4.1.25) we get

$$\left(\frac{\chi_{PM} + i\frac{\pi}{2}}{\chi_{PM} - i\frac{\pi}{2}}\right)^N = \frac{A_{PMP1 \ldots P(M-1)}}{A_{P1P2 \ldots PM}}. \tag{4.1.29}$$

Using Eq. (4.1.28) repeatedly we obtain

$$\left(\frac{\chi_{PM} + i\frac{\pi}{2}}{\chi_{PM} - i\frac{\pi}{2}}\right)^N = \prod_{j=1}^{M-1} \left(\frac{\chi_{PM} - \chi_{Pj} + i\pi}{\chi_{PM} - \chi_{Pj} - i\pi}\right). \tag{4.1.30}$$

Since this equation should be valid for all permutations P, we get the *Bethe Ansatz equations*

$$\left(\frac{\chi_j + i\frac{\pi}{2}}{\chi_j - i\frac{\pi}{2}}\right)^N = -\prod_{l=1}^{N} \left(\frac{\chi_j - \chi_l + i\pi}{\chi_j - \chi_l - i\pi}\right). \tag{4.1.31}$$

Also we see that

$$\lambda^N \equiv \prod_j \left(\frac{\chi_j + i\frac{\pi}{2}}{\chi_j - i\frac{\pi}{2}}\right)^N = 1. \tag{4.1.32}$$

Thus, for all M, the eigenvalue of $\tilde{P}$ is an N-th root of unit. For a given value of $S = \frac{N}{2} - M$, the Hilbert space with given S and $\tilde{P}$ has a huge size. It is thus generally unlikely that this basis will diagonalize a randomly chosen Hamiltonian. It is now known that systems that can be diagonalized in this basis, like the nearest neighbor Heisenberg chain, do so because they are completely integrable, i.e. have an infinite number of conservation laws [Faddeev 81].

4.1.3 The Spectrum

Let us now act with the Heisenberg Hamiltonian on a Bethe-Ansatz wave function. The result is

$$\begin{aligned} H\phi(p_1,\ldots,p_M) = J &\sum_{\substack{j=1 \\ p_j \neq p_{j+1}-1}}^{M} \phi(p_1,\ldots,p_j+1,\ldots,p_M) + \\ + J &\sum_{\substack{j=1 \\ p_j \neq p_{j-1}+1}}^{M} \phi(p_1,\ldots,p_j-1,\ldots,p_M) + \\ + (N-2M)J&\phi(p_1,\ldots,p_M) + \\ + 2J &\sum_{\substack{j=1 \\ p_j \neq p_{j+1}-1}}^{M} \phi(p_1,\ldots,p_j,\ldots,p_M) - NJ\phi(p_1,\ldots,p_M). \end{aligned} \tag{4.1.33}$$

The first and second terms come from acting with $\sum_n P_{n,n+1}$ on $\uparrow\downarrow$ and $\downarrow\uparrow$ pairs. The third and fourth terms come from acting with $\sum_n P_{n,n+1}$ on $\uparrow\uparrow$ and $\downarrow\downarrow$ pairs. Using the Bethe-Ansatz, Eq. (4.1.31), we can put Eq. (4.1.33) in the form

$$\begin{aligned} H\phi(p_1,\ldots,p_M) = J &\sum_{j=1}^{M} [\mu(\chi_j) + \mu^{-1}(\chi_j) - 2]\phi(p_1,\ldots,p_M) + \\ - J &\sum_{j=1}^{M} [\phi(\ldots p_j+1, p_j+1\ldots) + \phi(\ldots p_j, p_j\ldots) + \\ - 2&\phi(\ldots p_j, p_{j+1}\ldots)]. \end{aligned} \tag{4.1.34}$$

This last term (in brackets) is found to vanish. Thus the Bethe-Ansatz state, Eq. (4.1.24), is an eigenstate of the Heisenberg model with eigenvalue E given by

$$E = J\sum_{j=1}^{M} [\mu(\chi_j) + \mu^{-1}(\chi_j) - 2] = -J\sum_{j=1}^{M} \frac{\pi^2}{\chi_j^2 + (\frac{\pi}{2})^2}. \tag{4.1.35}$$

We must now find solutions to the Bethe-Ansatz equation, Eq. (4.1.31). Intuitively, if $J > 0$ (antiferromagnet), we expect the ground state to have $S_z = 0$ ("Néel ") and thus $\frac{M}{N} = \frac{1}{2}$. Let us assume that the solutions of the Bethe-Ansatz equations are *real roots* χ_j. By taking logarithms we can write the Bethe-Ansatz equations in the form

$$2N\tan^{-1}\left(\frac{\chi_j}{\frac{\pi}{2}}\right) - 2\sum_{i=1}^{M} \tan^{-1}\left(\frac{\chi_j - \chi_i}{\pi}\right) = 2\pi I_j \tag{4.1.36}$$

for $j = 1,\ldots,M$ and where I_j are integers (half-integers) for $N - M$ odd (even).

Let us now assume that $\{\chi_j\}$ is a set of real roots with $N-M$ odd. The *function* $J(\chi)$

$$J(\chi)=\frac{1}{2\pi}\left(2N\tan^{-1}\left(\frac{\chi}{\frac{\pi}{2}}\right)-2\sum_{i=1}^{M}\tan^{-1}\left(\frac{\chi-\chi_i}{\pi}\right)\right) \tag{4.1.37}$$

is a monotonically increasing function of χ. If J happens to take the value of one of the integers I_i, $J^{-1}(I_i)=\chi$ will be equal to the corresponding root χ_i. However, it may happen that for some integers the value of χ may not be in the set $\{\chi_j\}$. Such a χ is called a hole (not to be confused with the "holes" of a more general context). If the roots are closely spaced (i.e. their separation vanishes in the thermodynamic $N\to\infty$ limit), we should be able to define a distribution of roots and holes $\rho(\chi)$

$$\rho(\chi)=\frac{dJ(\chi)}{d\chi} \tag{4.1.38}$$

or, equivalently,

$$J(\chi)=J(-\infty)+\int_{-\infty}^{\chi}d\chi'\rho(\chi'). \tag{4.1.39}$$

Now $\frac{dJ}{d\chi}$ is given by differentiating Eq. (4.1.37)

$$\frac{dJ}{d\chi}=\rho(\chi)=\frac{N/2}{\chi^2+\left(\frac{\pi}{2}\right)^2}-\sum_{j=1}^{M}\frac{1}{(\chi-\chi_i)^2+\pi^2}. \tag{4.1.40}$$

Let $\{\theta_j\}_{j=1\ldots n}$ denote the positions of the holes. In the $N\to\infty$ limit the following approximation is valid

$$\sum_{i=1}^{M}f(\chi_i)=\int_{-\infty}^{+\infty}d\chi\rho(\chi)f(\chi)-\sum_{i=1}^{n}f(\theta_i) \tag{4.1.41}$$

where n is the number of holes. By using these results we find the integral equation

$$\rho(\chi)+\int_{-\infty}^{+\infty}d\chi'\frac{\rho(\chi')}{(\chi-\chi')^2+\pi^2}=\frac{N/2}{\chi^2+\left(\frac{\pi}{2}\right)^2}+\sum_{j=1}^{n}\frac{1}{(\chi-\theta_j)^2+\pi^2}. \tag{4.1.42}$$

Consider now the set $\{\chi_1,\ldots,\chi_M,\theta_1,\ldots,\theta_n\}$ of roots and holes and let ξ_k denote the k-th element in this set counting from left to right on the χ-axis. This element is defined by

$$\int_{-\infty}^{\xi_k}\rho(\chi)d\chi=J(\xi_k)-J(-\infty)=I_k-\frac{(M-N)}{2}. \tag{4.1.43}$$

The integral equation is solved by a Fourier transform

$$\rho(\chi)=\int_{-\infty}^{+\infty}\frac{dp}{2\pi}e^{ip\chi}\tilde{\rho}(p). \tag{4.1.44}$$

One finds the solution

$$\tilde{\rho}(p) = \tilde{\rho}_0(p) + \sum_{j=1}^{n} \frac{e^{-ip\theta_j - \frac{\pi|p|}{2}}}{2\cosh\frac{\pi p}{2}} \tag{4.1.45}$$

with

$$\tilde{\rho}_0(p) = \frac{N/2}{2\cosh\frac{p\pi}{2}}. \tag{4.1.46}$$

Thus

$$\rho(\chi) = \rho_0(\chi) + \sum_j \rho_{\text{hole}}(\chi - \theta_j) \tag{4.1.47}$$

and

$$\rho_0(\chi) = \frac{N}{2\cosh\chi}. \tag{4.1.48}$$

The total number of roots M in a state with n holes is

$$M = \int_{-\infty}^{+\infty} d\chi\rho(\chi) - n = \tilde{\rho}(0) - n = \frac{N-n}{2}. \tag{4.1.49}$$

Since M is an integer, n must be even (odd) for N even (odd). This state has the energy eigenvalue

$$E = -J\pi^2 \int d\chi \frac{\sigma(\chi)}{\chi^2 + \left(\frac{\pi}{2}\right)^2}. \tag{4.1.50}$$

Here I introduced the density of roots for the Bethe-Ansatz equations

$$\sigma(\chi) = \rho(\chi) - \sum_{i=1}^{n} \delta(\chi - \theta_i). \tag{4.1.51}$$

In Fourier space, we get

$$E = -J\pi \int dp\tilde{\sigma}(-p)e^{-\frac{\pi|p|}{2}} \tag{4.1.52}$$

with

$$\tilde{\sigma}(p) = \tilde{\rho}(p) - \sum_{i=1}^{n} e^{-ip\theta_i}. \tag{4.1.53}$$

We find the result

$$E = E_0 + \sum_{i=1}^{n} E_h(\theta_i) \tag{4.1.54}$$

where E_0, the ground energy state, is

$$E_0 = -2NJ\ln 2$$

and the "excitation energy" (i.e. "holes")

$$E_h(\theta) = \frac{\pi J}{\cosh\theta}. \tag{4.1.55}$$

Thus, we can minimize the energy by choosing the solution with real roots and no holes (complex roots are irrelevant to this issue [Lowenstein 81]). The

total spin S for this state when N is even is obtained from Eqs. (4.1.3) and (4.1.49)

$$S = \frac{N}{2} - M = 0. \tag{4.1.56}$$

Thus the ground state is a *singlet* ($S = 0$). The excitations are "holes" with energy $\frac{\pi J}{\cosh \theta}$. For a lattice with N sites, N even (odd), there are an even (odd) number of holes. A state with one hole constructed in this manner carries $S_z = +\frac{1}{2}$. The spin-reversed hole is found by acting with S^- on this state. These states are degenerate as required by the SU(2) symmetry.

The momentum of these states can be calculated by noting that the operator $\tilde{P}$ which translates the wave function by one lattice spacing is related to the total momentum $\bar{P}$ of the state by

$$\tilde{P}\phi(p_1, \ldots, p_M) = e^{i\bar{P}}\phi(p_1, \ldots, p_M). \tag{4.1.57}$$

Before we found that the eigenvalue of $\tilde{P}$ was λ. Hence

$$\bar{P} = -i \ln \lambda = -i \sum_{j=1}^{M} \ln \mu_j = -i \sum_{j=1}^{M} \ln \left(\frac{\chi_j + i\frac{\pi}{2}}{\chi_j - i\frac{\pi}{2}} \right). \tag{4.1.58}$$

We can also write

$$\bar{P} = -2 \sum_{j=1}^{M} \tan^{-1} \left(\frac{2\chi_j}{\pi} \right) + M\pi. \tag{4.1.59}$$

In terms of "holes" θ_i and the distribution $\rho(\chi)$ we can write $\bar{P}$ in the form

$$\bar{P} = \bar{P}_0 + \sum_{i=1}^{n} \bar{P}_i \tag{4.1.60}$$

where $\bar{P}_0$ is the total momentum of the ground state

$$\bar{P}_0 = -\int_{-\infty}^{+\infty} d\chi \rho_0(\chi) 2 \tan^{-1} \left(\frac{2\chi}{\pi} \right) + M\pi \tag{4.1.61}$$

and $\bar{P}_i$ is the contribution from the i-th "hole" (see Eq. (4.1.45))

$$\bar{P}_i = \int_{-\infty}^{+\infty} d\chi \int \frac{dp}{2\pi} 2 \tan^{-1} \left(\frac{2\chi}{\pi} \right) \frac{e^{ip(\chi - \theta_i)}}{1 + e^{-\pi|p|}}. \tag{4.1.62}$$

Since $\rho_0(\chi)$ is even (see Eq. (4.1.48)), the total momentum of the ground state is (mod 2π)

$$\bar{P}_0 = M\pi \tag{4.1.63}$$

as predicted by Marshall's theorem [Marshall 55].

What is the momentum of the first excited state? From Mean-Field theory, which yields a Néel state, we expect that the lowest excited state should be a spin wave with wave vector $Q = \pi$ (i.e. momentum $\bar{P} = \pi$) and vanishing energy. From the excitation energy, Eq. (4.1.54), we learn that there are massless excitations (i.e. $E \to E_0$) if $\theta \to \pm\infty$. But, in this limit, $\bar{P}_i$ has the

value

$$\begin{aligned}\lim_{\theta_i\to\pm\infty}\bar{P}_i &= +\lim_{\theta_i\to\pm\infty}\int d\chi\int\frac{dp}{2\pi}2\tan^{-1}\left(\frac{2\chi}{\pi}\right)\frac{e^{ip(\chi-\theta_i)}}{1+e^{-\pi|p|}}\\ &= +\lim_{\theta_i\to\pm\infty}\int d\chi\int\frac{dp}{2\pi}2\tan^{-1}\left(\frac{2}{\pi}(\chi+\theta_i)\right)\frac{e^{ip\chi}}{1+e^{-\pi|p|}}\\ &= \pm\pi\int d\chi\int\frac{dp}{2\pi}\frac{e^{ip\chi}}{1+e^{-\pi|p|}}\\ &= \pm\pi\int dp\frac{\delta(p)}{1+e^{-\pi|p|}}.\end{aligned} \tag{4.1.64}$$

Thus we get

$$\lim_{\theta_i\to\pm\infty}\bar{P}_i = \pm\frac{\pi}{2}. \tag{4.1.65}$$

This result means that the lowest excited state of a chain with N even, which has *two* "holes", has total momentum equal either to zero or π (mod 2π). In fact we can view this state as the sum of two "single" particle states (i.e. "hole"), each with momenta $\pm\frac{\pi}{2}$. In other words, this state is *not* a spin-wave with momentum π. Rather the system behaves as if its elementary excitations had momenta close to $\pm\frac{\pi}{2}$. This resembles the physics of one-dimensional *fermions* on a half-filled chain. The Fermi "surface" is just two points, $k_F = \pm\frac{\pi}{2}$. The elementary excitations are particle-hole pairs with momenta close to the Fermi points. We will see below that this system, with purely bosonic degrees of freedom, indeed has fermions in its spectrum.

4.2 Fermions and the Heisenberg Model.

4.2.1 The Jordan-Wigner Transformation

At first sight it may appear to be obvious that there should be fermions in the spectrum of the Heisenberg model. After all, we derived the Heisenberg model as the strong coupling limit of a purely fermionic system: the half-filled Hubbard model. However, the fermions found in the last section are *not* the "constituent" band (Hubbard) fermions. For one thing these states carry no electric charge. The spin up and spin down species are only degenerate precisely at the Heisenberg isotropic point. Furthermore, it is not possible to write the spin operators $S^{\pm}$ as local bilinears in those fermions.

One may also argue that the states of the spin system can be viewed as a collection of bosons with hard cores: a spin can only be flipped once. The algebra of the Pauli matrices, on the other hand, seems to have mixed properties: they *commute* on different sites and they *anticommute* on the same sites. The anticommutivity of the Pauli matrices guarantees that the bosons have indeed hard cores.

More formally, let us imagine that we are going to use a set of basis vectors in which $S_z \equiv S_3$ is diagonal. We can also consider the raising and lowering operators, at each site n, $S^{\pm}(n)$

$$S^{\pm}(n) = S_1(n) \pm iS_2(n) \tag{4.2.1}$$

where I am using the notation

$$S_i \equiv \frac{1}{2}\sigma_i \qquad i = 1,2,3 \tag{4.2.2}$$

and the $\sigma_i's$ are the three Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.2.3}$$

The operators $S^{\pm}(n)$ *commute* on different sites

$$\left[S^{+}(n), S^{+}(m)\right] = \left[S^{-}(n), S^{-}(m)\right] = \left[S^{+}(n), S^{-}(m)\right] = 0 \tag{4.2.4}$$

for $m \neq n$. But on the same sites they anticommute

$$\{S^{+}(n), S^{-}(n)\} = 1 \tag{4.2.5}$$

$$\{S^{+}(n), S^{+}(n)\} = \{S^{-}(n), S^{-}(n)\} = 0. \tag{4.2.6}$$

This last condition implies that, if $|F\rangle$ is an arbitrary state *not annihilated* by $S^{+}(n)$, then it is annihilated by $S^{+}(n)^2$

$$S^{+}(n)\left[S^{+}(n)|F\rangle\right] = 0. \tag{4.2.7}$$

In other words, $S^{+}(n)$ creates bosonic excitation at the n-th site but it is not possible to have two such excitations at the same site. This is the hard sphere condition.

Consider now the *kink* or *soliton* operators $K(n)$

$$K(n) = \exp\left(i\pi \sum_{j=1}^{n-1} S^{+}(j)S^{-}(j)\right). \tag{4.2.8}$$

In terms of $S_3(n)$ we can write

$$K(n) = \exp\left(i\pi \sum_{j=1}^{n-1}\left(S_3(j) + \frac{1}{2}\right)\right) \equiv i^{n-1} \exp\left(i\pi \sum_{j=1}^{n-1} S_3(j)\right). \tag{4.2.9}$$

Thus $K(n)$ is a unitary operator which, up to a phase factor, rotates the spin configurations by π around the z-axis on all sites to the left of the n-th site. Thus the state $|\frac{1}{2}\ldots\frac{1}{2}\rangle$, an eigenstate of S_1 on all sites, becomes

$$K(n)|\frac{1}{2}\ldots\frac{1}{2}\rangle = i^{n-1}|-\frac{1}{2}\ldots-\frac{1}{2},\frac{1}{2}\ldots\frac{1}{2}\rangle \tag{4.2.10}$$

where the last flipped spin is at the site $n-1$. The operator $K(n)$ is said to create a *kink* in the spin configuration. Clearly this operator cannot have a non-vanishing expectation value in any state exhibiting long-range-order. On

the other hand, it may have an expectation value on states without long-range-order. For this reason these operators are usually called disorder operators [Kadanoff 71] and [Fradkin 78]. Consider now the operators $c^\dagger(n)$ and $c(n)$ obtained from flipping a spin and creating a kink at the same place [Jordan 28]

$$\begin{aligned} c(n) &\equiv K(n)S^-(n) = e^{i\pi\sum_{j=1}^{n-1} S^+(j)S^-(j)}S^-(n), \\ c^\dagger(n) &\equiv S^+(n)K^\dagger(n) = S^+(n)e^{-i\pi\sum_{j=1}^{n-1} S^+(j)S^-(j)}. \end{aligned} \tag{4.2.11}$$

The following results are easy to prove [Lieb 61].

Firstly

$$\begin{aligned} c^\dagger(n)c(n) &= S^+(n)K^\dagger(n)K(n)S^-(n), \\ c(n)c^\dagger(n) &= K(n)S^-(n)S^+(n)K^\dagger(n). \end{aligned} \tag{4.2.12}$$

But, the kink operator is unitary

$$K^\dagger(n)K(n) = K(n)K^\dagger(n) = 1 \tag{4.2.13}$$

and because $S^\pm(n)$ and $K(n)$ commute, one finds

$$\begin{aligned} c^\dagger(n)c(n) &= S^+(n)S^-(n) = \frac{1}{2} + S_3(n), \\ c(n)c^\dagger(n) &= S^-(n)S^+(n) = \frac{1}{2} - S_3(n). \end{aligned} \tag{4.2.14}$$

Moreover, the hard-core condition $(S^\pm)^2 = 0$ implies that the same property holds for the $c's$

$$\left(c^\dagger(n)\right)^2 = (c(n))^2 = 0. \tag{4.2.15}$$

What are the commutation relations obeyed by the operators $c^\dagger(n), c(m)$? Let us compute the products $c(n)c(m)$ and $c(m)c(n)$, say for $m > n$. Clearly $S^-(n)$ commutes with all the operators in $K(m)$ except for those at the site $j = n$ and therefore

$$S^-(n)K(m) = \prod_{j=1, j\neq n}^{m-1} e^{i\pi S^+(j)S^-(j)}S^-(n)e^{i\pi S^+(n)S^-(n)}. \tag{4.2.16}$$

By making use of the identity

$$e^{\pm i\pi S^+(n)S^-(n)} = e^{\pm i\pi\left(\frac{1}{2}+S_3(n)\right)} = -2S_3(n) \tag{4.2.17}$$

we get

$$S^-(n)K(m) = -K(m)S^-(n) \tag{4.2.18}$$

since $\{S^-(n), S_3(n)\} = 0$ on the same site. Thus

$$\begin{aligned} c(n)c(m) &= K(n)S^-(n)K(m)S^-(m) \\ &= -K(n)K(m)S^-(n)S^-(m) \\ &= -K(m)S^-(m)K(n)S^-(n) \\ &= -c(m)c(n). \end{aligned} \tag{4.2.19}$$

Similarly we can also prove ($n \neq m$)

$$\begin{aligned} c^\dagger(n)c(m) &= S^+(n)K^\dagger(n)K(m)S^-(m) \\ &= -K(m)S^-(m)S^+(n)K^\dagger(n) \\ &= -c(m)c^\dagger(n). \end{aligned} \tag{4.2.20}$$

In summary, the operators $c^\dagger(n)$ and $c(n)$ obey canonical anticommutation relations

$$\{c(n), c(m)\} = \{c^\dagger(n), c^\dagger(m)\} = 0 \tag{4.2.21}$$

and

$$\{c(n), c^\dagger(m)\} = \delta_{n,m}. \tag{4.2.22}$$

Thus the operator $c^\dagger(n)$ ($c(n)$) creates (destroys) a fermion at site n. These operators are highly non-local. The states created by $c^\dagger(n)$ are fermions. Conversely we can also write the inverse of the Jordan-Wigner transformation

$$\begin{aligned} S^-(n) &= e^{-i\pi \sum_{j=1}^{n-1} c^\dagger(j)c(j)} c(n), \\ S^+(n) &= c^\dagger(n) e^{i\pi \sum_{j=1}^{n-1} c^\dagger(j)c(j)}. \end{aligned} \tag{4.2.23}$$

4.2.2 The Heisenberg Chain: Fermion Picture

Let us apply these results to the Heisenberg model. In terms of S^+ and S^-, the Heisenberg Hamiltonian (with anisotropy γ) is

$$\begin{aligned} H = &\frac{1}{2} J \sum_{j=1}^{N} \left(S^+(j)S^-(j+1) + S^-(j)S^+(j+1)\right) + \\ &+ \gamma J \sum_{j=1}^{N} \left(S^+(j)S^-(j) - \frac{1}{2}\right)\left(S^+(j+1)S^-(j+1) - \frac{1}{2}\right). \end{aligned} \tag{4.2.24}$$

For $\gamma = 1$ we recover the isotropic Heisenberg model. The case $\gamma = 0$ is known as the spin one-half XY model.

We can now use the Jordan-Wigner transformation, Eq. (4.2.23), to get

$$\begin{aligned} S^+(j)S^-(j+1) &= c^\dagger(j)e^{-i\pi c^\dagger(j)c(j)}c(j+1) \\ &= c^\dagger(j)\left(1 - 2c^\dagger(j)c(j)\right)c(j+1) \\ &= c^\dagger(j)c(j+1) \end{aligned} \tag{4.2.25}$$

and

$$\begin{aligned} S^-(j)S^+(j+1) &= c(j)e^{+i\pi c^\dagger(j)c(j)}c^\dagger(j+1) \\ &= c(j)\left(1-2c^\dagger(j)c(j)\right)c^\dagger(j+1) \\ &= c(j)c^\dagger(j+1) - 2c(j)c^\dagger(j)c(j)c^\dagger(j+1) \\ &= c(j+1)c(j). \end{aligned} \tag{4.2.26}$$

The Heisenberg Hamiltonian takes the simple form [Luther 75]

$$H = \frac{J}{2}\sum_{j=1}^{N}\left(c^\dagger(j)c(j+1)+\text{h.c.}\right) + \gamma J\sum_{j=1}^{N}\left(n(j)-\frac{1}{2}\right)\left(n(j+1)-\frac{1}{2}\right) \tag{4.2.27}$$

where $n(j)$ is the density (or occupation number) for spinless fermions

$$n(j) = c^\dagger(j)c(j). \tag{4.2.28}$$

What boundary conditions do the $c(j)$ operators obey? Suppose that the *spin* problem has periodic boundary conditions, i.e.

$$S_i(N+1) = S_i(1) \quad \text{for } i = 1,2,3. \tag{4.2.29}$$

In the fermion case, the periodic boundary conditions on the spin degrees of freedom imply

$$\begin{aligned} c(N+1) &= e^{i\pi\sum_{j=1}^{N}S^+(j)S^-(j)}S^-(N+1) \\ &= e^{i\pi\sum_{j=1}^{N}[\frac{1}{2}+S_3(j)]}S^-(1) \end{aligned} \tag{4.2.30}$$

where

$$c(1) \equiv S^-(1). \tag{4.2.31}$$

Thus, the boundary condition on the fermionic degrees of freedom is

$$c(N+1) = i^N e^{i\pi S_3}c(1) \tag{4.2.32}$$

where S_3 is the total z-component of the spin. But, $\sum_{j=1}^{N}S^+(j)S^-(j)$ is just the total fermion number N_F so that S_3 and N_F are related by

$$S_3 = \sum_{j=1}^{N}c^\dagger(j)c(j) - \frac{N}{2} = N_F - \frac{N}{2}. \tag{4.2.33}$$

Hence, the $S_3 = 0$ sector maps into the half-filled sector for the fermions under the Jordan-Wigner transformation:

$$S_3 = 0 \Rightarrow N_F = \frac{N}{2}, \tag{4.2.34}$$

provided that N is even. Conversely, the state with $S_3 = \frac{1}{2}$ has $N_F = \frac{N+1}{2}$ provided that N is odd. The boundary condition, Eq. (4.2.32), depends on the z-component of the total spin S_3 or, alternatively, on the total number of fermions N_F

$$c(N+1) = e^{i\pi N_F}c(1). \tag{4.2.35}$$

For a lattice with N even and $S_3 = 0$ (i.e. $N_F = \frac{N}{2}$) we get *periodic (antiperiodic) boundary conditions* if $\frac{N}{2}$ is even (odd). Thus the many-body fermion wave functions obey different boundary conditions depending on whether N_F is even or odd.

The Hamiltonian, Eq. (4.2.27), has quadratic terms and is not readily solvable except, of course, by Bethe's method. We can gain some insight by considering the case $\gamma = 0$, the XY model.

For $\gamma = 0$, the Hamiltonian is simply

$$H_0 = \frac{J}{2} \sum_{j=1}^{N} \left(c^\dagger(j) c(j+1) + \text{h.c} \right). \tag{4.2.36}$$

This is a trival problem. *The fermions are free.* As we saw before, this problem can be solved by Fourier transform. Let $c(k)$ denote the Fourier modes, with $|k| \leq \pi$. The eigenvalues for a system with periodic boundary conditions are

$$H_0 = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \epsilon(k) c^\dagger(k) c(k) \tag{4.2.37}$$

where

$$\epsilon(k) = J \cos k. \tag{4.2.38}$$

The ground state is found by filling up the negative energy modes. In the case of $N_F = \frac{N}{2}$, we get two Fermi points, $k_F = \pm\frac{\pi}{2}$. The negative energy states have k in the interval $\pi > |k| \geq \frac{\pi}{2}$.

This system is *gapless.* In fact there are no massive excitations in the one-dimensional spin one-half spin chain. This system is *critical* in the sense that *all* its correlation functions fall off as a power of the distance. We will discuss this issue below. Also, there is no long-range-order in the sense that (at equal times)

$$\lim_{|m-n| \to \infty} \langle S^+(n) S^-(m) \rangle \approx (-1)^{m-n} \frac{\text{const}}{|m-n|^\eta} \longrightarrow 0 \tag{4.2.39}$$

with an exponent η that will be computed below. Thus there is no Néel order for the chain (Kennedy, Lieb and Shastri [Kennedy 88] have shown that for the square lattice the spin one-half XY model *has* long range order $\langle S^+ \rangle \neq 0$).

4.2.3 Continuum Limit

We are interested in the physics at large distances compared with the lattice constant and at frequencies much lower than, say, J. In this limit some sort of continuum theory should emerge. We will see now that the continuum theory associated with this one-dimensional system of fermions looks like a theory of "relativistic" fermions moving at the speed of "light" (with $c = v_F$, the Fermi velocity). These results apply not only to the Hamiltonian of Eq. (4.2.27) but, in fact, to *all* one-dimensional Fermi systems with local hopping Hamiltonians. A similar situation developes for fermions in a flux phase in two dimensions as we will see in chapter 6.

Consider first the non-interacting problem

$$H_0 = \frac{J}{2}\sum_{n=1}^{N}\left(c^\dagger(n)c(n+1) + \text{h.c.}\right) \tag{4.2.40}$$

which is equivalent to the XY model. We are assuming periodic boundary conditions. The dispersion law for this system is

$$\epsilon(k) = J\cos k \tag{4.2.38}$$

with Fermi points at $k_F = \pm\frac{\pi}{2}$. The elementary excitations will have a characteristic momentum of $\pm k_F$ and we should expect that the correlation functions of the fermions should have a rapid variation of the type $e^{ik_F n} = i^n$ with a *slow* variation on top. It is then natural to define new fermionic variables $a(n)$ which should only exhibit a slow variation in n and hence should have a simple continuum limit. Define

$$a(n) = i^{-n}c(n). \tag{4.2.41}$$

The Hamiltonian H_0 now reads

$$\begin{aligned} H_0 &= \frac{J}{2}\sum_{n=1}^{N}\left(i^{-n}a^\dagger(n)i^{(n+1)}a(n+1) + \text{h.c.}\right) \\ &= \frac{J}{2}\sum_{n=1}^{N}\left(ia^\dagger(n)a(n+1) + \text{h.c.}\right) \\ &= \frac{J}{2}\sum_{n=1}^{N} ia^\dagger(n)\left[a(n+1) - a(n-1)\right]. \end{aligned} \tag{4.2.42}$$

where we have used the periodic boundary conditions in the last step. By separating the sum into even and odd sites, one finds for N even

$$H_0 = \frac{J}{2}\sum_{s=1}^{N/2} i\{a^\dagger(2s)[a(2s+1) - a(2s-1)] + a^\dagger(2s+1)[a(2s+2) - a(2s)]\}. \tag{4.2.43}$$

We see that even sites couple to odd sites (and vice versa) but there is no even-even or odd-odd coupling.

Define now the *spinor* field $\phi_\alpha(\alpha = 1, 2)$, by

$$\phi_\alpha(n) = \begin{cases} \phi_1(n) = a(2s) & \text{on even sites,} \\ \phi_2(n) = a(2s+1) & \text{on odd sites.} \end{cases} \tag{4.2.44}$$

Thus we can write

$$\begin{aligned} H_0 = i\frac{J}{2}\sum_{s=1}^{N/2}\{&\phi_1^\dagger(2s)[\phi_2(2s+1) - \phi_2(2s-1)] + \\ &+\phi_2^\dagger(2s+1)[\phi_1(2s+2) - \phi_1(2s)]\}. \end{aligned} \tag{4.2.45}$$

A Fermi field $\psi_\alpha(x)$ in the continuum is expected to obey the equal- time canonical anticommutation relations

$$\{\psi_\alpha^\dagger(x), \psi_{\alpha'}(x')\} = \delta_{\alpha\alpha'}\delta(x - x'). \tag{4.2.46}$$

The $\phi_\alpha(n)$ fields obey

$$\{\phi_\alpha^\dagger(n), \phi_{\alpha'}(n')\} = \delta_{\alpha\alpha'}\delta_{n,n'} \tag{4.2.47}$$

since they are defined on a lattice. We can make these relations compatible by defining

$$\psi_\alpha(x) = \frac{1}{\sqrt{2a_0}}\phi_\alpha(n) \tag{4.2.48}$$

for $x = 2sa_0$ and a_0 the lattice spacing, which will be the unit of length. Thus ψ_α has dimensions of $[\text{length}]^{-\frac{1}{2}}$ whereas $\phi_\alpha$is dimensionless. We have assumed that the *distribution* $\delta(x - x')$ is defined by the limit

$$\delta(x - x') = \lim_{a_0 \to 0} \frac{\delta_{n,n'}}{2a_0} \tag{4.2.49}$$

which, of course, makes sense only as a limit.

By expanding ϕ, in Eq. (4.2.45), in a Taylor series expansion:

$$\begin{aligned} \phi_2(2s+1) - \phi_2(2s-1) &\approx 2a_0(2a_0)^{1/2}\partial_x\psi_2(x), \\ \phi_1(2s+2) - \phi_1(2s) &\approx 2a_0(2a_0)^{1/2}\partial_x\psi_1(x), \end{aligned} \tag{4.2.50}$$

and using the fact that

$$\lim_{a_0 \to 0} \sum_s 2a_0 f(s) = \int dx f(x), \tag{4.2.51}$$

one gets the effective Hamiltonian in the continuum $\tilde{H}_0$ to be given by

$$\tilde{H}_0 = \int dx \psi^\dagger(x)\alpha i\partial_x\psi(x) \tag{4.2.52}$$

where

$$\tilde{H}_0 = \frac{H_0}{Ja_0} \tag{4.2.53}$$

and the matrix

$$\alpha \equiv \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.2.54}$$

This is just the Hamiltonian for a Dirac spinor field $\psi_\alpha(x)$ in units in which $\hbar = 1$ and the Fermi velocity $v_F = 1$. The upper (lower) component of ψ_α represents the amplitude on even (odd) sites. Alternatively we could have used a basis in which σ_1 is diagonal. In this basis, the upper (lower) component $R(L)$ represents fermions moving toward the right (left) with speed $v_F = 1$. It will be, in fact, more convenient to work in this so-called *chiral* basis.

$$\begin{aligned} \psi_1(x) &= \frac{1}{\sqrt{2}}\left(-R(x) + L(x)\right), \\ \psi_2(x) &= \frac{1}{\sqrt{2}}\left(\ \ R(x) + L(x)\right). \end{aligned} \tag{4.2.55}$$

We get

$$\psi_1^\dagger i\partial_x\psi_2 + \psi_2^\dagger i\partial_x\psi_1 = -(R^\dagger i\partial_x R - L^\dagger i\partial_x L). \tag{4.2.56}$$

In the Dirac theory in (1+1) dimensions one defines the γ-matrices γ_0, γ_1 and γ_5 by requiring that they satisfy

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu} \quad \text{and} \quad \gamma_5 = i\gamma_0\gamma_1. \tag{4.2.57}$$

We can choose the chiral representation, in which

$$\begin{aligned} \gamma_5 &= \gamma_0\gamma_1 = \sigma_3, \\ \gamma_0 &= \sigma_1, \\ \gamma_1 &= -i\sigma_2. \end{aligned} \tag{4.2.58}$$

It is convenient to define a field $\bar{\psi}$ by

$$\bar{\psi} = \psi^\dagger\gamma_0. \tag{4.2.59}$$

The Hamiltonian $\tilde{H}_0$ now is

$$\tilde{H}_0 = \int dx\bar{\psi}(x)i\gamma_1\partial_x\psi(x). \tag{4.2.60}$$

Let us write the interaction terms of Eq. (4.2.27) in this formalism.

First, we note that we can rewrite

$$H_{\text{int}} = \gamma J\sum_{j=1}^{N}\left(c^\dagger(j)c(j) - \frac{1}{2}\right)\left(c^\dagger(j+1)c(j+1) - \frac{1}{2}\right) \tag{4.2.61}$$

in the form

$$H_{\text{int}} = -\frac{\gamma J}{2}\sum_{j=1}^{N}\left(c^\dagger(j)c(j) - c^\dagger(j+1)c(j+1)\right)^2 + \frac{1}{4}\gamma JN. \tag{4.2.62}$$

Following the same steps which led to Eq. (4.2.60), we find that $\tilde{H}_{\text{int}}$, defined by

$$\tilde{H}_{\text{int}} = \frac{H_{\text{int}}}{Ja_0} \tag{4.2.63}$$

has the form, up to the irrelevant additive constant $\frac{\gamma N}{4a_0}$,

$$\tilde{H}_{\text{int}} = -2\gamma\int dx\left(\bar{\psi}(x)\psi(x)\right)^2 \tag{4.2.64}$$

which is usually referred to as a Gross-Neveu interaction. The expression $\bar{\psi}\psi$ is the continuum limit of

$$\begin{aligned} \frac{1}{2a_0}(n(2s+1) - n(2s)) &\approx -\left(\psi_1^\dagger(x)\psi_1(x) - \psi_2^\dagger(x)\psi_2(x)\right) \\ &= \left(R^\dagger L + L^\dagger R\right) \equiv \bar{\psi}\psi. \end{aligned} \tag{4.2.65}$$

Thus a non-zero average for $\bar{\psi}\psi$ breaks chiral (i.e. left-right) symmetry. We see that this is equivalent to the development of a periodic density modulation

of the lattice fermion system. Tracing our steps backwards, we interpret this state as an antiferromagnet.

Equation (4.2.64) can also be written in the form, up to an additive constant,

$$\tilde{H}_{\text{int}} = \gamma \int dx j_\mu j^\mu - 2\gamma \int dx \left((R^\dagger L)^2 + (L^\dagger R)^2 \right) \tag{4.2.66}$$

where we have used the fermionic current j_μ

$$j_\mu = \bar{\psi} \gamma_\mu \psi \tag{4.2.67}$$

which, in the chiral basis, has components

$$j_0 = R^\dagger R + L^\dagger L \tag{4.2.68}$$

and

$$j_1 = R^\dagger R - L^\dagger L. \tag{4.2.69}$$

Thus j_0 measures the total number of fermions, i.e. the total density, and j_1 is the difference of the number of left and right movers. A system with the first term of Eq. (4.2.66) as its only interaction is known as the (massless) Thirring or Luttinger model.

The last term in Eq. (4.2.66) is peculiar. On the one hand, it appears to be superficially zero, since it is a sum of squares of Fermi fields and Fermi statistics may seem to imply that it is zero. However, all these expressions, written in the continuum, are to be interpreted as a product of operators at short distances. Furthermore, when inserted in the calculation of any expectation value, there should be singular contributions due to the presence of this operator. We are supposed to keep the leading singular term in the product. Thus, an expression such as $(\bar{\psi}\psi)^2$ and the like are to be taken in the sense of an *operator product expansion* [Kadanoff 69] and [Wilson 69] in which only the leading singularity is kept.

What is more important the operators $(R^\dagger L)^2$ and $(L^\dagger R)^2$ break the continuous left-right (chiral) symmetry down to a discrete subgroup. Terms of this sort arise from Umklapp scattering processes [Emery 79] [Haldane 82]. In the language of Feynman diagrams, these terms give contributions of the type shown in Fig. 4.1. Such processes violate momentum conservation by $4k_F$, which equals 2π for a half-filled system. Thus $4k_F$ is a reciprocal lattice vector and hence the process is allowed, since on a lattice momentum is conserved mod 2π.

There is a *continuous* chiral symmetry

$$\psi_\alpha = \left(e^{i\gamma_5 \theta} \right)_{\alpha\beta} \psi'_\beta \tag{4.2.70}$$

where γ_5 is given by Eq. (4.2.58), and θ is an arbitrary *constant* angle.

It is easy to check that operators, such as the current $\bar{\psi}\gamma_\mu\psi$ and $\bar{\psi} i \gamma^\mu \partial_\mu \psi$ are invariant. Indeed upon a chiral transformation, $\bar{\psi}$ transforms like

$$\bar{\psi} = \psi^\dagger \gamma_0 = \psi'^\dagger e^{-i\gamma_5\theta} \gamma_0 = \bar{\psi}' e^{+i\gamma_5\theta} \tag{4.2.71}$$

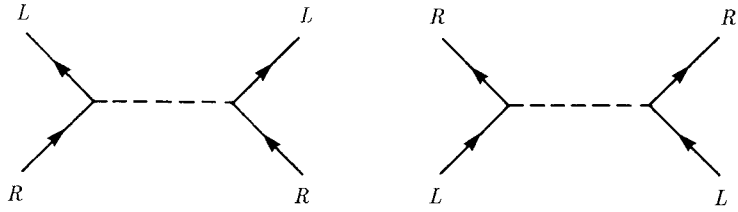


Figure 4.1 Umklapp processes.

since γ_0 and γ_5 anticommute. Thus

$$\bar{\psi}\gamma_\mu\psi = \bar{\psi}'\gamma_\mu\psi' \tag{4.2.72}$$

and

$$\bar{\psi}i\gamma_\mu\partial^\mu\psi = \bar{\psi}'i\gamma_\mu\partial^\mu\psi' \tag{4.2.73}$$

again, since $\{\gamma_5, \gamma_\mu\} = 0$. However $\bar{\psi}\psi$ is not invariant since

$$\bar{\psi}\psi = \bar{\psi}'e^{i2\gamma_5\theta}\psi' \equiv \cos(2\theta)\bar{\psi}'\psi' + i\sin(2\theta)\bar{\psi}'\gamma_5\psi'. \tag{4.2.74}$$

In particular $(\bar{\psi}\psi)^2$ only has the discrete invariance

$$\bar{\psi}\psi = -\bar{\psi}'\psi'. \tag{4.2.75}$$

In other words $\theta = \frac{\pi}{2}$. This is not so surprising. The chiral symmetry originates from the two-sublattice structure. There is always an arbitrariness in how we choose a given sublattice. Thus the discrete symmetry is genuine, but the continuous symmetry is a consequence of a carelessly taken continuum limit.

4.3 Bosonization

We are now going to discuss some subtle but very important properties of one-dimensional Fermi systems. To date, these properties are known not to generalize to higher dimensions. Some superficially similar ideas have been recently discussed in the context of "anyon superfluids" (which we will discuss later). The physics is quite different, though.

A very important tool for the understanding of one-dimensional Fermi systems is the *bosonization transformation*. In its Abelian form this transformation was first discussed by F. Bloch and S. Tomonaga [Bloch 34]. It was rediscovered by Mattis and Lieb [Lieb 65] in the 60's and by Coleman [Coleman 75], Luther [Luther 75] and Mandelstam [Mandelstam 75] in the 70's. Witten [Witten 84] solved the non-Abelian problem in 1984. We will only consider the Abelian case.

Let us consider first the non-interacting theory with Hamiltonian H_0 given by

$$H_0 = \int dx \psi^\dagger i\alpha \partial_x \psi \tag{4.3.1}$$

where $\alpha = \gamma_5$, with canonically quantized Fermi fields, i.e.

$$\begin{aligned} \{\psi^\dagger_\alpha(x), \psi_{\alpha'}(x')\} &= \delta_{\alpha\alpha'}\delta(x-x'), \\ \{\psi_\alpha(x), \psi_{\alpha'}(x')\} &= \{\psi^\dagger_\alpha(x), \psi^\dagger_{\alpha'}(x')\} = 0 \end{aligned} \tag{4.3.2}$$

at equal times. The Hamiltonian H_0 and the canonical anticommutation relations follow from canonical quantization (for fermions!) of the system with Lagrangian density

$$\mathcal{L}_0 = \bar{\psi} i\gamma^\mu \partial_\mu \psi = \bar{\psi} i\gamma^0 \partial_0 \psi - \bar{\psi} i\gamma^1 \partial_1 \psi \tag{4.3.3}$$

which has the Dirac form. All along I have assumed that the metric tensor $g_{\mu\nu}$ is

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.3.4}$$

This Lagrangian density is clearly invariant under global continuous chiral transformations. In fact, the Hamiltonian, in the chiral basis, is

$$H_0 = -\int dx (R^\dagger i\partial_x R - L^\dagger i\partial_x L) \tag{4.3.5}$$

which implies that the right (left) moving component $R(L)$ moves towards the right (left) at speed 1 (in units in which $v_F = 1$).

4.3.1 Anomalous Commutators

Consider now the "vacuum states" $|0\rangle$ and $|G\rangle$, where $|0\rangle$ is the *empty* state and $|G\rangle$ is the *filled Fermi sea* obtained by having occupied all the negative energy one-particule eigenstates of the Hamiltonian Eq. (4.3.5). The Hamiltonian H_0 relative to both vacua differs by normal ordering terms. Indeed, for any eigenstate $|F\rangle$ of H_0 one can write

$$H_0 =: H_0 : + E_F |F\rangle\langle F| \tag{4.3.6}$$

where $: H_0 :$ is the Hamiltonian normal ordered with respect to $|F\rangle$, i.e.

$$: H_0 : |F\rangle = \langle F| : H_0 := 0 \tag{4.3.7}$$

and E_F is the energy of $|F\rangle$

$$H_0|F\rangle = E_F|F\rangle. \tag{4.3.8}$$

Clearly, if we choose $|0\rangle$ or $|G\rangle$ as the reference state, E_F will be different.

The currents and densities also need to be normal ordered. This is equivalent to the subtraction of the (infinite) background charge of the reference state, say of the filled Fermi sea. We will see that these apparently "formal" manipulations have a profound effect on the physics.

Let us compute the commutator of the charge density and current operators at equal times $[j_0(x), j_1(x')]$. Relative to the empty state $|0\rangle$, both operators are already normal ordered since a state with no fermions has neither charge nor current, i.e.

$$\begin{aligned} j_0(x)|0\rangle &= 0, \\ j_1(x)|0\rangle &= 0. \end{aligned} \tag{4.3.9}$$

It will be useful to consider the right and left components of the current $j_\pm$ defined by

$$j_\pm = \frac{1}{2}(j_0 \pm j_1). \tag{4.3.10}$$

Clearly, we get that

$$j_+ = R^\dagger R \tag{4.3.11}$$

is the right moving current, and

$$j_- = L^\dagger L \tag{4.3.12}$$

is the left-moving current. In Fourier components, we find

$$j_+(p) = \frac{1}{\sqrt{L_0}} \sum_k R^\dagger(k) R(k+p) \tag{4.3.13}$$

which annihilates the empty state $|0\rangle$. In fact, for any state $|\phi\rangle$ with a *finite* number of particles, the result is

$$[j_\pm(p), j_\pm(p')]|\phi\rangle = 0. \tag{4.3.14}$$

Consider now the filled Fermi sea, $|G\rangle$. Explicitly we can write

$$|G\rangle = \prod_{p<0} R^\dagger(p) \prod_{q>0} L^\dagger(q)|0\rangle. \tag{4.3.15}$$

In other words, in $|G\rangle$ all right moving states with negative momentum and all left moving states with positive momentum are filled (see Fig. (4.2)).

Let us compute the commutator $[j_+(x), j_+(x')]$ at equal times (see, for instance, I. Affleck [Affleck 86]). The operator $j_+(x)$ is formally equal to a product of fermion operators at the same point. Since we anticipate divergencies, we should "point-split" the product

$$j_+(x) = R^\dagger(x)R(x) = \lim_{\epsilon \to 0} R^\dagger(x+\epsilon)R(x-\epsilon) \tag{4.3.16}$$

and write j_+ in terms of a normal ordered operator $: j_+ :$ and a vacuum expectation value

$$j_+(x) =: j_+(x) : + \lim_{\epsilon \to 0} \langle G|R^\dagger(x+\epsilon)R(x-\epsilon)|G\rangle. \tag{4.3.17}$$

The singularities are absorbed in the expectation value.
Consider a system on a segment of length L_0 with periodic boundary conditions and expand $R(x)$ in Fourier series

$$R(x) = \frac{1}{\sqrt{L_0}} \sum_{p=-\infty}^{+\infty} R_p e^{i\frac{2\pi x p}{L_0}}. \tag{4.3.18}$$

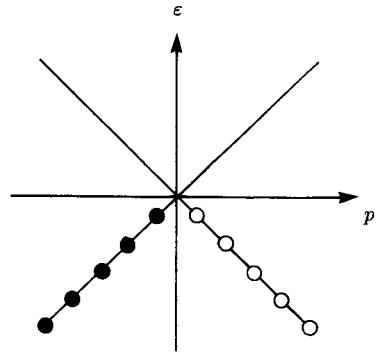


Figure 4.2 Vacuum $|G\rangle$ is obtained by filling the right moving states with negative momentum (filled cicles) and filling the left moving states with positive momentum (empty circles).

The vacuum expectation value to be computed is

$$\langle G|R^\dagger(x+\epsilon)R(x-\epsilon)|G\rangle = \frac{1}{L_0}\sum_{p,p'=-\infty}^{+\infty} e^{i\frac{2\pi}{L_0}[(x-\epsilon)p'-(x+\epsilon)p]}\langle G|R_p^\dagger R_{p'}|G\rangle. \tag{4.3.19}$$

Using the definition of the filled Fermi sea, we get

$$\langle G|R_p^\dagger R_{p'}|G\rangle = \delta_{p,p'}\theta(-p), \tag{4.3.20}$$

$$\langle G|L_p^\dagger L_{p'}|G\rangle = \delta_{p,p'}\theta(+p). \tag{4.3.21}$$

Hence

$$\langle G|R^\dagger(x+\epsilon)R(x-\epsilon)|G\rangle = \frac{1}{L_0}\sum_{p=-\infty}^{0} e^{-i\frac{2\pi p}{L_0}(2\epsilon)}. \tag{4.3.22}$$

This is a conditionally convergent series. In order to make it convergent, we will regulate this series by damping out the contributions due to states deep below the Fermi energy. We can achieve this if we analytically continue ϵ to the upper half of the complex plane (i.e. $\epsilon \equiv \epsilon + i\eta$) to get the convergent

expression

$$\begin{aligned}\langle G|R^\dagger(x+\epsilon)R(x-\epsilon)|G\rangle &= \lim_{\eta\to 0}\frac{1}{L_0}\sum_{p=0}^{\infty} e^{i\frac{\pi p}{L_0}(\epsilon+i\eta)} \\ &= \lim_{\eta\to 0}\frac{1}{L_0}\frac{1}{1-e^{i\frac{4\pi}{L_0}(\epsilon+i\eta)}} \\ &= \lim_{\eta\to 0}\frac{1}{L_0}\frac{1}{-i\frac{4\pi}{L_0}(\epsilon+i\eta)} \\ &= \frac{i}{4\pi\epsilon}.\end{aligned} \tag{4.3.23}$$

Thus, the result is

$$\langle G|R^\dagger(x+\epsilon)R(x-\epsilon)|G\rangle = \frac{i}{4\pi\epsilon}. \tag{4.3.24}$$

Similarly, the expectation value $\langle G|L^\dagger(x+\epsilon)L(x-\epsilon)|G\rangle$, is found to be given by

$$\langle G|L^\dagger(x+\epsilon)L(x-\epsilon)|G\rangle = \frac{-i}{4\pi\epsilon}. \tag{4.3.25}$$

The current commutator can now be readily evaluated

$$\begin{aligned}[j_+(x), j_+(x')] &= \lim_{\epsilon,\epsilon'\to 0}\left[R^\dagger(x+\epsilon)R(x-\epsilon), R^\dagger(x'-\epsilon')R(x'+\epsilon')\right] \\ &= \lim_{\epsilon,\epsilon'\to 0}\{\delta(x'-x+\epsilon'+\epsilon)R^\dagger(x+\epsilon)R(x'-\epsilon')+ \\ &\qquad -\delta(x-x'+\epsilon'+\epsilon)R^\dagger(x'+\epsilon')R(x-\epsilon)\}.\end{aligned} \tag{4.3.26}$$

The contributions from normal ordered products cancel (since they are regular). The only non-zero terms are, using Eq. (4.4.22),

$$[j_+(x), j_+(x')] = \lim_{\epsilon,\epsilon'\to 0}\left(\frac{i\delta(x'-x+\epsilon'+\epsilon)}{2\pi(x-x'+\epsilon+\epsilon')} - \frac{i\delta(x'-x+\epsilon+\epsilon')}{2\pi(x'-x+\epsilon+\epsilon')}\right). \tag{4.3.27}$$

Thus, in the limit we find

$$[j_+(x), j_+(x')] = -\frac{i}{2\pi}\partial_x\delta(x-x') \tag{4.3.28}$$

and

$$[j_-(x), j_-(x')] = +\frac{i}{2\pi}\partial_x\delta(x-x'). \tag{4.3.29}$$

In terms of Lorentz components, we get

$$[j_0(x), j_1(x')] = -\frac{i}{\pi}\partial_x\delta(x-x') \tag{4.3.30}$$

whereas

$$[j_0(x), j_0(x')] = [j_1(x), j_1(x')] = 0. \tag{4.3.31}$$

The commutator $[j_0(x), j_1(x')]$ has a non-vanishing right-hand side which is a c-number. These terms are generally known as *Schwinger terms*. They are

pervasive in theories of relativistic fermions. But terms of this sort are also found in non-relativistic systems of fermions at finite densities [Pines 66]. In fact, these terms are the key to the derivation of the f-sum rule.

4.3.2 The Bosonization Rules

We thus notice that the equal-time current commutator $[j_0(x), j_1(x')]$ acquires a Schwinger term if the currents and densities are normal ordered relative to the filled Fermi sea. The identity of Eq. (4.3.30) suggests that there should be a connection between a canonical Fermi field ψ with a filled Fermi sea and canonical Bose field ϕ. Let $\Pi(x)$ be the canonical momentum conjugate to ϕ, i.e. at equal times

$$[\phi(x), \Pi(x')] = i\delta(x - x'). \tag{4.3.32}$$

If we identify the normal ordered operators

$$j_0(x) = \frac{1}{\sqrt{\pi}}\partial_x\phi(x) \tag{4.3.33}$$

and

$$j_1(x) = -\frac{1}{\sqrt{\pi}}\partial_t\phi(x) \equiv -\frac{1}{\sqrt{\pi}}\Pi(x), \tag{4.3.34}$$

we see that Eq. (4.3.32) implies

$$\frac{1}{\pi}[\partial_x\phi(x), \Pi(x')] = -\frac{i}{\pi}\delta'(x - x') \tag{4.3.35}$$

which is consistent with the Schwinger term. These equations can be written in the more compact form

$$j_\mu = \frac{1}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu\phi \tag{4.3.36}$$

where $\epsilon_{\mu\nu}$ is the (antisymmetric) Levi-Civita tensor and we are using *from now on* the notation $t \to x_0$, $x \to x_1$ and $x \equiv (x_0, x_1)$. We then arrive at the conclusion that the current commutator with a Schwinger term, Eq. (4.3.30), is equivalent to the statement that there exists a canonical Bose field ϕ whose *topological current*, Eq. (4.3.36), coincides with the normal ordered fermion current.

The fermion current j_μ is conserved, i.e.

$$\partial_\mu j^\mu = 0 \tag{4.3.37}$$

which is automatically satisfied by Eq. (4.3.36). In the case of the *free* theory, the number of left and right movers are separately conserved. This means that not only should j_μ be conserved, but j^5_μ, defined by

$$j^5_\mu = \bar{\psi}\gamma_\mu\gamma^5\psi \tag{4.3.38}$$

should also be conserved. Using the identity

$$\gamma_\mu\gamma_5 = \epsilon_{\mu\nu}\gamma^\nu \tag{4.3.39}$$

we see that j_μ and j_μ^5 are in fact related by

$$j_\mu^5 = \epsilon_{\mu\nu} j^\nu . \tag{4.3.40}$$

The divergence of j_μ^5 can be computed in terms of the Bose field ϕ as follows

$$\partial_\mu j^{5\mu} = \epsilon^{\mu\nu}\partial_\mu j_\nu = \frac{1}{\sqrt{\pi}}\epsilon^{\mu\nu}\epsilon_{\nu\lambda}\partial_\mu\partial^\lambda\phi = \frac{1}{\sqrt{\pi}}\partial^2\phi . \tag{4.3.41}$$

Thus, the conservation of the *axial* current j_μ^5 implies that ϕ should be a free canonical Bose field

$$\partial_\mu j^{5\mu} = 0 \;\Rightarrow\; \partial^2\phi = 0 \tag{4.3.42}$$

where

$$\partial^2 \equiv \partial_0^2 - \partial_1^2 . \tag{4.3.43}$$

The Lagrangian for these bosons is simply given by

$$\mathcal{L}_B = \frac{1}{2}\left(\partial_\mu\phi\right)^2 . \tag{4.3.44}$$

Conversely, if ϕ is not free j_μ^5 should not be conserved. We will see below that this is indeed what happens in the Thirring-Luttinger model.

Before doing that, let us consider a set of identities, originally derived by Mandelstam [Mandelstam 75]. By analogy with the Jordan-Wigner transformation of Section 4.2, we should expect that these identities should be highly non-local, although they should have local anti-commutation relations. These identities, like all others derived within the Bosonization approach, only make sense within the Operator Product Expansion: the operators so identified give rise to the same leading singular behavior when arbitrary matrix elements are computed. Also, from the Jordan-Wigner analogy, we should expect that the fermion operators, as seen from their representation in terms of bosons, should act like operators which create *solitons.*

The free Bose field ϕ can be written in terms of creation and annihilation operators. Let $\phi^+(x)$ ($\phi^-(x)$) denote the piece of $\phi(x)$ which depends on the creation (annihilation) operators only

$$\phi(x) = \phi^+(x) + \phi^-(x) \tag{4.3.45}$$

where $\phi(x)$ is a Heisenberg operator ($x \equiv (x_0, x_1)$, see Eq. (4.3.36)). Obviously, ϕ^- annihilates the vacuum of the Bose theory. The operators ϕ^+ and ϕ^- obey the commutation relations

$$[\phi^+(x_0,x_1), \phi^-(x_0',x_1')] = \lim_{\epsilon\to 0}\Delta_+(x_0 - x_0', x_1 - x_1') \tag{4.3.46}$$

where Δ_+ is given by

$$\Delta_+(x_0 - x_0', x_1 - x_1') = -\frac{1}{4\pi}\ln\left[(c\mu)^2((x_1 - x_1')^2 - (x_0 - x_0' + i\epsilon)^2)\right] . \tag{4.3.47}$$

The arbitrary constant μ has dimensions of mass (i.e. $(\text{length})^{-1}$) and it is necessary to make the argument of the logarithm dimensionless. It is customary to do this calculation by adding a small mass μ and to consider the limit

$|x_1 - x_1'| \ll \mu^{-1}$. In this case the numerical constant c is related to Catalan's constant.

Consider now the operators $\mathcal{O}_\alpha(x)$ and $\mathcal{Q}_\beta(x)$ defined by

$$\mathcal{O}_\alpha(x) = e^{i\alpha\phi(x)} \tag{4.3.48}$$

and

$$\mathcal{Q}_\beta(x) = e^{i\beta\int_{-\infty}^{x_1} dx_1'\partial_0\phi(x_0,x_1')} \equiv e^{i\beta\int_{-\infty}^{x_1} dx_1'\Pi(x_0,x_1')}. \tag{4.3.49}$$

When acting on a state $|\{\phi(x')\}\rangle$, $\mathcal{O}_\alpha(x)$ simply multiplies the state by $e^{i\alpha\phi(x)}$. The operator $\mathcal{Q}_\beta(x)$ has quite a different effect. Since $\Pi(x)$ and $\phi(x)$ are conjugate pairs, $\mathcal{Q}_\beta(x)$ will shift the value of $\phi(x_0,x_1')$ to $\phi(x_0,x_1')+\beta$ for all $x_1' < x_1$. Thus, $\mathcal{Q}_\beta(x)$ creates a coherent state which we can call a *soliton*

$$\mathcal{Q}_\beta(x)|\{\phi(x_0,x_1')\}\rangle = |\{\phi(x_0,x_1') + \beta\theta(x_1 - x_1')\}\rangle. \tag{4.3.50}$$

Consider now the operator $\psi_{\alpha,\beta}(x)$ of the form

$$\psi_{\alpha,\beta}(x) = \mathcal{O}_\alpha(x)\mathcal{Q}_\beta(x) = e^{i\alpha\phi(x)+i\beta\int_{-\infty}^{x_1} dx_1'\partial_0\phi(x_0,x_1')} \tag{4.3.51}$$

and compute the product $\psi_{\alpha,\beta}(x)\psi_{\alpha,\beta}(x')$ at equal times ($x_0' = x_0$). Using the Baker-Hausdorff formula

$$e^{\hat{A}}e^{\hat{B}} = e^{\hat{B}}e^{\hat{A}}e^{-[\hat{A},\hat{B}]} = e^{\hat{A}+\hat{B}-\frac{1}{2}[\hat{A},\hat{B}]} \tag{4.3.52}$$

where $[\hat{A},\hat{B}]$ is a complex valued distribution, we get

$$\psi_{\alpha,\beta}(x)\psi_{\alpha,\beta}(x') = \psi_{\alpha,\beta}(x')\psi_{\alpha,\beta}(x)e^{-i\Phi(x,x')} \tag{4.3.53}$$

where $\Phi(x,x')$ is given by (all the commutators are understood to be at equal times and $x_0 = x_0'$ but $x_1' \neq x_1$)

$$\begin{aligned} i\Phi(x,x') &= -\alpha^2[\phi(x),\phi(x')] - \beta^2\int_{-\infty}^{x_1} dy_1\int_{-\infty}^{x_1'} dy_1'[\Pi(y),\Pi(y')]+ \\ &- \alpha\beta\int_{-\infty}^{x'} dy_1'[\phi(x),\Pi(y')] - \alpha\beta\int_{-\infty}^{x_1} dy_1[\Pi(y),\phi(x')] \\ &= -i\alpha\beta. \end{aligned} \tag{4.3.54}$$

For the operators $\psi_{\alpha,\beta}(x)$ to have *fermion commutation relations* we need to choose $\alpha\beta = \pm\pi$. It is useful to write left and right components of the Fermi field in the form [Mandelstam 75]

$$R(x) = -i\left(\frac{c\mu}{2\pi}\right)^{1/2} e^{\frac{\mu}{8\epsilon}} : e^{-i\frac{2\pi}{\beta}\int_{-\infty}^{x_1} dx_1'\Pi(x_0,x_1')+\frac{i\beta}{2}\phi(x)} : \tag{4.3.55}$$

$$L(x) = \left(\frac{c\mu}{2\pi}\right)^{1/2} e^{\frac{\mu}{8\epsilon}} : e^{-i\frac{2\pi}{\beta}\int_{-\infty}^{x_1} dx_1'\Pi(x_0,x_1')-\frac{i\beta}{2}\phi(x)} : . \tag{4.3.56}$$

The phase factor and constants in front of the operators R and L are chosen so that γ-matrices have the canonical form. The constant β is arbitrary and it can be chosen by demanding that the currents satisfy the formula

$$j_\mu = \frac{1}{\sqrt{\pi}}\epsilon_{\mu\nu}\partial^\nu\phi. \tag{4.3.57}$$

Mandelstam found that Eqs. (4.3.55) and (4.3.56), imply that

$$j_\mu = \frac{\beta}{2\pi}\epsilon_{\mu\nu}\partial^\nu\phi. \tag{4.3.58}$$

Thus, we must choose $\beta = \sqrt{4\pi}$ for the fermion problem.

It is interesting to consider products of the form $\lim_{y_1\to x_1} R^\dagger(x)L(y)$ and $\lim_{y_1\to x_1} L^\dagger(y)R(x)$ at equal times. We will use Mandelstam's formulas to derive an operator product expansion for $R^\dagger L$ and $L^\dagger R$, both to leading order. We find

$$\lim_{y_1\to x_1} R^\dagger(x)L(y) = i\left(\frac{c\mu}{2\pi}\right) e^{\frac{\mu}{4\epsilon}} : e^{i\frac{2\pi}{\beta}\int_{-\infty}^{x_1} dx_1'\partial_0\phi(x_0,x_1') - \frac{i\beta}{2}\phi(x_0,x_1)} : \\ \times : e^{-\frac{i2\pi}{\beta}\int_{-\infty}^{y_1} dy_1'\partial_0\phi(x_0,y_1') - \frac{i\beta}{2}\phi(x_0,y_1)} : . \tag{4.3.59}$$

We can make use of the Baker-Hausdorff formula once again, now in the form

$$: e^{\hat{A}} :: e^{\hat{B}} := e^{[\hat{A}^+,\hat{B}^-]} : e^{\hat{A}+\hat{B}} : \tag{4.3.60}$$

and write down a *bosonic* expression for $R^\dagger L$. The normal ordered operator in Eq. (4.3.60) is, by definition, regular. Thus we can take the limit readily to find

$$\lim_{y\to x} : e^{\hat{A}+\hat{B}} :=: e^{-i\beta\phi(x)} : . \tag{4.3.61}$$

This operator is multiplied by a singular coefficient which compensates for the fact that $R^\dagger L$ and $e^{-i\beta\phi}$ have, in principle, different dimensions. An explicit calculation gives the asymptotic result

$$\lim_{y_1\to x_1} R^\dagger(x)L(y) = \lim_{y_1\to x_1} f(|x_1-y_1|) : e^{-i\beta\phi(x)} : \tag{4.3.62}$$

where the *singular* function $f(|x_1-y_1|)$ is computed as follows

$$f(|x_1-y_1|) = i\left(\frac{c\mu}{2\pi}\right) e^{\frac{\mu}{4\epsilon}} \exp\{-\pi\int_{-\infty}^{y_1} dy_1'[\phi^+(x),\partial_0\phi^-(y')]|_{y_0'=x_0} + \\ + \pi\int_{-\infty}^{x_1} dx_1'[\partial_0\phi^+(x'),\phi^-(y)]|_{y_0=x_0'=x_0} + \\ - \frac{\beta^2}{4}[\phi^+(x),\phi^-(y)]|_{y_0=x_0} + \\ + \frac{4\pi^2}{\beta^2}\int_{-\infty}^{x_1} dx_1'\int_{-\infty}^{y_1} dy_1'[\partial_0\phi^+(x'),\partial_0\phi^-(y')]|_{y_0'=x_0'=x_0}\}. \tag{4.3.63}$$

By inserting expressions for the commutators in Eq. (4.3.63) and, afterwards, evaluating the integrals, we get

$$f(|x_1-y_1|) = \frac{c\mu}{2\pi}(|x_1-y_1|)^\sigma \tag{4.3.64}$$

with an exponent σ given by

$$\sigma = \frac{\beta^2}{8\pi} - \frac{2\pi}{\beta^2}. \tag{4.3.65}$$

Thus, to leading order, we must identify the operator $R^\dagger L$ with

$$\lim_{|x_1-y_1|\approx a_0} R^\dagger(x_0,x_1)L(x_0,y_1) \approx \frac{c\mu}{2\pi}(c\mu a_0)^\sigma : e^{-i\beta\phi(x)} : \tag{4.3.66}$$

where a_0 is a short-distance cutoff. Similarly, one finds the identification

$$\lim_{|x_1-y_1|\approx a_0} L^\dagger(x_0,y_1)R(x_0,x_1) \approx \frac{c\mu}{2\pi}(c\mu a_0)^\sigma : e^{+i\beta\phi(x)} : . \tag{4.3.67}$$

To sum up, the order-parameter field $\bar\psi\psi$ at $\beta = \sqrt{4\pi}$ is given by

$$\lim_{|x_1-y_1|\approx a_0} \bar\psi(x_0,x_1)\psi(x_0,y_1) \approx \frac{c\mu}{\pi} : \cos(\sqrt{4\pi}\phi(x)) : \tag{4.3.68}$$

since for $\beta = \sqrt{4\pi}$, Eq. (4.3.65) yields $\sigma = 0$.

In the Ising regime of the Heisenberg model, we expect $\langle\bar\psi\psi\rangle$ to be different from zero and therefore the bosonic theory should have a ground state such that the expectation value $\langle\cos(\sqrt{4\pi}\phi\rangle$ is not zero. Under a chiral transformation by $\theta = \pi/2$, $\bar\psi\psi$ transforms like

$$\bar\psi\psi \to -\bar\psi\psi \tag{4.3.69}$$

which is equivalent to a sublattice exchange. In bosonic language, this transformation amounts to

$$\phi \to \phi + \frac{\pi}{\sqrt{4\pi}}. \tag{4.3.70}$$

The Umklapp operators play a crucial role here [Haldane 82] [Nijs 81] [Emery 79]. These operators enter in the interaction Hamiltonian, through terms of the form (see Eq. (4.2.66))

$$\int dx_1 \left((R^\dagger L)^2 + (L^\dagger R)^2\right). \tag{4.3.71}$$

These terms can be bosonized using the Mandelstam identities Eqs. (4.3.66) and (4.3.67). Indeed, we get the equal-time operator expansion

$$\begin{aligned}\lim_{|x_1-y_1|\approx a_0} (R^\dagger(x)L(y))^2 &\approx (\frac{c\mu}{2\pi})^2 : e^{-i\sqrt{4\pi}\phi(x)} :: e^{-i\sqrt{4\pi}\phi(y)} : \\ &\approx (\frac{c\mu}{2\pi})^2 : e^{-4\pi[\phi^+(x),\phi^-(y)]} :: e^{-i2\sqrt{4\pi}\phi(x)} : \\ &\approx (\frac{c\mu}{2\pi})^2 : e^{-4\pi\Delta_+(0^+,x_1-y_1)} :: e^{-i2\sqrt{4\pi}\phi(x)} :\end{aligned} \tag{4.3.72}$$

where Eqs. (4.3.60) and (4.3.46) have been used. In short, the bosonized version of the Umklapp terms is (at $\beta = \sqrt{4\pi}$)

$$\lim_{|x_1-y_1|\approx a_0} (R^\dagger(x)L(y))^2 \approx (\frac{c\mu}{2\pi})^2 (c\mu a_0)^2 : e^{-i4\sqrt{\pi}\phi(x)} : \tag{4.3.73}$$

and likewise

$$\lim_{|x_1-y_1|\approx a_0} (L^\dagger(y)R(x))^2 \approx (\frac{c\mu}{2\pi})^2 (c\mu a_0)^2 : e^{+i4\sqrt{\pi}\phi(x)} : . \tag{4.3.74}$$

4.3.3 The Sine-Gordon Theory

Now that we have done all the hard work and derived the necessary identities, we are in position to write down the bosonized form of the Lagrangian. The *fermionic* Lagrangian density (see Eqs. (4.3.3) and (4.2.66))

$$\mathcal{L}_F = \bar{\psi} i\gamma^\mu \partial_\mu \psi - \gamma(\bar{\psi}\gamma_\mu\psi)^2 + 2\gamma\left((R^\dagger L)^2 + (L^\dagger R)^2\right) \tag{4.3.75}$$

which we showed was equivalent to the Heisenberg model (in the continuum limit), is thus equivalent to a *bosonic* theory with Lagrangian density (see Eqs. (4.3.44), (4.3.57) and (4.3.73-74))

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{\gamma}{\pi}\epsilon_{\mu\nu}\partial^\nu\phi\epsilon^{\mu\lambda}\partial_\lambda\phi + 4\gamma\left(\frac{c\mu}{2\pi}\right)^2 (c\mu a_0)^2 : \cos(4\sqrt{\pi}\phi) : . \tag{4.3.76}$$

Using the identity

$$\epsilon_{\mu\nu}\epsilon^{\mu\lambda} = -\delta^\lambda_\nu \tag{4.3.77}$$

we can write

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\phi)^2 + \frac{\gamma}{\pi}(\partial_\mu\phi)^2 + 4\gamma\left(\frac{c\mu}{2\pi}\right)^2 (c\mu a_0)^2 : \cos(4\sqrt{\pi}\phi) : . \tag{4.3.78}$$

Thus, the interactions in the fermions give rise to (a) a rescaling of the Bose field ϕ and (b) a non-linear term.

This Lagrangian density can be brought into the canonical form by a simple rescaling of the field $\phi(x)$

$$\left(1 + \frac{2\gamma}{\pi}\right)^{\frac{1}{2}} \phi(x) \equiv \varphi(x). \tag{4.3.79}$$

If we define β by the expression

$$\beta^2 = \frac{16\pi}{1 + \frac{2\gamma}{\pi}} \tag{4.3.80}$$

we can write the Lagrangian in the Sine-Gordon form

$$\mathcal{L}_B = \frac{1}{2}(\partial_\mu\varphi)^2 + g : \cos(\beta\varphi) : \tag{4.3.81}$$

where g, the Sine-Gordon coupling constant, is given by

$$g \approx \frac{\gamma}{\pi^2 a_0^2} \tag{4.3.82}$$

up to a finite non-universal multiplicative constant determined by the short-distance cutoff (i.e. we have arbitrarily set $c\mu a_0 = 1$).

The rescaling of ϕ implies that the canonical momentum Π should also be rescaled so as to keep the form of the canonical commutation relations. Thus Π is scaled as

$$\Pi = \left(1 + \frac{2\gamma}{\pi}\right)^{-\frac{1}{2}} \partial_0\varphi. \tag{4.3.83}$$

The Mandelstam operators now read (see Eqs. (4.3.55-56))

$$\begin{aligned} R(x) &= -i\left(\frac{c\mu}{2\pi}\right)^{\frac{1}{2}} e^{\frac{\mu}{8\epsilon}} :e^{-i\frac{4\pi}{\beta}\int_{-\infty}^{x_1} dx_1' \partial_0\varphi(x_0,x_1')+i\frac{\beta}{4}\varphi(x)}: \\ L(x) &= \left(\frac{c\mu}{2\pi}\right)^{\frac{1}{2}} e^{\frac{\mu}{8\epsilon}} :e^{-i\frac{4\pi}{\beta}\int_{-\infty}^{x_1} dx_1' \partial_0\varphi(x_0,x_1')-i\frac{\beta}{4}\varphi(x)}: \end{aligned} \tag{4.3.84}$$

with β given by Eq. (4.3.80).

Similarly, the order parameter field $\bar{\psi}\psi$ now has the asymptotic behavior (see Eqs. (4.3.68) and (4.3.79-80))

$$\bar{\psi}(x)\psi(x) \approx \frac{c\mu}{\pi} :\cos\left(\frac{\beta}{2}\varphi\right): . \tag{4.3.85}$$

This formula will help us to determine the correlation function of the staggered longitudinal order parameter at long distances. We can also find bosonized expressions for the *transverse* components of the order parameter, i.e. $S^{\pm}(2s+1) - S^{\pm}(2s)$. The same procedure which led to the relation between $S_z(2s+1) - S_z(2s)$ and $\bar{\psi}\psi$, Eq. (4.2.65), now gives the correspondence

$$\begin{aligned} M^+(x) &\approx S^+(2s+1) - S^+(2s) \\ &\approx \sqrt{2a_0} e^{-i\pi\int_{-\infty}^{x_1} dx_1' :\psi^\dagger(x_0,x_1')\psi(x_0,x_1'):} \left(\psi_1^\dagger(x') - i\psi_2^\dagger(x)\right) \end{aligned} \tag{4.3.86}$$

which, in the chiral basis, has the form

$$M^+(x) \approx \sqrt{2a_0} e^{-i\pi\int_{-\infty}^{x_1} dx_1' :j_0(x_0,x_1'):} \left(R^\dagger(x)e^{-i\pi/4} + L^\dagger(x)e^{i\pi/4}\right) . \tag{4.3.87}$$

The other transverse component, M^-, is just the hermitian conjugate of M^+.

We can use the Bosonization "technology" to find an expression for M^+ in terms of the Bose field φ. The result is, up to singular coefficients,

$$\begin{aligned} M^+(x) \approx &: e^{i\frac{4\pi}{\beta}\int_{-\infty}^{x_1} dx_1' \partial_0\varphi(x_0,x_1') - \frac{i\beta}{4}\varphi(x_0,-\infty)} : \\ &\times : e^{i\frac{4\pi}{\beta}\int_{-\infty}^{x_1} dx_1' \partial_0\varphi(x_0,x_1') + \frac{i\beta}{2}\varphi(x_0,x_1) - \frac{i\beta}{4}\varphi(x_0,-\infty)} : . \end{aligned} \tag{4.3.88}$$

The Sine-Gordon potential $\cos(\beta\varphi)$ does not affect the behavior at long distances unless the operator is *relevant*, in the sense of the renormalization group. This means that the dimension Δ of this operator should be less than or equal to two, the dimension of space time. The dimension Δ_A of an operator $A(x)$ is found by considering the correlation function, say at equal times,

$$\langle A(x)A(x')\rangle \sim \frac{1}{|x_1 - x_1'|^{\eta_A}} . \tag{4.3.89}$$

The critical exponent η_A and the dimension Δ_A are related

$$\eta_A \equiv 2\Delta_A . \tag{4.3.90}$$

Thus, adding the operator $A(x)$ to the Lagrangian density of the free theory, $\mathcal{L}_0 = \frac{1}{2}(\partial_\mu\varphi)^2$, does not alter the infrared behavior unless $\Delta_A \leq 2$. For $\Delta_A \leq 2$, the infrared divergences grow more and more severe with the order of

perturbation theory in g_A, the coupling constant for the operator $A(x)$. Conversely, for $\Delta_A > 2$ the infrared behavior is, at every order of perturbation theory in g_A, the same as that of a theory with $g_A = 0$.

In addition to the fermions themselves, two operators $\mathcal{O}_a(x)$ and $\mathcal{Q}_b(x)$ are of importance to us

$$\begin{aligned} \mathcal{O}_a(x) &= e^{ia\varphi(x)}, \\ \mathcal{Q}_b(x) &= e^{ib\int_{-\infty}^{x_1} dx_1' \partial_0 \varphi(x_0, x_1')}. \end{aligned} \tag{4.3.91}$$

The equal-time correlation functions for $\mathcal{O}_a$ and $\mathcal{Q}_b$ are

$$\langle G| : \mathcal{O}_a(x) :: \mathcal{O}_a^\dagger(y) : |G\rangle = \text{const} \times e^{a^2[\varphi^+(x_0,x_1),\varphi^-(x_0,y_1)]}. \tag{4.3.92}$$

Similarly we get

$$\langle G| : \mathcal{Q}_b(x) :: \mathcal{Q}_b^\dagger(y) : |G\rangle = \text{const} \times e^{b^2 \int_{-\infty}^{x_1} dx_1 \int_{-\infty}^{y_1} dy_1 [\partial_0\varphi^+(x_0,x_1),\partial_0\varphi^-(x_0,y_1)]}. \tag{4.3.93}$$

After a short computation, we get for the equal-time correlation functions

$$\langle G| : \mathcal{O}_a(x) :: \mathcal{O}_a^\dagger(y) : |G\rangle \approx \frac{\text{const}}{|x_1 - y_1|^{\frac{a^2}{2\pi}}} \tag{4.3.94}$$

and

$$\langle G| : \mathcal{Q}_b(x) :: \mathcal{Q}_b^\dagger(y) : |G\rangle \approx \frac{\text{const}}{|x_1 - y_1|^{\frac{b^2}{2\pi}}}. \tag{4.3.95}$$

Thus, the dimension Δ of the operator $:\cos(\beta\varphi):$ is equal to

$$\Delta = \frac{\beta^2}{4\pi}. \tag{4.3.96}$$

For $\Delta \leq 2$ (i.e. $\beta^2 \leq 8\pi$) this interaction is relevant in the infrared and for $\beta^2 \geq 8\pi$ it is infrared trivial. Thus, for values of the anisotropy γ stronger than a critical value $\gamma_c \simeq \frac{\pi}{2}$, we expect the non-linear term to be dominant. In this regime, the field φ has small fluctuations around the classical value, determined by its equations of motion. The order parameter field $\bar{\psi}\psi$ has a non-zero expectation value and the ground state is two-fold degenerate. This is the Ising regime of the Heisenberg model.

For the lattice theory one expects, and this is confirmed by a Bethe-Ansatz calculation, that γ_c should be equal to one [Luther 75]. In other words, the quantum Heisenberg antiferromagnet should be at this critical point. For $\gamma < \gamma_c$, XY anisotropy should dominate and the Mermin-Wagner theorem would prohibit the spontaneous breaking of the continuous symmetry of the XY model. The domain $\gamma < \gamma_c$ is a line (or segment) of critical points. A detailed theory of this phase transition, in connection with the Kosterlitz-Thouless transition can be found in [Amit 80].

The correlation functions for all interesting operators on the domain $\gamma \leq \gamma_c$, can be calculated. All the expressions listed below get logarithmic corrections to scaling at $\gamma = \gamma_c$. The dimensions of the fermion $\Delta(\psi)$, longitudinal $\Delta(\bar{\psi}\psi)$ and transverse $\Delta(M^\pm)$ components of the staggered order

parameter are found to be

$$\Delta(\psi) = \frac{\left(\frac{4\pi}{\beta}\right)^2}{4\pi} + \frac{\left(\frac{\beta}{4}\right)^2}{4\pi} = \frac{4\pi}{\beta^2} + \frac{\beta^2}{64\pi} \tag{4.3.97}$$

$$\Delta(\bar{\psi}\psi) = \frac{\beta^2}{16\pi} \tag{4.3.98}$$

$$\Delta(M^{\pm}) = \frac{\left(\frac{4\pi}{\beta}\right)^2}{4\pi} = \frac{4\pi}{\beta^2} \tag{4.3.99}$$

where we have kept only the contributions with smallest dimension (more relevant). Thus, at γ_c, we find ($\beta_c^2 = 8\pi$)

$$\Delta(\psi) = \frac{5}{8} \tag{4.3.100}$$

$$\Delta(\bar{\psi}\psi) = \frac{1}{2} \tag{4.3.101}$$

$$\Delta(M^{\pm}) = \frac{1}{2}. \tag{4.3.102}$$

Conversely, at the XY point ($\gamma = 0$ or $\beta^2 = 16\pi$) we get

$$\Delta(\psi) = \frac{1}{2} \tag{4.3.103}$$

$$\Delta(\bar{\psi}\psi) = 1 \tag{4.3.104}$$

$$\Delta(M^{\pm}) = \frac{1}{4}. \tag{4.3.105}$$

From these results we conclude that the anisotropy disappears at $\gamma = \gamma_c$ since the longitudinal and transverse components of the order paramater, the staggered magnetization, have the *same* correlations functions at the critical point $\gamma = \gamma_c$, which behave like

$$\langle G|M^+(x)M^-(y)|G\rangle|_{\gamma_c} = \langle G|\bar{\psi}(x)\psi(x)\bar{\psi}(y)\psi(y)|G\rangle|_{\gamma_c} \sim \frac{\text{const}}{|x_1 - y_1|} \tag{4.3.106}$$

up to logarithmic corrections to scaling.

For $\gamma < \gamma_c$ the correlation functions are different although both exhibit an algebraic decay (i.e. power law behaviour) with exponents η_z and $\eta_\pm$ satisfying $\eta_z > \eta_\pm$. These exponents are universal in the sense that their numerical values are independent of the short-distance cutoff. However, the coupling constant itself does depend on the precise definition of the cutoff. Thus the value of γ_c, which is equivalent to one in the lattice system, turned out to be close to $\frac{\pi}{2}$ for the continuum model. Nevertheless, it is possible to find a relationship between the continuum and lattice coupling constants [Luther 75].

The fact that the correlation functions exhibit a power law behavior means that the system, for $\gamma < \gamma_c$, is critical. It has been argued [den Nijs 81] that this is a line of critical points ending at γ_c, the Heisenberg point. That the system

is critical means that there are no energy gaps, that is, *all* the excitations are gapless. For $\gamma > \gamma_c$ a gap is known to develop [den Nijs 81]. This is the regime with $\beta^2 < 8\pi$ in the Sine-Gordon theory. Renormalization group arguments imply that $e^{i\frac{\beta}{2}\phi}$ has long range order, i.e. $\bar{\psi}\psi$ has a vacuum expectation value different from zero.

It is natural to ask whether the fact that the spin one-half Heisenberg chain is at a critical point with gapless (neutral) fermions in the spectrum does generalize to other situations such as higher spin or higher dimensions. We will see below that the spin one-half chain is very special and that, for example, *integer* spin chains are not critical.

References

[Affleck 85] I. Affleck. Phys. Rev. Lett. 54, 966 (1985).

[Affleck 86] I. Affleck. Nucl. Phys. B265, 409 (1986).

[Amit 80] D.J. Amit J. Goldschmidt and G. Grinstein. J. Phys. A 13, 585 (1980).

[Bethe 31] H. Bethe. Z. Physik 71, 205 (1931).

[Bloch 34] F. Bloch. Helv. Phys. Acta 7, 385 (1934).

[Coleman 75] S. Coleman. Phys Rev. D 11, 2088 (1975).

[den Nijs 81] M.P.M. den Nijs. Phys. Rev. B 23, 6111 (1981).

[Emery 79] V.J. Emery. In *Highly Conducting One-Dimensional Solids.* Edited by J.T. Devreese, R.P. Evrard and V.E. van Doren, Plenum, 1979.

[Faddeev 81] L. Faddeev. In *Recent Advances in Field Theory and Statistical Mechanics* Les Houches Summer School 1982, session XXXIX, edited by L.B. Zuber and R. Stora, North Holland, 1984.

[Fradkin 78] E. Fradkin and L. Susskind. Phys. Rev D 17, 2637 (1928).

[Haldane 82] F.D.M. Haldane. Phys. Rev. B 25, 4925 (1982).

[Haldane 83] F.D.M. Haldane. Phys. Lett. 93A, 464 (1983) and Phys. Rev. Lett. 50, 1153 (1983).

[Jordan 28] P. Jordan and E. Wigner. Z. Phys. 47, 631 (1928).

[Kadanoff 69] L.P. Kadanoff. Phys. Rev. Lett. 23, 1430 (1969).

[Kadanoff 71] L.P. Kadanoff and H. Ceva. Phys. Rev. B 3, 3918 (1971).

[Kennedy 88] T. Kennedy, E. Lieb and S. Shastri. Phys. Rev. Lett. 61, 2582 (1988).

[Lieb 61] E. Lieb, T. Schultz and D. Mattis. Ann. Phys. 16, 407 (1961).

[Lieb 65] E. Lieb and D. Mattis. J. Math. Phys 6, 304 (1965).

[Lowenstein 81] J. Lowenstein. In *Recent Advances in Field Theory and Statistical Mechanics* Les Houches Summer School 1982, session XXXIX, edited by L.B. Zuber and R. Stora, North Holland, 1984.

[Luther 75] A. Luther and I. Peshel. Phys. Rev. B 12, 3908 (1975).

[Mandelstam 75] S. Mandelstam. Phys Rev. D 11, 3026 (1975).

[Marshall 55] W. Marshall. Proc. Roy. Soc. A 232, 48 (1955).

[Pines 66] D. Pines and P. Nozieres. *The Theory of Quantum Liquids*, Benjamin, 1966.

[Polyakov 84] A.M. Polyakov and P. Wiegmann. Phys. Lett. B141, 223 (1984).

[Tomonaga 50] S. Tomonaga. Progr. of Theor. Phys. 5, 544 (1950).

[Wilson 69] K.G. Wilson. Phys. Rev. 179, 1499 (1969).

[Witten 84] E. Witten. Comm. Math. Phys. 92, 455 (1984).

[Yang 69] C.N. Yang and C.P. Yang. Journ. Mat. Phys. 10, 1115 (1969).

CHAPTER 5

Sigma Models, Topological Terms and Spin Liquids

5.1 Generalized Spin Chains and Haldane's Conjecture

The phenomenology which emerges from the spin one-half Heisenberg antiferromagnetic chain is quite striking: no long range order, there are gapless states in particular, gapless spinless fermions (which, in the Heisenberg picture are solitons). From the point of view of the Hubbard model, the Heisenberg model occurs at infinite coupling where the charge-bearing degrees of freedom get a gap infinitely large. Thus spin and charge degrees of freedom are separated and the spin sector is at a critical point. This phenomenology inspired Anderson [Anderson 87] to propose a similar picture for the two-dimensional systems, the Resonating Valence Bond (RVB) picture. However, most of this picture surely should not generalize. Critical points are not generic and, in general, it is not possible to have gapless states without the spontaneous breaking of a continuous symmetry. Thus, the one-dimensional spin one-half case may be more the exception than the rule. For instance, it may be possible that the system is in a state without long-range order, which is likely to be massive. For this reason, it is important to consider generalizations of the Heisenberg model. This problem has been studied extensively. Two different approaches have been considered in one dimension (a) enlarging the representation (higher spin, same symmetry group SU(2)) and (b) higher symmetry groups (SU(N), for instance).

Haldane considered the generalization to higher spin but keeping the symmetry group to be SU(2) [Haldane 83] and [Haldane 85]. He first considered the *large spin limit* which should have semiclassical character. He showed that in this limit the effective Lagrangian was *almost* the Lagrangian of the quantum non-linear sigma model. That the non-linear sigma model should appear in a *semiclassical* ($S \to \infty$) limit should be of no surprise: one finds the same answer in mean-field theory. But there is something wrong with this picture.

The non-linear sigma model is known to have no long-range order and, in fact, it has a finite correlation length [Polyakov 75]. Thus if the sigma model truly was the infrared limit of the Heisenberg model, it could not possibly be a critical system, at least for S sufficiently large. Haldane found that this is indeed the case for spin systems in which S is an *integer*. For *half-integer* spins, he found that, in addition to the sigma model, there is an extra term which changes the physics drastically. The extra term turned out to be proportional to a topological invariant; the winding number or Pontryagin index of the (smooth) spin configuration. Thus it would appear that integer and half-integer *spin-chains* behave rather differently.

"Spin" systems with other symmetry groups have also been considered. These include SU(N) generalizations of the (SU(2)) Heisenberg model for various representations of the group. Affleck studied *a large-N limit* in which he was able to show that the ground state does not have long-range order and that there are no gapless states [Affleck 85]. However, other SU(N) generalizations of the Heisenberg model have been considered. For special choices of parameters, these systems are integrable (in the Bethe-Ansatz sense), and they are also at a critical point [Babudjian 1986]. Their critical behavior is however different from the one we discussed in the Heisenberg case. Thus, it appears that, at least in one dimension, these systems are either critical or in a *disordered state*, i.e. a state without long-range order and with only short-range spin correlations.

Let us first discuss the spin-s quantum Heisenberg chain. I will do so by introducing a path-integral method for spin systems which *does* generalize to higher dimensions, groups, representations, etc.

5.2 Path-Integrals for Spin Systems: The Single Spin Problem

In Chapter 3 we developed a path-integral method for Fermi systems of the Hubbard type (i.e. with local interactions). Using a Hubbard-Stratonovich transformation we were able to derive an effective action for the low-energy degrees of freedom, the spin fluctuations. The result was a path-integral representation of the long-range spin fluctuations, the quantum mechanical non-linear sigma model.

We also showed that, in the strong coupling limit, the half-filled Hubbard model maps onto the quantum Heisenberg model. In this limit the "band" fermions are bound into localized spins. There is no motion of fermions, the gap for charge fluctuations is infinitely large. It is natural to ask for an alternative derivation of the effective action for the spin fluctuations which should not be based on the weak coupling mean-field theory, as we did in Chapter 3. Also we will now be careful enough to keep terms of topological significance.

We begin with the discussion of an extremely simple system: a spin s degree of freedom coupled to an external field through a Zeeman term. From

the standard treatment in elementary quantum mechanics [Baym 74] we know that the $(2s+1)$-fold degeneracy is lifted by the Zeeman interaction, resulting in $(2s+1)$ non-degenerate levels. The path-integral will enable us to study the evolution operator between arbitrary initial and final states.

There are several published path-integral treatments of spin degrees of freedom. They all share the feature that they deal with coherent states rather than the more familiar complete states [Schulman 81]. The method of coherent states has been extensively reviewed by A. Perelonov [Perelomov 1986]. We will use a special version of the method of coherent states which keeps the spin symmetry intact, first introduced by Wiegmann [Wiegmann 88] and by Fradkin and Stone [Fradkin 88].

Let us begin by describing the Hilbert space. It is very simple. We have $(2s+1)$ states which transform like a spin-s representation of SU(2). Let $|0\rangle$ denote the *highest weight state* in this representation

$$|0\rangle = |S, S\rangle. \tag{5.2.1}$$

This state is an eigenstate of both S_3, the (only) diagonal generator of SU(2), and of the quadratic Casimir invariant $\vec{S}^2$

$$S_3|0\rangle = s|0\rangle, \tag{5.2.2}$$

$$\vec{S}^2|0\rangle = s(s+1)|0\rangle. \tag{5.2.3}$$

Consider now the state $|\vec{n}\rangle$ labeled by the unit vector $\vec{n}$ which is obtained by the rotation

$$|\vec{n}\rangle = e^{i\theta(\vec{n}_o \times \vec{n})\cdot\vec{S}}|S, S\rangle \tag{5.2.4}$$

where $\vec{n}_0$ is a unit vector along the quantization axis, θ is the co-latitude

$$\vec{n}\cdot\vec{n}_0 = \cos\theta \tag{5.2.5}$$

and S_i $(i = 1, 2, 3)$ are the (three) generators of SU(2) in the spin-s representation (see Fig. 5.1). For a review of SU(2) and its representations, see, for instance, Georgi [Georgi 82].

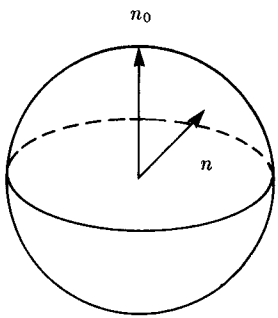


Figure 5.1 The unit sphere S_2 and the unit vectors $\vec{n}_0$ and $\vec{n}$.

The state $|\vec{n}\rangle$ can be expanded in a complete basis of the spin-s irreducible representation $\{|S,M\rangle\}$ where M labels the eigenvalue of S_3

$$S_3|S,M\rangle = m|S,M\rangle, \tag{5.2.6}$$

$$\vec{S}^2|S,M\rangle = s(s+1)|S,M\rangle \tag{5.2.7}$$

and $-s \leq m \leq s$. The coefficients of the expansion are the representation matrices $D^{(S)}(\vec{n})_{MS}$

$$|\vec{n}\rangle = \sum_{M=-S}^{S} D^{(S)}(\vec{n})_{MS}|S,M\rangle. \tag{5.2.8}$$

Clearly, there are many other rotations, differing from one another by multiplication on the right by rotations about the z-axis. This will give rise to the same state, except for an overall phase. In more formal terms, the observable states are in a one-to-one correspondence with the right co-sets SU(2)/U(1) where U(1), the phase, is generated by the *diagonal* generators of SU(2), that is S_3. In the language of differential geometry, the coherent states form a Hermitian line bundle associated with the Hopf, or monopole, principal bundle. The matrices $D^{(S)}$ do not form a group but rather satisfy

$$D^{(S)}(\vec{n}_1)D^{(S)}(\vec{n}_2) = D^{(S)}(\vec{n}_3)e^{i\Phi(\vec{n}_1,\vec{n}_2,\vec{n}_3)S_3} \tag{5.2.9}$$

where $\vec{n}_1, \vec{n}_2$ and $\vec{n}_3$ are three arbitrary unit vectors and $\Phi(\vec{n}_1,\vec{n}_2,\vec{n}_3)$ is the area of the spherical triangle with vertices at $\vec{n}_1$, $\vec{n}_2$ and $\vec{n}_3$ (see Fig. 5.2). Equation (5.2.9) is simply saying that the $D^{(S)}$-matrices form a group up to an element generated by the diagonal generators, the Cartan subalgebra. Since the sphere S_2 is a manifold without boundaries, the area of a spherical triangle is not uniquely defined. The shaded areas of the sphere in Fig. 5.2 (a)

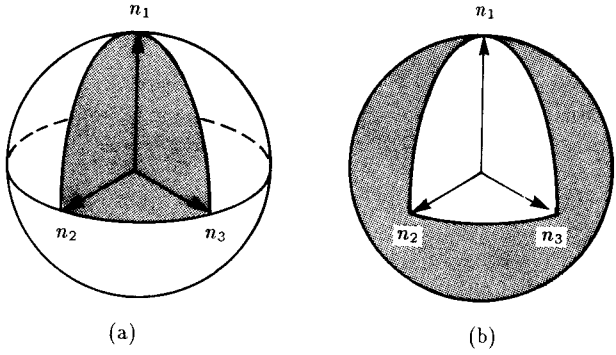


Figure 5.2 The spherical triangle with vertices at $\vec{n}_1$, $\vec{n}_2$ and $\vec{n}_3$. Its area is not unambiguously defined. "Inner" area (a) and "outer" area (b).

and (b) are equally good definitions of the area. The difference of the *oriented areas* is 4π. Since S_3 has eigenvalues equal to m, which is either an integer or a half-integer, this ambiguity has no physical manifestation since

$$e^{i4\pi m} = 1. \tag{5.2.10}$$

We can regard the requirement that the ambiguity in the definition of the area should lead to no physical consequences as the origin of the quantization of spin.

Other useful properties are the inner product $\langle \vec{n}_1 | \vec{n}_2 \rangle$

$$\begin{aligned} \langle \vec{n}_1 | \vec{n}_2 \rangle &= \langle 0 | D^{(S)\dagger}(\vec{n}_1) D^{(S)}(\vec{n}_2) | 0 \rangle \\ &= e^{i\Phi(\vec{n}_1, \vec{n}_2, \vec{n}_3)s} \left(\frac{1 + \vec{n}_1 \cdot \vec{n}_2}{2} \right)^s , \end{aligned} \tag{5.2.11}$$

the diagonal matrix elements of the generators $\vec{S}$

$$\langle \vec{n} | \vec{S} | \vec{n} \rangle = s\vec{n} \tag{5.2.12}$$

and the "resolution of unity" which is an expression of the identity operator $\hat{I}$ in terms of the coherent state operators $|\vec{n}\rangle\langle\vec{n}|$

$$\hat{I} = \int d\mu(\vec{n}) |\vec{n}\rangle\langle\vec{n}|. \tag{5.2.13}$$

The integration measure $d\mu(\vec{n})$ is given by the invariant measure

$$d\mu(\vec{n}) = \left(\frac{2s+1}{4\pi} \right) d^3n \delta(\vec{n}^2 - 1). \tag{5.2.14}$$

We are now in a position to write down an expression for the path-integral in this coherent state representation. Its generalization to other groups is straightforward and has been given by Wiegmann. Let $H(\vec{S}) = \vec{B} \cdot \vec{S}$ be the Zeeman-like Hamiltonian for a spin system with one spin-s degree of freedom. I will consider the representation of the evolution operator in imaginary time

$$Z = \mathrm{Tr}\ e^{iHT} = \mathrm{Tr}\ e^{-\beta H}. \tag{5.2.15}$$

In other words, we are assuming that the initial and final states are identified. Let us split the imaginary time interval into N_t steps each of length δt and consider the limit $N_t \to \infty$ and $\delta t \to 0$ while keeping $N_t \delta t = \beta$ constant. As usual we make use of the Trotter formula

$$Z = \mathrm{Tr}\ e^{-\beta H} = \lim_{\substack{N_t \to \infty \\ \delta t \to 0}} \left(e^{-\delta t H} \right)^{N_t} \tag{5.2.16}$$

and insert the "resolution of identity", Eq. (5.2.13), at every intermediate time

$$Z = \lim_{\substack{N_t \to \infty \\ \delta t \to 0}} \left(\prod_{j=1}^{N_t} \int d\mu(\vec{n}_j) \right) \left(\prod_{j=1}^{N_t} \langle \vec{n}(t_j) | e^{-\delta t H} | \vec{n}(t_{j+1}) \rangle \right) \tag{5.2.17}$$

with periodic boundary conditions. Here $\{t_j\}$ is a set of intermediate times in the imaginary–time interval $[0, \beta]$. Since δt is small we can approximate Eq.

(5.2.17)

$$Z = \lim_{\substack{N_t\to\infty \\ \delta t\to 0}} \left(\prod_{j=1}^{N_t} \int d\mu(\vec{n}_j) \right) \left(\prod_{j=1}^{N_t} [\langle \vec{n}(t_j)|\vec{n}(t_{j+1})\rangle - \delta t \langle \vec{n}(t_j)|H|\vec{n}(t_{j+1})\rangle] \right). \tag{5.2.18}$$

Within the same approximation we can write

$$\frac{\langle \vec{n}(t_j)|H|\vec{n}(t_{j+1})\rangle}{\langle \vec{n}(t_j)|\vec{n}(t_{j+1})\rangle} \approx \langle \vec{n}(t_j)|H|\vec{n}(t_j)\rangle + \mathcal{O}(\delta t). \tag{5.2.19}$$

Using the inner-product formula, Eq.(5.2.11), we get

$$\langle \vec{n}(t_j)|\vec{n}(t_{j+1})\rangle = e^{i\Phi(\vec{n}(t_j),\vec{n}(t_{j+1}),\vec{n}_0)s} \left(\frac{1+\vec{n}(t_j)\cdot\vec{n}(t_{j+1})}{2} \right)^s. \tag{5.2.20}$$

We now insert Eqs. (5.2.19) and (5.2.20) into Eq. (5.2.18) to find the (formal) path-integral

$$Z = \lim_{\substack{N_t\to\infty \\ \delta t\to 0}} \int \mathcal{D}\vec{n} e^{-\mathcal{S}_E[\vec{n}]} \tag{5.2.21}$$

where the measure $\mathcal{D}\vec{n}$ is given by

$$\mathcal{D}\vec{n} = \prod_{j=1}^{N_t} d\mu(\vec{n}(t_j)) \tag{5.2.22}$$

and the (exponentiated) Euclidean action is equal to

$$-\mathcal{S}_E[\vec{n}] = is \sum_{j=1}^{N_t} \Phi(\vec{n}(t_j), n(t_{j+1}), \vec{n}_0) + s \sum_{j=1}^{N_t} \ln \left(\frac{1+\vec{n}(t_j)\cdot\vec{n}(t_{j+1})}{2} \right) - \sum_{j=1}^{N_t} \langle \vec{n}(t_j)|H|\vec{n}(t_j)\rangle. \tag{5.2.23}$$

In this derivation, we have assumed that the sequence of unit vectors $\{\vec{n}(t_j)\}$ are closed trajectories (because $\vec{n}(t_0) = \vec{n}(t_{N+1})$) on the sphere S_2 which are sufficiently smooth so that all the approximations of Eq. (5.2.19) make sense (see Fig. 5.3). This is not quite the case, as emphasized by Klauder [Klauder 79]. But these technicalities, as well as operator ordering problems, can be taken care of without affecting the physics. We will ignore these difficulties from now on. Our path integral will be as good a mathematical object as any other path integral.

The first term of the effective Euclidean action is complex. It leads to a sum over trajectories weighted by phases (even though we are working in imaginary time!) of the form

$$e^{is\mathcal{A}[\vec{n}]} \tag{5.2.24}$$

where $\mathcal{A}[\vec{n}]$ is the limit

$$\mathcal{A}[\vec{n}] = \lim_{\substack{N_t \to \infty \\ \delta t \to 0}} \sum_{j=1}^{N_t} \Phi(\vec{n}(t_j), \vec{n}(t_{j+1}), \vec{n}_0). \tag{5.2.25}$$

Since each term of this sum is the area of the spherical triangle with vertices at $\vec{n}(t_j)$, $\vec{n}(t_{j+1})$ and $\vec{n}_0$, the sum, i.e. the sum of these areas, is just equal to the total area of the cap Σ bounded by the trajectory Γ parametrized by $\vec{n}(t)$ (see Fig. 5.4). Once again, since S_2 has no boundaries, there are two caps Σ^+ and Σ^-. The oriented areas of Σ^+ and Σ^- also differ by 4π

$$\mathcal{A}(\Sigma^+) + \mathcal{A}(\Sigma^-) = 4\pi. \tag{5.2.26}$$

This is the same ambiguity we encountered before. It does not lead us to any observable effects since s is restricted to be an integer or a half-integer. The area of the cap Σ, say, Σ^+, is given by (in the limit $N_t \to \infty$, $\delta t \to 0$)

$$\mathcal{A}(\Sigma^+) = \int_0^1 d\tau \int_0^\beta dt \vec{n}(t,\tau) \cdot (\partial_t \vec{n}(t,\tau) \times \partial_\tau \vec{n}(t,\tau)) \equiv \mathcal{S}_{\rm wz}[\vec{n}] \tag{5.2.27}$$

where $\vec{n}(t,\tau)$ is an arbitrary, smooth parametrization of the cap Σ^+ bounded by Γ which satisfies the boundary conditions

$$\vec{n}(t,0) \equiv \vec{n}(t), \quad \vec{n}(t,1) \equiv \vec{n}_0, \quad \vec{n}(0,\tau) = \vec{n}(\beta,\tau), \tag{5.2.28}$$

where $t \in [0,\beta]$ and $\tau \in [0,1]$. Terms of this sort are generically called Wess-Zumino terms although sometimes they are also referred to as Chern-Simons terms.

We now proceed to take a naive continuum limit ($N_t \to \infty, \delta t \to 0$) and find from Eqs. (5.2.27) and (5.2.23) the Euclidean action

$$\mathcal{S}_{\rm E}[\vec{n}] = -is\mathcal{S}_{\rm wz}[\vec{n}] + \frac{s\delta t}{4} \int_0^\beta dt (\partial_t \vec{n}(t))^2 + s \int_0^\beta dt \vec{B} \cdot \vec{n}(t) \tag{5.2.29}$$

where $\vec{B}$ is an external magnetic field.

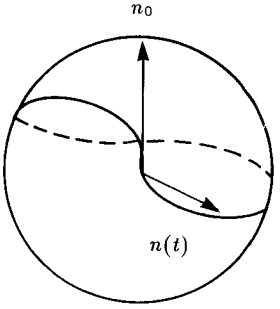


Figure 5.3 Closed smooth trajectories on S_2.

We can get back to real time x_0, with

$$t = ix_0 \quad \text{and} \quad \beta = iT \tag{5.2.30}$$

by writing

$$Z = \int \mathcal{D}\vec{n} e^{i\mathcal{S}_{\mathrm{M}}[\vec{n}]} \tag{5.2.31}$$

where $\mathcal{S}_{\mathrm{M}}[\vec{n}]$ is given by

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s\mathcal{S}_{\mathrm{wz}}[\vec{n}] + \left(\frac{s\delta t}{4}\right) \int_0^T dx_0 (\partial_0 \vec{n}(x_0))^2 - s \int_0^T dx_0 \vec{B} \cdot \vec{n}(x_0). \tag{5.2.32}$$

This expression has a simple mechanical analogy. Let us imagine that $\vec{n}(x_0)$ is the position vector of a charged particle at time x_0. The particle has a small mass $m = \frac{s\delta t}{2}, (m \to 0)$ and is constrained to move on the surface of the unit sphere, S_2. A magnetic monopole with magnetic charge s is placed at the center of the sphere. The usual electromagnetic coupling gives a contribution to the action of the form [Landau 75].

$$\mathcal{S}_{\mathrm{em}} = \oint dx_0 \vec{A} \cdot \frac{\partial \vec{n}}{\partial x_0} \tag{5.2.33}$$

where $\vec{A}$ is the vector potential at $\vec{n}$. In order to represent a monopole, the vector potential has to have a singular piece which describes the Dirac string. We can use Stoke's theorem to write $\mathcal{S}_{\mathrm{em}}$ in terms of a two-form instead of the one-form $\vec{A}$. Stoke's theorem simply says that $\mathcal{S}_{\mathrm{em}}$ is given by the flux of the magnetic monopole through the area of S_2 bounded by the trajectory Γ (see Fig. 5.5). This is nothing but the magnetic charge s of the monopole multiplied by the area of S_2 bounded by Γ, in other words, the cap Σ of Fig. 5.4. This is precisely identical to the first term in the action Eq. (5.2.32). Ideas of this sort were first popularized by Witten [Witten 83] in his discussion of Wess-Zumino terms (see also Stone [Stone 86]).

The magnetic monopole gives rise to a uniform radial magnetic field on the surface of the sphere with total flux equal to the magnetic charge s. It is well known that the eigenstates of such a particle are monopole spherical

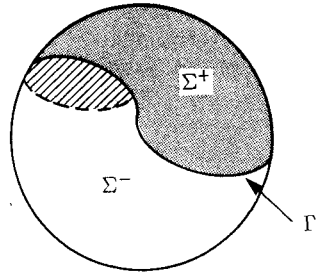


Figure 5.4 The trajectory Γ and the caps Σ^+ and Σ^-.

harmonics. The ground state is $(2s+1)$-fold degenerate and it is separated from the higher angular momentum states (i.e. Landau "orbits") by an energy gap which scales with the mass of the particle like $\frac{1}{m}$. Thus in the small mass limit ($m \to 0$) the system is projected onto the ground state. In this way the subspace of the "lowest-Landau orbit" on a spherical geometry becomes identical to the space of the spin-s representation of SU(2). In retrospect, it would have been possible to describe spin in terms of the path-integral with Eq. (5.2.32) for its action directly, without reference to coherent states.

5.3 Path-Integral for Many-Spin Systems

It is trivial to generalize the one-spin problem to a many (or infinitely many!) spin situation. Once again, I will follow the treatment of Fradkin and Stone.

The Hilbert space of a many-spin system is just the tensor product of the Hilbert space of the individual spins. Let H be the Hamiltonian for a spin-s system on an *arbitrary lattice*

$$H = J \sum_{(\vec{r},\vec{r}')} \vec{S}(\vec{r}\,) \cdot \vec{S}(\vec{r}\,') \tag{5.3.1}$$

where $(\vec{r},\vec{r}\,')$ are pairs of sites on that lattice. We can now use the identity $\langle\vec{n}|\vec{S}|\vec{n}\rangle = s\vec{n}$, to write down the imaginary-time action for the many-spin

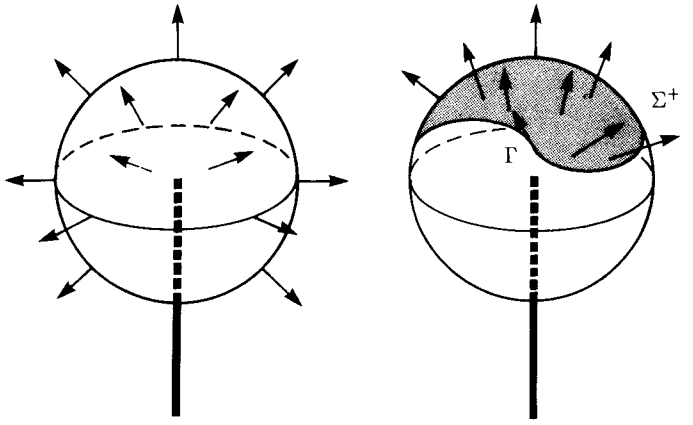


Figure 5.5 A magnetic monopole at the center of the unit sphere in (a) and the flux through the cap Σ^+ bounded by the trajectory Γ in (b). The thick line represents an infinitely long solenoid of infinitesimal thickness (Dirac string).

system

$$\mathcal{S}_{\mathrm{E}}[\vec{n}] = -is\sum_{\vec{r}}\mathcal{S}_{\mathrm{wz}}[\vec{n}(\vec{r}\,)] + \frac{m}{2}\int_0^\beta dt\sum_{\vec{r}}(\partial_t\vec{n}(\vec{r},t))^2 + \int_0^\beta dt\sum_{(\vec{r},\vec{r}')}Js^2\vec{n}(\vec{r},t)\cdot\vec{n}(\vec{r}\,',t) \tag{5.3.2}$$

where we are supposed to take the limit $m \to 0$ (it will be dropped from now on). The sums in Eq. (5.3.2) run over all the sites of the lattice. The first term is just the sum of the Wess-Zumino terms of the individual spins. Note that the only real time dependence enters through the Wess-Zumino terms.

We can Wick-rotate back to real time, $t = ix_0$, $\beta = iT$, and write the real-time action, $\mathcal{S}_{\mathrm{M}}$

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s\sum_{\vec{r}}\mathcal{S}_{\mathrm{wz}}[\vec{n}(\vec{r}\,)] - \int_0^T dx_0\sum_{(\vec{r},\vec{r}')}Js^2\vec{n}(\vec{r},x_0)\cdot\vec{n}(\vec{r}\,',x_0). \tag{5.3.3}$$

The effective action $\mathcal{S}_{\mathrm{M}}[\vec{n}]$ scale like s, the spin representation. Thus, in the large spin limit $s \to \infty$, the path-integral

$$Z = \int \mathcal{D}\vec{n}e^{i\mathcal{S}_{\mathrm{M}}[\vec{n}]} \tag{5.3.4}$$

should be dominated by the stationary points of the action $\mathcal{S}_{\mathrm{M}}[\vec{n}]$. This is the semiclassical limit. Corrections to the large-s limit can be arranged in an expression in powers of $\frac{1}{s}$. This is the content of the Holstein-Primakoff expansion [Holstein 40]. Note however that we did not make use of the semiclassical limit in order to derive the path-integral. Let us consider a number of cases of interest.

5.4 Quantum Ferromagnets

In this case we set $J = -|J|$. I will consider the case of an hypercubic lattice and restrict the sum over pairs of sites to nearest neighbors. The results can be generalized very easily to any other lattice.

I first make use of the constraint $\vec{n}^2 = 1$ to write the action in the form

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s\sum_{\vec{r}}\mathcal{S}_{\mathrm{wz}}[\vec{n}(\vec{r}\,)] - \frac{|J|s^2}{2}\sum_{<\vec{r},\vec{r}'>}\int_0^T dx_0[\vec{n}(\vec{r},x_0) - \vec{n}(\vec{r}\,',x_0)]^2 \tag{5.4.1}$$

up to an additive constant. Consider now the long-wavelength limit, in which $\vec{n}(\vec{r},x_0)$ is a smooth function of the spatial coordinates. If we denote by a_0 a short-distance cutoff (i.e. the lattice spacing) we can write the continuum action

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = \int d^dx\frac{s}{a_0^d}\mathcal{S}_{\mathrm{wz}}[\vec{n}] - \frac{|J|s^2}{2a_0^{d-2}}\int d^dx\int_0^T dx_0\,(\nabla_i\cdot\vec{n}(\vec{x},x_0))^2\,. \tag{5.4.2}$$

It is important to stress that the effective continuum action for the *quantum ferromagnet* does not have the non-linear sigma model form. This is natural since the non-linear sigma model leads to Goldstone bosons with a *linear* dispersion law. This is the correct result for antiferromagnetic magnons but not for ferromagnetic magnons. It is well known that ferromagnetic magnons have a quadratic dispersion relation [Bloch 30].

To see how does all this come about we will derive the classical equations of motion. We take care of the *local constraint*

$$\vec{n}^2(\vec{x}, x_0) = 1 \tag{5.4.3}$$

by introducing the Lagrangian multiplier field $\lambda(\vec{x}, x_0)$ which enforces the constraint in the path-integral through an extra term in the action

$$\mathcal{S}_{\text{extra}}[\vec{n}, \lambda] = \int d^d x \int_0^T dx_0 \frac{\lambda(\vec{x}, x_0)}{2} \left(\vec{n}^2(\vec{x}, x_0) - 1\right). \tag{5.4.4}$$

The classical equations of motion result from demanding that the total action

$$\mathcal{S}_{\text{tot}}[\vec{n}, \lambda] = \mathcal{S}_{\text{M}}[\vec{n}] + \mathcal{S}_{\text{extra}}[\vec{n}, \lambda] \tag{5.4.5}$$

be stationary

$$\delta \mathcal{S}_{\text{tot}} = 0. \tag{5.4.6}$$

The variation of the Wess-Zumino action is very simple. Indeed $\mathcal{S}_{\text{wz}}$ is essentially the area of the sphere bounded by the trajectory $\vec{n}(\vec{x}, x_0)$ (at each point $\vec{x}$) on the spin manifold (the two-sphere S_2). Thus the variation $\delta \mathcal{S}_{\text{wz}}$ due to a small change in the trajectory $\delta \vec{n}$ is simply equal to

$$\delta \mathcal{S}_{\text{wz}} = \delta \vec{n} \cdot (\vec{n} \times \partial_0 \vec{n}). \tag{5.4.7}$$

Hence, we get the classical equations of motion

$$\frac{\delta \mathcal{S}_{\text{tot}}}{\delta \vec{n}} = \nabla_i \left(\frac{\delta \mathcal{S}_{\text{tot}}}{\delta \nabla_i \vec{n}} \right) \tag{5.4.8}$$

supplemented by the constraint Eq. (5.4.3). More explicitly, we get

$$\frac{s}{a_0^d} \vec{n} \times \partial_0 \vec{n} + \lambda \vec{n} = -\frac{|J| s^2}{a_0^{d-2}} \nabla^2 \vec{n}. \tag{5.4.9}$$

The classical value of the Lagrange field λ can be evaluated by computing the scalar product of Eq. (5.4.9) with $\vec{n}$. The result is

$$\lambda = -\frac{|J| s^2}{a_0^{d-2}} (\vec{n} \cdot \nabla^2 \vec{n}). \tag{5.4.10}$$

Substituting Eq. (5.4.10) back in Eq. (5.4.9) I get the equation of motion for the quantum ferromagnet

$$\frac{s}{a_0^d} \vec{n} \times \partial_0 \vec{n} + \frac{|J| s^2}{a_0^{d-2}} \left(\nabla^2 - (\vec{n} \cdot \nabla^2 \vec{n}) \right) \vec{n} = 0. \tag{5.4.11}$$

By using elementary algebra as well as Eq. (5.4.3), this equation can be brought to the form

$$\partial_0 \vec{n} = |J| s a_0^2 \vec{n} \times \nabla^2 \vec{n}. \tag{5.4.12}$$

This equation is known as the Landau-Lifshitz equation. The derivation shown here is due to M. Stone.

The Landau-Lifshitz equation has several interesting properties. It is a non-linear equation with first-order time derivatives and second-order space derivatives. Thus the solutions of Eq. (5.4.12) have a quadratic dispersion law, as they should. The spins move in a precessional fashion with an angular velocity $\vec{\Omega}$ given by

$$\vec{\Omega} = -|J|sa_0^2\nabla^2\vec{n}. \tag{5.4.13}$$

The Landau-Lifshitz equations can be solved in the linear regime. Let us parametrize $\vec{n}$ by the components

$$\vec{n} = \begin{pmatrix} \vec{\pi} \\ \sigma \end{pmatrix} \tag{5.4.14}$$

where σ and π_i, $(i = 1, 2)$ satisfy the constraint

$$\sigma^2 + \vec{\pi}^2 = 1. \tag{5.4.15}$$

The (linearized) Landau-Lifshitz equations are

$$\begin{aligned} \partial_0\pi_1 &\approx -|J|sa_0^2\nabla^2\pi_2, \\ \partial_0\pi_2 &\approx +|J|sa_0^2\nabla^2\pi_1, \end{aligned} \tag{5.4.16}$$

to leading order in $\vec{\pi}$. From Eq. (5.4.16) we find the dispersion relation for ferromagnetic spin-waves

$$|p_0| \approx |J|sa_0^2|\vec{p}|^2 \tag{5.4.17}$$

which is known as Bloch's law.

5.5 Effective Action for One-Dimensional Quantum Antiferromagnets

We will not consider here frustrated systems. Thus and for the sake of simplicity, we will consider the case of quantum antiferromagnets on *bipartite* lattices, such as the hypercubic lattice. We will see that, unlike the case of the ferromagnets, the effective low-energy action is *different* for one-dimensional systems and for higher dimensional cases such as the square and cubic lattices. In all cases we will find a non-linear sigma model, in agreement with our previous discussion (see Chapter 3) based on a mean-field weak coupling treatment of the Hubbard model. But we will get more. For the spin-chain case we will find that the action has an extra term, a topological term.

The starting point will be, once again, the real-time action of Eq. (5.3.3) with a nearest neighbor *antiferromagnetic* coupling constant $J > 0$. Since we expect that *at least* the short-range order should have Néel character, it is natural to consider the *staggered* and *uniform* components of the spin-field $\vec{n}$. This construction, as is, works only for two-sublattice systems close to a Néel state, although it is possible to generalize it to other cases.

Consider a spin chain with an *even* number of sites N occupied by spin–s degrees of freedom. The sites of the lattice are labeled by an integer $j = 1, \ldots, N$. The real time action is

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s \sum_{j=1}^{N} \mathcal{S}_{\mathrm{wz}}[\vec{n}(j)] - \int_0^T dx_0 \sum_{j=1}^{N} Js^2 \vec{n}(j, x_0) \cdot \vec{n}(j+1, x_0) \qquad (5.5.1)$$

where we have assumed periodic boundary conditions. Since we expect to be close to a Néel state, we will stagger the configuration

$$\vec{n}(j) \to (-1)^j \vec{n}(j). \qquad (5.5.2)$$

On a bipartite lattice, the substitution of Eq. (5.5.2) will change the sign of the *exchange term* of the action to a *ferromagnetic* one. The Wess-Zumino terms are *odd* under the replacement of Eq. (5.5.2) and thus get *staggered*. Thus, it is the Wess-Zumino term, a purely quantum mechanical effect, which will distinguish ferromagnets from antiferromagnets. After staggering the spins we get, up to an additive constant,

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s \sum_{j=1}^{N} (-1)^j \mathcal{S}_{\mathrm{wz}}[\vec{n}(j)] - \frac{Js^2}{2} \int_0^T dx_0 \sum_{j=1}^{N} \left(\vec{n}(j, x_0) - \vec{n}(j+1, x_0)\right)^2 . \qquad (5.5.3)$$

We now split the (staggered) spin field $\vec{n}$ into a slowly varying piece $\vec{m}(j)$, the order parameter field, and a small rapidly varying part, $\vec{l}(j)$, which roughly represents the average spin [Affleck 88]. Hence, we write

$$\vec{n}(j) = \vec{m}(j) + (-1)^j a_0 \vec{l}(j). \qquad (5.5.4)$$

The constraint $\vec{n}^2 = 1$ and the requirement that the order-parameter field $\vec{m}$ should obey the same constraint, $\vec{m}^2 = 1$, demand that $\vec{m}$ and $\vec{l}$ be orthogonal vectors

$$\vec{m} \cdot \vec{l} = 0. \qquad (5.5.5)$$

The Wess-Zumino term is rewritten

$$s \sum_{j=1}^{N} (-1)^j \mathcal{S}_{\mathrm{wz}}[\vec{n}(j)] = s \sum_{r=1}^{N/2} \left(\mathcal{S}_{\mathrm{wz}}[\vec{n}(2r)] - \mathcal{S}_{\mathrm{wz}}[\vec{n}(2r-1)]\right) \qquad (5.5.6)$$

which, by making use of the approximation

$$\begin{aligned} \vec{n}(2r) - \vec{n}(2r-1) =& \vec{m}(2r) - \vec{m}(2r-1) + a_0(\vec{l}(2r) + \vec{l}(2r-1)) \\ =& a_0 \left(\partial_1 \vec{m}(2r) + 2\vec{l}(2r)\right) + \mathcal{O}(a_0^2), \end{aligned} \qquad (5.5.7)$$

becomes

$$s\sum_{j=1}^{N}(-1)^j \mathcal{S}_{\rm wz}[\vec{n}(j)] \approx s\sum_{r=1}^{N/2}\int_0^T dx_0\,\delta\vec{n}(2r,x_0)\cdot(\vec{n}(2r,x_0)\times\partial_0\vec{n}(2r,x_0))$$
$$\approx s\sum_{r=1}^{N/2}\int_0^T dx_0\left(a_0\partial_1\vec{m}(2r,x_0)+2a_0\vec{l}(2r,x_0)\right)$$
$$\times(\vec{m}(2r,x_0)\times\partial_0\vec{m}(2r,x_0))\,. \tag{5.5.8}$$

Thus, in the continuum limit, one finds

$$\lim_{a_0\to 0} s\sum_{j=1}^{N}(-1)^j \mathcal{S}_{\rm wz}[\vec{n}(j)] \approx \frac{s}{2}\int d^2x\;\vec{m}\cdot(\partial_0\vec{m}\times\partial_1\vec{m})+s\int d^2x\;\vec{l}\cdot(\vec{m}\times\partial_0\vec{m})\,. \tag{5.5.9}$$

Similarly, the continuum limit of the potential energy terms can also be found to be given by

$$\lim_{a_0\to 0}\frac{Js^2}{2}\sum_{j=1}^{N}\int_0^T dx_0\,(\vec{n}(j,x_0)-\vec{n}(j+1,x_0))^2 \approx$$
$$\frac{a_0Js^2}{2}\int d^2x\left((\partial_1\vec{m})^2+4\vec{l}^2\right). \tag{5.5.10}$$

Collecting terms we find a Lagrangian density involving both the order-parameter field $\vec{m}$ and the local spin-density $\vec{l}$

$$\mathcal{L}_{\rm M}(\vec{m},\vec{l}\,) = -2a_0Js^2\vec{l}^2+s\vec{l}\cdot(\vec{m}\times\partial_0\vec{m})-\frac{a_0Js^2}{2}(\partial_1\vec{m})^2+\frac{s}{2}\vec{m}\cdot(\partial_0\vec{m}\times\partial_1\vec{m})\,. \tag{5.5.11}$$

The fluctuations in the average spin density $\vec{l}$ can be integrated out. The result is the Lagrangian density of the non-linear sigma model

$$\mathcal{L}_{\rm M}(\vec{m}) = \frac{1}{2g}\left(\frac{1}{v_s}(\partial_0\vec{m})^2-v_s(\partial_1\vec{m})^2\right)+\frac{\theta}{8\pi}\epsilon_{\mu\nu}\vec{m}\cdot(\partial_\mu\vec{m}\times\partial_\nu\vec{m}) \tag{5.5.12}$$

where g and v_s are the coupling constant and spin-wave velocity respectively:

$$g=\frac{2}{s}, \tag{5.5.13}$$

$$v_s = 2a_0Js. \tag{5.5.14}$$

The last term has topological significance. We have chosen the normalization so that θ is given by

$$\theta = 2\pi s. \tag{5.5.15}$$

The tensor $\epsilon_{\mu\nu}$ is the usual Levi-Civita antisymmetric tensor in two dimensions.

Thus, apart from an anisotropy determined by the spin-wave velocity v_s and for the topological term, we find that the effective action for the low-frequency, long-wavelength fluctuation about a state with *short-range* Néel

order is given by the non-linear sigma model. We have reached the same results within the weak-coupling mean-field theory of the half-filled Hubbard model of Chapter 3. Indeed, using that approach, it is possible to get even the topological term [Wen 88].

5.6 The Role of Topology

In the past section we reached the conclusion that the low-energy excitations of a one-dimensional quantum antiferromagnet with short-range Néel order can be described by the path-integral of a non-linear sigma model with a topological term

$$Z = \int \mathcal{D}\vec{m}\delta(\vec{m}^2 - 1)e^{iS_{\text{eff}}[\vec{m}]} \tag{5.6.1}$$

with the effective action obtained from Eq. (5.5.12). Before considering the role of local quantum fluctuations, which are of fundamental importance here, we look at the role of the last term in the action, the topological term $\mathcal{S}_t^M$

$$\mathcal{S}_t^M = \frac{\theta}{8\pi}\int d^2x\epsilon_{\mu\nu}\vec{m}\cdot(\partial_\mu\vec{m}\times\partial_\nu\vec{m})\,. \tag{5.6.2}$$

Let us consider first the Euclidean sector of the theory (i.e. we are back to imaginary time $x_2 = ix_0$) with the Lagrangian density $\mathcal{L}_{\text{E}}$

$$\mathcal{L}_{\text{E}} = \frac{1}{2g}\left(v_s(\partial_1\vec{m})^2 + \frac{1}{v_s}(\partial_2\vec{m})^2\right) + i\frac{\theta}{8\pi}\epsilon_{ij}\vec{m}\cdot(\partial_i\vec{m}\times\partial_j\vec{m})\,. \tag{5.6.3}$$

We now define the Pontryagin index (or Winding Number) $\mathcal{Q}$ of the Euclidean space spin configuration $\{\vec{m}(x)\}$ by the expression

$$\mathcal{Q} = \frac{1}{8\pi}\int d^2x\epsilon_{ij}\vec{m}\cdot(\partial_i\vec{m}\times\partial_j\vec{m})\,. \tag{5.6.4}$$

We impose the boundary condition that the Euclidean action $\int d^2x\mathcal{L}_{\text{E}}[\vec{m}]$ is finite. This is equivalent to the requirement that asymptotically $\vec{m}$ becomes a *constant* vector $\vec{m}_0$ at space-time infinity

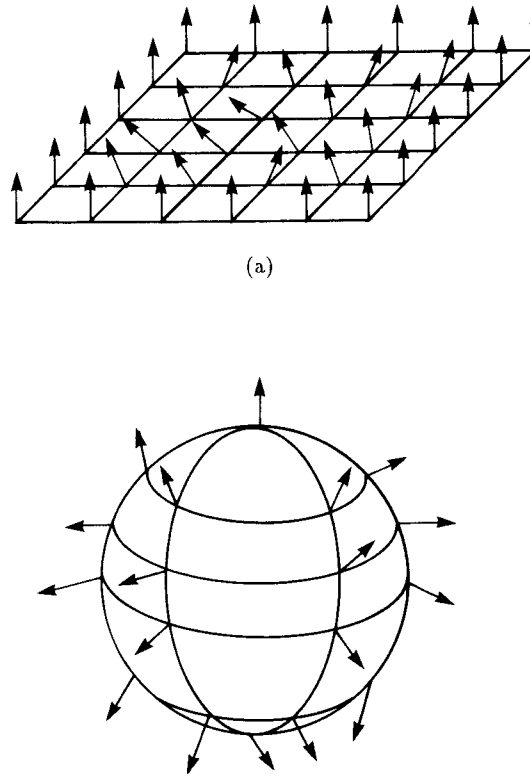
$$\lim_{|\vec{x}|\to\infty}\vec{m}(\vec{x}) = \vec{m}_0. \tag{5.6.5}$$

Thus, topologically, Euclidean space-time is a sphere S_2 since the fields are identified with $\vec{m}_0$ at the point of infinity (Fig. 5.6). The order parameter manifold is also isomorphic to a sphere S_2, since the constraint $\vec{m}^2 = 1$ has to be satisfied everywhere. A field configuration $\vec{m}(x)$ with finite action is thus a smooth (differentiable) mapping from the S_2 of Euclidean space-time to the S_2 of the order parameter manifold (Fig. 5.7).

The Pontryagin Index $\mathcal{Q}[\vec{m}]$ is the winding number in the sense that it counts how many times the spin configuration $\vec{m}$ has wrapped around the sphere, as can be checked by comparing the definition of $\mathcal{Q}$, Eq. (5.6.4), with

the area formula of Eq. (5.2.27). We can make these ideas more concrete by considering a configuration $\vec{m}(x)$ representing an *instanton* (Fig. 5.8). Let the field at infinity point parallel to $\vec{m}_0$, the north pole of S_2. In the case of an instanton, the field near the origin points *opposite* to $\vec{m}_0$, i.e. in the direction of the south pole. Alternatively, we can look at the configuration on S_2. Here it looks like a magnetic monopole or a hairy ball (Fig. 5.9). The winding number Q of this configuration is determined by the area of the sphere divided by 4π (i.e. the "magnetic flux")

$$Q = \left(\frac{1}{4\pi}\right) 4\pi = +1. \tag{5.6.6}$$



(a)

(b)

Figure 5.6 A finite action spin configuration in Euclidean space-time (a) is isomorphic to one on the sphere S_2 (b).

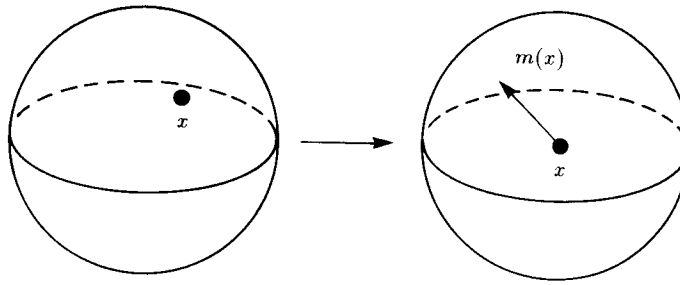


Figure 5.7 The mapping $\vec{m}(\vec{x})$.

Thus, an instanton has winding number +1. An anti-instanton has $Q = -1$. It is also possible to find multi-instanton configurations with arbitrary *integral* winding number Q.

We conclude that the smooth configurations $\vec{m}(x)$ can be classified according to their winding number. In other words, the configurations $\vec{m}(x)$ are mappings of S_2 into S_2 with *homotopy* classes classified by an integer, the Pontryagin Index Q. In mathematical terms the statement is

$$\Pi_2(S_2) = \mathcal{Z}. \tag{5.6.7}$$

Back in real time we can consider *soliton* configurations, such as the half-twist of Fig. 5.10. As time goes by, each spin traces a closed path on the sphere S_2 and hence it sweeps an area bound by the path. If we define that the area swept by a spin at $-\infty$ is equal to zero, we see that as we move from left to right the spins sweep an increasingly large area. At $+\infty$ the area swept is that of a full sphere, 4π. It is easy to see that Q is also equal to one for

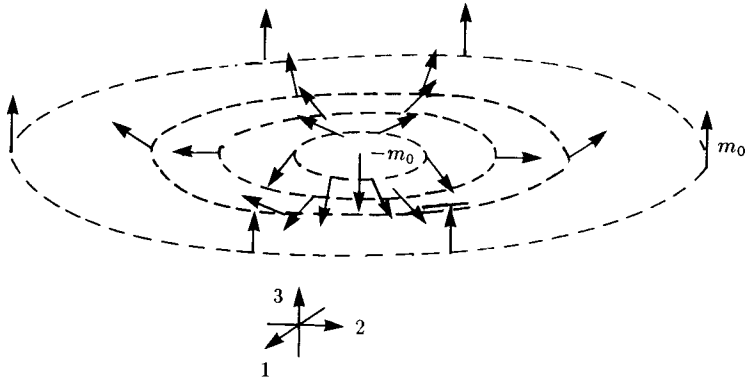


Figure 5.8 An instanton configuration in Euclidean space-time.

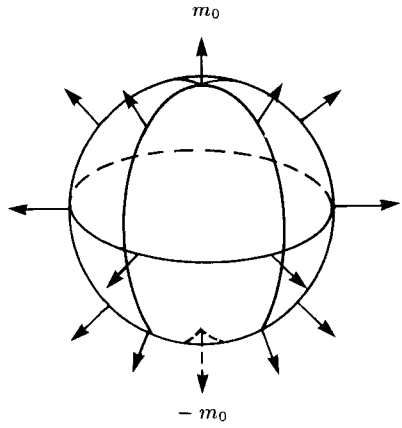


Figure 5.9 An instanton on S_2 has the same topology as a monopole.

the half-twist. At each point in space, the spins are coherently precessing and keeping their relative angles constant. In other words, the spins trace lines of longitude on a sphere. The global configuration still looks like a monopole and hence also has winding number +1 (Fig. 5.11).

The final conclusion is that the topological term, Eq. (5.6.2), is proportional to an integer $\mathcal{Q}$. The action in the path-integral of Eq. (5.6.1) has a contribution equal to $(2\pi s)\mathcal{Q}$ which should be added to the sigma-model term. Since s is an integer or a half-integer, we find that the extra, topological, term gives conbribution

$$e^{i2\pi s\mathcal{Q}} = (-1)^{2s\mathcal{Q}}. \tag{5.6.8}$$

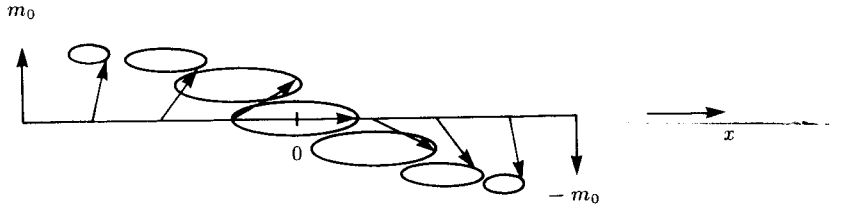


Figure 5.10 Half-twist soliton. The circles represent the precession of the spins.

Thus, if s is an integer, the spin-chain is described at low energies by the non-linear sigma model. For a half-integral s, each topological *class* contributes with a sign which is positive (negative) if the winding number Q is even (odd). Note that the sign does not depend on the actual value of the spin s, but only on whether it is an integer or half-integer. This means that the physics of this problem is *not* analytic in s: the integer and half-integer spin-chains fall in different universality classes. We will now see that this property implies a very important result, known as Haldane's conjecture, which states that the integer spin chains are massive (i.e. have a gap) while the half-integer chains are massless as in the spin one-half case.

5.7 Quantum Fluctuations: Renormalization Group

In the previous section we saw that the configuration space of the non-linear sigma model can be partitioned into classes classified by their winding numbers

$$Q = \frac{1}{8\pi}\int d^2x \epsilon_{abc}\epsilon_{ij} m_a \partial_i m_b \partial_j m_c \tag{5.7.1}$$

which are topological invariants. Thus, the partition function is a sum over distinct topological sectors

$$Z = \int \mathcal{D}\vec{m} e^{-\mathcal{S}_E[\vec{m}]} = \sum_{Q=0}^{\infty} \int_Q \mathcal{D}\vec{m} e^{-\mathcal{S}_0^E[\vec{m}]} e^{i2\pi s Q} \tag{5.7.2}$$

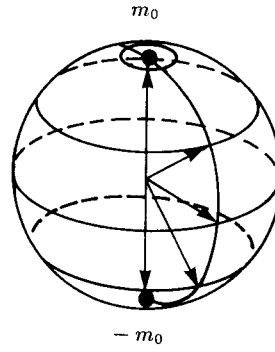


Figure 5.11 History of a half-twist.

where the subindex Q indicates that the path-integral is to be performed over configurations with a fixed winding number Q and $\mathcal{S}_0^E[\vec{m}]$ is the standard action of the non-linear sigma model

$$\mathcal{S}_0^E[\vec{m}] = \int d^2x \frac{1}{2g} (\nabla_i \vec{m})^2 \tag{5.7.3}$$

where space and time have been rescaled so as to have $v_s = 1$.

In this section we will consider the role of quantum fluctuations. We can do so by considering each topological class separately since these quantum fluctuations are local and do not alter the winding number. In other words, the winding number of a class of configurations cannot be changed by local fluctuations, since the former is a global property whereas the latter are purely local. Naturally, for this picture to hold it is necessary that the short distance (ultraviolet) and the long distance (infrared) of the theory remain separate. We will see that this is not the case in one-space-dimension. The behavior of the non-linear sigma model is dominated by *infrared* fluctuations. Thus the actual role, in detail, of topological sectors is unclear.

We will pretend that the fluctuations are local and reasonably small. This assumption amounts to a semiclassical treatment of the path-integral. Formally, this can be achieved only if the coupling constant g is small, i.e. in the limit $s \to \infty$. The standard perturbative treatment of the non-linear sigma model is thus equivalent, at low energies, to the $\frac{1}{s}$ expansion of the Heisenberg antiferromagnet [Haldane 83]. The classical action of the non-linear sigma model, Eq. (5.7.3), has a very important property: it is *scale invariant*. In other words the scale transformation

$$\begin{aligned} (x,t) &\to \lambda(x,t), \\ \vec{m} &\to \vec{m}, \end{aligned} \tag{5.7.4}$$

leaves the *action* invariant. Recall that $\vec{m}$ is dimensionless. The coupling constant g is also *dimensionless* in $(1+1)$ dimension. In higher dimensions, g is *dimensionful*. Let us define the *dimensionless coupling constant* u

$$u = g a_0^{2-d} \tag{5.7.5}$$

where d is the dimension of space-time. Thus the action now reads

$$\mathcal{S}_0^E[\vec{m}] = \frac{1}{2u a_0^{d-2}} \int d^d x \, (\nabla_i \vec{m})^2 \tag{5.7.6}$$

where $i = 1, \ldots, d$. For the sake of simplicity the discussion will be carried out in Euclidean space (i.e. imaginary time).

In Renormalization Group Theory [Wilson 74] the fact that the classical action is scale invariant means that $g = 0$ is a fixed point of the renormalization group (RG). I will define *a renormalization group transformation* by progressively integrating out the faster modes and obtaining an effective theory for the slower modes. In general, the sigma-field $\vec{m}$ will have Fourier components with momenta $\vec{p}$ ranging from the infrared ($|\vec{p}| \approx 0$) to the ultraviolet ($|\vec{p}| \approx \frac{1}{a_0}$). We can also use the constraint $\vec{m}^2 = 1$ to demand that one of the

components of the field $\vec{m}$, say m_3, has only fast components and be small [Kogut 79] . Let m_1 and m_2 be parametrized by m_3 and $\theta, (0 \leq \theta \leq 2\pi)$

$$m_1 = \sqrt{1-m_3^2}\cos\theta, \quad m_2 = \sqrt{1-m_3^2}\sin\theta \tag{5.7.7}$$

so as to solve the constraint $\vec{m}^2 = 1$. The Euclidean Lagrangian density now reads

$$\begin{aligned}\mathcal{L}_0^E &= \frac{1}{2ua_0^{d-2}}(\nabla_i \vec{m})^2 \\ &= \frac{1}{2ua_0^{d-2}}\left[(\nabla_i m_3)^2 + (1-m_3^2)(\nabla_i\theta)^2 + \frac{(m_3\nabla_i m_3)^2}{1-m_3^2}\right].\end{aligned} \tag{5.7.8}$$

Let us rescale m_3

$$m_3 = \sqrt{ua_0^{d-2}}\varphi \tag{5.7.9}$$

and write

$$\mathcal{L}_0^E = \frac{1}{2}(\nabla_i\varphi)^2 + \frac{1}{2ua_0^{d-2}}(1-ua_0^{d-2}\varphi^2)(\nabla_i\theta)^2 + \frac{1}{2}\left(\frac{ua_0^{d-2}}{1-ua_0^{d-2}}\right)(\varphi\nabla_i\varphi)^2. \tag{5.7.10}$$

We will be interested in the behavior for small g (i.e. small u). In this limit we can approximate $\mathcal{L}_0^E$ by the expression

$$\begin{aligned}\mathcal{L}_0^E =& \frac{1}{2}(\nabla_i\varphi)^2 + \frac{1}{2ua_0^{d-2}}(\nabla_i\theta)^2 - \frac{1}{2}\varphi^2(\nabla_i\theta)^2 + \\ &+ \frac{1}{2}ua_0^{d-2}(\varphi\nabla_i\varphi)^2 + \frac{1}{2}u^2a_0^{2(d-2)}\varphi^2(\varphi\nabla_i\varphi)^2 + 0(u^3).\end{aligned} \tag{5.7.11}$$

Both φ and θ have Fourier components all the way from zero momentum up to the cutoff $\Lambda \sim \frac{1}{a_0}$ The behavior at large momenta $|\vec{p}| \sim \Lambda$ should not affect very strongly phenomena taking place for small values of $\vec{p}$. It is then natural to integrate out such fluctuations.

Consider the momentum shell $b\Lambda < |\vec{p}| < \Lambda$ with $b < 1$ and the fluctuations with momenta inside that shell (fast modes). We now will carry out the functional integral

$$\begin{aligned}&\int_{b\Lambda<|\vec{p}|<\Lambda} \mathcal{D}\varphi e^{-\mathcal{S}_0^E[\varphi,\theta]} = \\ &\int_{b\Lambda<|\vec{p}|<\Lambda} \mathcal{D}\varphi(\vec{p}) e^{-\frac{1}{2}\int d^dx[(\nabla_i\varphi)^2 + \frac{1}{ua_0^{d-2}}(\nabla_i\theta)^2 - \varphi^2(\nabla_i\theta)^2 + \mathcal{O}(u)]}\end{aligned} \tag{5.7.12}$$

I will assume that θ is slowly varying and, hence, that its gradient, $(\nabla_i\theta)^2$, is small and it does not have Fourier components in the shell $b\Lambda < |\vec{p}| < \Lambda$

provided that $b \to 1$. Thus

$$\int_{b\Lambda<|\vec{p}|<\Lambda} \mathcal{D}\varphi(\vec{p}) e^{-\frac{1}{2}\int \frac{d^d p}{(2\pi)^d}\vec{p}^2|\varphi(\vec{p})|^2 + \frac{1}{2}\int \frac{d^d p}{(2\pi)^d}|\varphi(\vec{p})|^2(\nabla_i\theta)^2} \approx \prod_{b\Lambda<|\vec{p}|<\Lambda} \left[\frac{2\pi}{\vec{p}^2 - (\nabla\theta)^2}\right]^{\frac{1}{2}}. \tag{5.7.13}$$

The right-hand side of Eq. (5.7.13) can be exponentiated and approximated by the expression

$$\exp\left[\frac{1}{2}\int_{b\Lambda<|\vec{p}|<\Lambda} \frac{d^d p}{(2\pi)^d} \ln\frac{2\pi}{\vec{p}^2} + \frac{1}{2}(\nabla_i\theta)^2 \int_{b\Lambda<|\vec{p}|<\Lambda} \frac{d^d p}{(2\pi)^d}\frac{1}{\vec{p}^2}\right]. \tag{5.7.14}$$

To lowest order in u, the main effects of integrating out the fast modes in φ are twofold: (a) a shift of energy and (b) a shift, or renormalization, of the coupling constant u. Indeed, we can recast Eqs. (5.7.12) and (5.7.13) into the effective Lagrangian density

$$\begin{aligned}\mathcal{L}^E_{\text{eff}} = &-\frac{1}{2}\int_{b\Lambda<|\vec{p}|<\Lambda} \frac{d^d p}{(2\pi)^d} \ln\frac{2\pi}{\vec{p}^2} + \\ &+ \frac{1}{2}(\nabla_i\varphi)^2 + \frac{1}{2}\left(\frac{1}{u a_0^{d-2}} - \int_{b\Lambda<|\vec{p}|<\Lambda} \frac{d^d p}{(2\pi)^d}\frac{1}{\vec{p}^2}\right)(\nabla_i\theta)^2 - \frac{1}{2}\varphi^2(\nabla_i\theta)^2 + \\ &+ \mathcal{O}(u),\end{aligned} \tag{5.7.15}$$

with a momentum cutoff Λ' which has been *reduced* by b. Equivalently, the spatial cutoff a_0' has been *increased* by $\frac{1}{b}$:

$$\Lambda' = b\Lambda, \quad a_0' = \frac{a_0}{b}. \tag{5.7.16}$$

The effective Lagrangian density for the slower modes $\mathcal{L}^E_{\text{eff}}$ has the same form as the old Lagrangian density except for a constant shift, a new cutoff a_0' ($a_0' > a_0$) and a new renormalized coupling constant u' defined by

$$\frac{1}{u' a_0'^{d-2}} = \frac{1}{u a_0^{d-2}} - \int_{b\Lambda<|\vec{p}|<\Lambda} \frac{d^d p}{(2\pi)^d}\frac{1}{\vec{p}^2}. \tag{5.7.17}$$

After evaluating the integral, we get

$$\frac{1}{u' a_0'^{d-2}} = \frac{1}{u a_0^{d-2}} - \frac{S_d}{(2\pi)^d}\frac{1}{d-2}(1 - b^{d-2})\Lambda^{d-2} \tag{5.7.18}$$

where S_d is the area of the d-dimensional unit sphere. Since $b \to 1$ and $a_0' = \frac{a_0}{b}$ we can write

$$-\ln b = \frac{da_0}{a_0} \tag{5.7.19}$$

and find the β-function

$$\beta(u) = a_0\frac{du}{da_0} \tag{5.7.20}$$

to be given by

$$\beta(u) = -\epsilon u + \frac{u^2}{2\pi} + \mathcal{O}(u^3) \tag{5.7.21}$$

for $\epsilon = d - 2$ *small.*

In particular in $1 + 1$ dimensions ($d = 2$) we find a *positive* β-function [Polyakov 75].

$$\beta(u) = \frac{u^2}{2\pi}. \tag{5.7.22}$$

This result means that as the cutoff a_0 is increased, and we look at longer and longer distances, the fluctuations *increase* the effective value of the coupling constant at such scales (see Fig. (5.12)). Thus, even though the *bare* coupling constant $u_0 \propto \frac{1}{s}$ may be initially small, as we consider the effective theory at lower energies we find that the effective coupling ("effective s") increases (decreases). From classical statistical mechanics we know that the sigma model at strong coupling (i.e. the classical Heisenberg ferromagnet at high temperatures) is disordered and has a finite correlation length. Thus, in the language of the quantum spin chains, we get that as the "effective s" decreases the semi-classical behavior gets wiped-out. Instead we find a state *without* spontaneous symmetry breaking and with short range correlations.

5.8 Asymptotic Freedom and Haldane's Conjecture

In the last section we found the result that the effective coupling constant of the non-linear sigma model in $(1 + 1)$ dimensions increases with the length scale. We have chosen to present this result in the form of a β-function, Eq. (5.7.22), which measures the change of the coupling constant u as the cutoff a_0 (the lattice constant) is increased and the fast degrees of freedom of the system are progressively integrated out. Alternatively, we could have kept the cutoff fixed and varied a physical scale such as the length L of the chain or an energy scale such as the temperature T.

At finite temperature T, the system can be viewed (in imaginary time) as a non-linear sigma model on a strip of length L (the linear size of the chain) and width $\frac{1}{T}$ with periodic boundary conditions in imaginary time. This is the standard statement that the partition function of a quantum system, with a

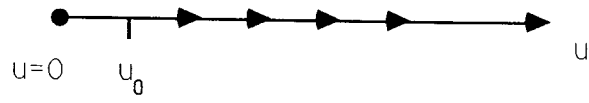


Figure 5.12 Renormalization Group Infrared Flow for one-dimensional Quantum Spin-Chains; u_0 is the bare coupling constant ($u_0 = \frac{2}{s} a_0^{2-d}$).

global symmetry, in d space dimensions is equivalent to a classical mechanics problem in $(d+1)$ dimensions with imaginary time being the extra dimension [Fradkin 78] (Fig. 5.13). The renormalization group of the last section can easily be generalized to an anisotropic system with spin wave velocity $v_s \neq 1$ which is kept fixed in the RG process.

We begin our RG process with some lattice constant a_0, bare coupling $u_0 \propto \frac{1}{s}$ and spin wave velocity v_s. As we integrate out degrees of freedom the effective coupling grows and the spatial cutoff increases. At some point, the cutoff a becomes of the order of $\frac{v_s}{T}$. At this point the quantum fluctuations are negligible since the cutoff is as large as the width of the strip and we have effectively a *non-linear* sigma model at finite temperature T. In turn the non-linear sigma-model, in imaginary time, is identical to the *classical* Heisenberg model in d space-time Euclidean dimensions. It can be easily proven that a classical Heisenberg model (or non-linear sigma model) in one dimension, like all one-dimensional classical systems with short-range interactions, has a finite correlation length ξ_c at all temperatures [Landau 75].

We can now ask how much does the effective coupling u differ from the bare coupling u_0 if the cutoff is changed from a_0 to $\bar{a}_0 \sim \frac{v_s}{T}$. The β-function tells us what is the dependence of u on the cutoff, at least for small enough u. The result of integrating the differential equation

$$\beta(u) \equiv a_0 \frac{du}{da_0} = \frac{u^2}{2\pi} \tag{5.8.1}$$

is

$$\frac{1}{u(\bar{a}_0)} = \frac{1}{u(a_0)} + \frac{1}{2\pi} \ln\left(\frac{a_0}{\bar{a}_0}\right). \tag{5.8.2}$$

By choosing $\bar{a}_0$ to be of the order of $\frac{1}{T}$

$$\bar{a}_0 = \frac{v_s}{T} \tag{5.8.3}$$

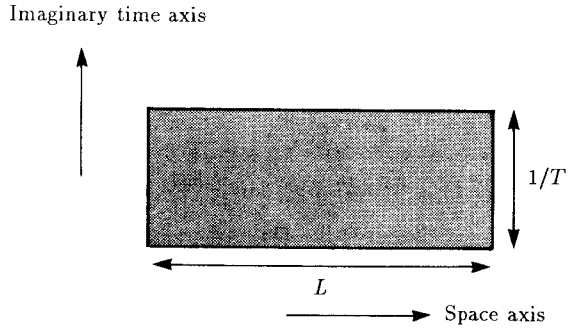


Figure 5.13 Euclidean space-time for a system of length L at temperature T. With periodic boundary conditions in imaginary time, Euclidean space-time is a cylinder.

we find the temperature dependence of u to be

$$\frac{1}{u(T)} = \frac{1}{u_0} + \frac{1}{2\pi} \ln\left(\frac{a_0 T}{v_s}\right). \tag{5.8.4}$$

Equivalently, we can write

$$u(T) = \frac{u_0}{1 + \frac{u_0}{2\pi} \ln(\frac{a_0 T}{v_s})}. \tag{5.8.5}$$

Thus, at high temperatures, $T \gg v_s/a_0$, we find that $u(T)$ becomes small

$$u(T) \approx \frac{2\pi}{\ln(\frac{a_0 T}{v_s})} \to 0 \quad \text{for} \quad T \to \infty. \tag{5.8.6}$$

In other words, the effective coupling at short distances or at high temperatures is *small*. This result is known as asymptotic freedom and, in this context, was first discussed by Polyakov [Polyakov 75].

Conversely, as the temperature is lowered, the effective coupling u becomes large (Fig. 5.14). Equation (5.8.5) exhibits an apparent divergence at the temperature T_0 where

$$T_0 \approx \frac{v_s}{a_0} e^{-\frac{2\pi}{u_0}} = \frac{v_s}{a_0} e^{-\pi s}. \tag{5.8.7}$$

The meaning of T_0 is that of the temperature at which the weak coupling (i.e. $\frac{1}{s}$) expansion breaks down. To continue down to lower temperatures, we must take into account the fact that for $T \leq T_0$, the sigma model has a large effective coupling. At this point we notice that, at large values of the coupling constant, the sigma model is disordered no matter the dimensionality of space-time. Thus we expect a finite, and short, correlation length ξ. The effective coupling should saturate due to lattice effects and the constraint $\vec{m}^2 = 1$. These ideas have been confirmed by Monte Carlo RG studies [Shenker 80].

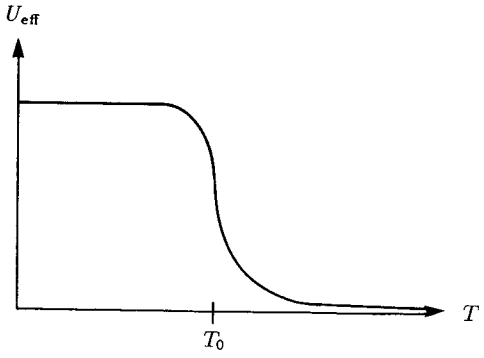


Figure 5.14 Crossover of the effective coupling $u_{eff}(T)$.

We can also use the RG to estimate the dependence of the correlation length ξ on the bare coupling constant $u_0 = \frac{2}{s}$. Under the RG, the correlation length, like all other physical observables, remains invariant. From dimensional analysis we expect ξ, which is a length scale, to have the form

$$\xi(u) = a_0 f(u) \tag{5.8.8}$$

where f is a function of u, the coupling constant *at the scale* a_0. Being an RG invariant, it must obey

$$a_0 \frac{d\xi}{da_0} = 0 \tag{5.8.9}$$

which implies that $f(u)$ satisfies the differential equation

$$\beta(u)\frac{df}{du} + f(u) = 0. \tag{5.8.10}$$

The solution to Eq. (5.8.10) is

$$f(u) = f(u')e^{-\int_{u'}^{u} \frac{dz}{\beta(z)}} = f(u')e^{2\pi(\frac{1}{u} - \frac{1}{u'})} \tag{5.8.11}$$

where u and u' are connected by the RG flow.

Consider now the correlation length ξ at *two* different values of u, u_1 and u_2 for the *same* value of the lattice constant a_0. Let u^* be a large reference value of the coupling u. From Eqs. (5.8.8) and (5.8.11), we find

$$\xi(u_i) = a_0 f(u_i) = a_0 f(u^*) e^{-\int_{u^*}^{u_1} \frac{du}{\beta(u)}} \tag{5.8.12}$$

for both $i = 1, 2$. Thus the *ratio* of values of ξ for two different couplings and equal lattice spacing a_0 is given by

$$\frac{\xi(u_1)}{\xi(u_2)} = \frac{a_0 f(u^*) e^{-\int_{u^*}^{u_1} \frac{dz}{\beta(z)}}}{a_0 f(u^*) e^{-\int_{u^*}^{u_2} \frac{dz}{\beta(z)}}}. \tag{5.8.13}$$

Thus, we get

$$\frac{\xi(u_1)}{\xi(u_2)} = e^{-\int_{u_2}^{u_1} \frac{dz}{\beta(z)}}. \tag{5.8.14}$$

The integral can be easily evaluated to find

$$\frac{\xi(u_1)}{\xi(u_2)} = e^{(\frac{2\pi}{u_1} - \frac{2\pi}{u_2})}. \tag{5.8.15}$$

We find for the case in which $u_1 = u_0 = \frac{2}{s}$ and u_2 is large that

$$\xi(u_0) \approx \xi(u_2) e^{\pi s}. \tag{5.8.16}$$

What value should we assign to $\lim_{u_2 \to \infty} \xi(u_2)$? The answer depends on whether the spin is integer or half-integer.

(a) Integer spin: In this case we do not get a topological term. As was emphasized above, the sigma model is always disordered at strong coupling. Thus, we expect $\xi(u_2) \approx a_0$ and we find a *finite* correlation length

$$\xi_0 = \xi(u_0) \approx a_0 e^{\pi s}. \tag{5.8.17}$$

There is no long-range order (i.e. no Néel state). The spectrum has a gap

$$\Delta = \frac{v_s}{\xi_0} \tag{5.8.18}$$

and the ground state is unique. Equation (5.8.17) shows that the correlation length is non-perturbative in the $\frac{1}{s}$ expansion.

(b) Half-integer spin: The sigma model coupling constant u still scales to strong coupling but the topological term remains unchanged at the value $\theta = 2\pi s \pmod{2\pi}$. However, the coupling constant $g \propto u$ is related to the spin through $s = \frac{2}{g}$. Thus strong coupling is equivalent to low spin. Hence the behavior of *all* half-integral spin chains is qualitatively identical to the spin one-half case for which $u_0 \propto 4$. The spin one-half case is *gapless*, as we saw from the Bethe-Ansatz and other approaches. Thus, $\xi(\infty)$ is still infinite. All half-integral spin chains are at a critical point with infinite correlation length. At first sight, this result seems to be paradoxical. We started with smooth configurations with well defined winding numbers and a weak coupling g. As the energy scale was lowered the effective coupling of the sigma model grew *but* the topological coupling was unaffected. Thus, at low energies, the configurations become *rough* and it is unclear what is the actual meaning of the topological term in this situation. This poses no problems for the *integer* spin chains since the topological term does not contribute in this case ($\theta = 2\pi s$). In contrast, for half-integral chains, this result simply means that *all* $s > \frac{1}{2}$ systems behave *qualitatively* in the same way as the $s = \frac{1}{2}$ case.

This result, s integer is disordered, and s half-integer is critical, is known

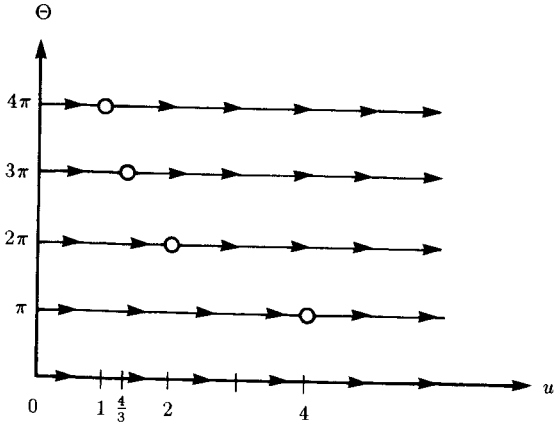


Figure 5.15 RG flows of QHA chains for $s = \frac{1}{2}, 1, \frac{3}{2}, 2$. The open circles represent their σ-model bare coupling constant $u \approx \frac{2}{s}$. They all iterate to $u^* = \infty$. The difference in their behaviour is a consequence of the presence of a $\theta = \pi$ term in the σ-model for the half-integer spin chains.

as Haldane's conjecture [Haldane 83]. It has also been checked by accurate numerical calculations using exact diagonalization on finite (but large) chains [Moreo 87], [Ziman 87] and by Green's function Monte Carlo simulations [Liang 90]. Affleck and Haldane [Affleck 87] have also found the same result using non-Abelian bosonization [Witten 84]. The RG flows are shown in Fig. 5.15.

5.9 Hopf Term or No Hopf Term?

The one-dimensional spin chains have a very unusual behavior: disorder (integer spin) or critical (half-integer spin) ground states, neutral fermions which are massless for the half-integer case and massive for s integer, etc. There is nothing in this picture which is remotely close to the physics that emerges from the mean-field theory of Chapter 3. It is then natural to ask whether or not this picture is peculiar to one-dimensional systems or if there is a natural generalization to higher dimensions. It is a trivial matter to generalize the one-dimensional formalism to the case of a square lattice. The lattice action is a simple generalization of Eq. (5.5.1). Let $\vec{r}$ span a square lattice of size $N \times N$: $\vec{r} = (x_1, x_2)$, where $x_1, x_2 = 1 \ldots N$. I will assume that N is even. The action is

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s \sum_{\vec{r}} \mathcal{S}_{\mathrm{wz}}[\vec{n}(\vec{r}\,)] - \int_0^T dx_0 \sum_{\langle \vec{r}, \vec{r}' \rangle} J s^2 \vec{n}(\vec{r}, x_0) \vec{n}(\vec{r}\,', x_0) seqnu$$

where $\vec{r}$ and $\vec{r}\,'$ are nearest neighboring sites on the square lattice. Since the square lattice is bipartite and we expect short-range Néel order, we will once again stagger the field configurations and find

$$\mathcal{S}_{\mathrm{M}}[\vec{n}] = s \sum_{\vec{r}} (-1)^{x_1 + x_2} \mathcal{S}_{\mathrm{wz}}[\vec{n}(\vec{r}\,)] + \int_0^T dx_0 \sum_{\langle \vec{r}, \vec{r}' \rangle} J s^2 \vec{n}(\vec{r}, x_0) \vec{n}(\vec{r}\,', x_0). \tag{5.9.1}$$

It is straightforward, but tedious, to derive the effective action for the slow varying fields. Once again, on the basis of symmetry, we expect a non-linear sigma model. The issue is whether or not there is a topological term in the effective action.

Before deriving the effective action by an explicit calculation, let us consider what topological terms are possible. In the $(1+1)$-dimensional case we saw that the configurations were classified in terms of an index, the topological charge, which labels the homotopy class of the configuration. The existence of such an index was guaranteed by the fact that the configurations fall into homotopy classes which form the group $\Pi_2(S_2)$ of smooth maps of the two-dimensional Euclidean space-time S_2 into the S_2 of the order parameter manifold. This homotopy group $\Pi_2(S_2)$ is isomorphic to $\mathcal{Z}$, the group of integers, i.e. the winding numbers of the topological classes. In $(2+1)$–dimensions the situation is rather different. Once again, the Euclidean space-time can be

regarded as a sphere S_3 and the configurations are maps of S_3 into S_2. However, there are no smooth solutions of the classical Euclidean equations of motion with non-trivial winding numbers. There are *singular* solutions which are known as hedgehogs (Fig. 5.16) but which have linearly divergent action. Haldane has recently speculated that these hedgehogs may become relevant if the sigma model becomes disordered by some mechanism [Haldane 88].

On the other hand, there are non-trivial configurations in Minkowski space-time (i.e. real time). Consider at some time $t = t_0$ a configuration of sigma-model fields identical to one of the instantons of Section 5.6. Now it represents the snapshot of an eigenstate and it is called a soliton. Thus the *configuration space* of a two-dimensional quantum non-linear sigma model is also a sphere S_2 and is usually denoted by $\Omega_2 S_2$. Consider now the real-time evolution of such a state with periodic boundary conditions in time, i.e. we consider histories in which the initial state is the same as the final state. Thus, a history is a closed curve in the configuration space $\Omega_2 S_2$. In quantum mechanics we are told to sum over all histories and to assign a phase to each history, i.e. to each curve in $\Omega_2 S_2$. Since a phase is an element of S_1 (the unit circle) we have constucted the set of maps $\Pi_1(\Omega_2 S_2)$. However, we know that the configurations at any given time are maps of S_2 (space) into S_2 (field), i.e. Homotopy classes of $\Pi_2(S_2)$ which we saw was isomorphic to the group of integers $\mathcal{Z}$. Hence the configuration space $\Omega_2 S_2$ is decomposed into a disjoint union of path-connected pieces, each characterized by the winding number or soliton number $\mathcal{Q}$. Thus each disconnected piece of the Hilbert space will have a separate time evolution and will have to be summed with separate phases. Since the classical paths are continuous curves in $\Omega_2 S_2$ classified by $\Pi_3(S_2) = \mathcal{Z}$, the relevant issue is now what topolgical invariant is associated with such histories.

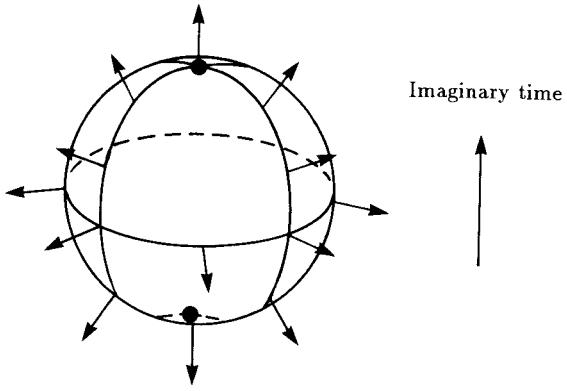


Figure 5.16 A hedgehog.

Consider a history of the order parameter field $\vec{m}(\vec{x},t)$ in $(2+1)$ dimensions. We can define a *topological current* J_μ by

$$J_\mu = \frac{1}{8\pi}\epsilon_{\mu\nu\lambda}\epsilon_{abc}m_a\partial^\nu m_b\partial^\lambda m_c \tag{5.9.2}$$

with $\mu = 0,1,2$ and $a,b,c,= 1,2,3$. The topological current J_μ is clearly conserved

$$\partial^\mu J_\mu = 0. \tag{5.9.3}$$

Therefore the total *topological charge* $Q = \int d^2x J^0(\vec{x},t)$ is constant in time

$$Q = \int d^2x J^0(\vec{x},t) = \int d^2x \frac{1}{8\pi}\epsilon_{0ij}\epsilon_{abc}m_a\partial^i m_b\partial^j m_c. \tag{5.9.4}$$

Clearly Q is identical to the winding number $\mathcal{Q}$ of Eq. (5.6.4).

Consider now a soliton state with $\mathcal{Q} = 1$ (Fig. 5.8). Imagine a time evolution in which the soliton rotates slowly around its center and executes exactly n turns during its lifespan. Each point on the equator of the soliton traces a curve ("worldline") which wraps n-times around the other curves traced by the other points (Fig. 5.17). An easy way to compute the winding number of this history is to imagine that each worldine is a wire carrying the unit of current. As the soliton rotates, the worldlines ("wires") are *braided*. The natural topological invariant is the *linking number* of these worldlines (Fig. 5.18). If we denote by $\vec{j}$ the current carried by the wires and by $\vec{B}$ the magnetostatic field they create, the linking number is simply given by Ampere's law

$$\int d^3x \vec{j}\cdot\vec{B} = 2\pi n \tag{5.9.5}$$

where n is the number of turns.

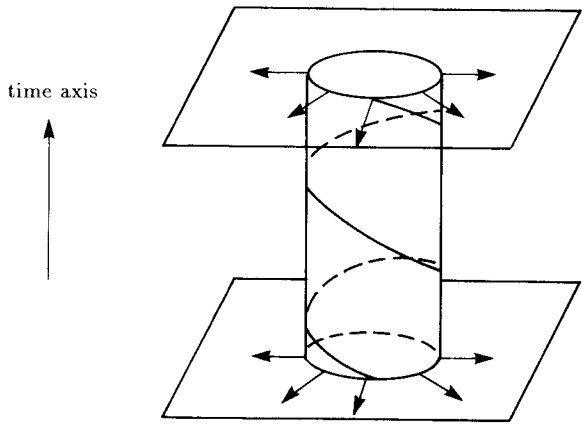


Figure 5.17 A world tube with linking number +2.

We can make this analogy more precise by using the *Hopf map* which maps the three sphere S_3 onto S_2. Let z_1 and z_2 be two complex number satisfying

$$|z_1|^2 + |z_2|^2 = 1. \tag{5.9.6}$$

Clearly (z_1, z_2) span S_3. Define now the *spinor* $z_\alpha, (\alpha = 1, 2)$. The order parameter field $\vec{m}$ is related to z_α through the map

$$m^a = z^*_\alpha \sigma^a_{\alpha\beta} z_\beta \tag{5.9.7}$$

where $\{\sigma^a\}_{a=1,2,3}$ are the Pauli matrices. The order parameter $\vec{m}$ also satisfies $\vec{m}^2 = 1$. This is the Hopf map.

It is clear that (z_1, z_2) has three independent parameters whereas $\vec{m}$ only has two. But one of these parameters, or degrees of freedom, is unobservable since a global chage of phase of the spinor z_α

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \rightarrow e^{i\phi} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{5.9.8}$$

does not lead to any observable effects since $\vec{m}$ is *invariant* under such *gauge transformations.*

Furthermore, the action of the non-linear sigma model itself can be written in terms of the spinor field z_α. This is the CP^1 model. To simplify matters I will consider the problem with spin-wave velocity $v_s = 1$. Let z_α be a CP^1

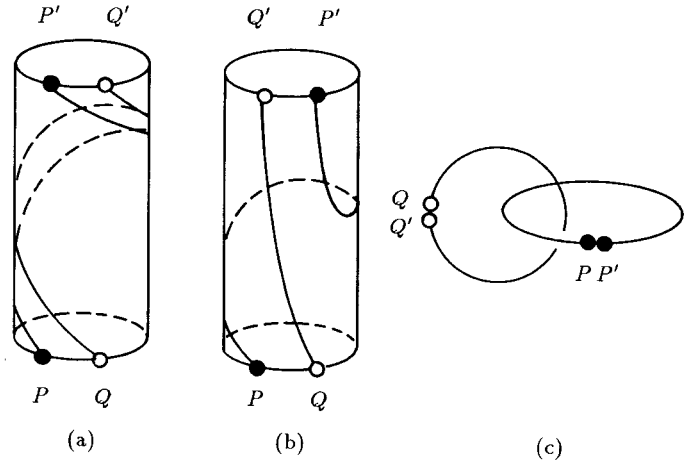


Figure 5.18 (a) Two world lines PP' and QQ' with linking number $+1$. (b) Smooth deformation of the world line PP' and QQ'. (c) Periodic boundary conditions in time are enforced, P and P' as well as Q and Q' are identified and the world lines form a braid.

field and A_μ an unconstrained gauge field, with Lagrangian density

$$\mathcal{L}_{\mathrm{CP}^1} = \frac{1}{2g}|D_\mu z|^2 \tag{5.9.9}$$

where g is a coupling constant and D_μ is the covariant derivative

$$D_\mu = \partial_\mu - iA_\mu. \tag{5.9.10}$$

The functional integral is

$$Z = \int \mathcal{D}\bar{z}\mathcal{D}z\delta(|z|^2 - 1)\int \mathcal{D}A e^{i\mathcal{S}_{\mathrm{CP}^1}[z,A]}. \tag{5.9.11}$$

Since $\mathcal{L}_{\mathrm{CP}^1}$ is quadratic in A_μ, it can be integrated out exactly by a saddle point calculation. The saddle point condition

$$\frac{\delta\mathcal{L}_{\mathrm{CP}^1}}{\delta A_\mu} = 0 \tag{5.9.12}$$

determines the gauge field in terms of the CP^1 field

$$A_\mu = \frac{i}{2}(z_\alpha^*\partial_\mu z_\alpha - z_\alpha\partial_\mu z_\alpha^*) \equiv -\frac{i}{2}z_\alpha^*\partial_\mu^{\leftrightarrow} z_\alpha. \tag{5.9.13}$$

By substituting Eq. (5.9.14) in the Lagrangian density, Eq. (5.9.10), one finds

$$\frac{1}{2g}(\partial_\mu\vec{m})^2 = \frac{1}{g}|D_\mu z|^2. \tag{5.9.14}$$

In other words, the CP^1 model and the $O(3)$ non-linear sigma model are equivalent.

The topological invariant, or *Hopf invariant*, has a very simple and natural form in terms of the vector potentials A_μ. Consider a term in the Lagrangian density of the form

$$\mathcal{L}_{\mathrm{CS}} = \frac{\theta}{4\pi}\epsilon_{\mu\nu\lambda}A^\mu F^{\nu\lambda} \tag{5.9.15}$$

which is known as a Chern-Simons term. The gauge field A_μ is *constrained* to be given by Eq. (5.9.14) and its field strength can be related back to the sigma model field $\vec{m}$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \equiv \vec{m}\cdot(\partial_\mu\vec{m}\times\partial_\nu\vec{m}). \tag{5.9.16}$$

Thus, the flux associated with the gauge field A_μ is simply related to the topological current. The *Hopf invariant* H is simply

$$H = \frac{\theta}{8\pi^2}\int d^3x\epsilon_{\mu\nu\lambda}A^\mu F^{\nu\lambda} \tag{5.9.17}$$

with A_μ and $F_{\nu\lambda}$ given by Eqs. (5.9.14) and (5.9.17). We will see in chapter 7 that a non-zero value of θ will change the statistics of the solitons.

But, is there a Hopf term in the effective action of the quantum Heisenberg antiferromagnet in two dimensions? The only way to determine that is to compute the effective action carefully. [Dzyaloshinskii 88] and [Wiegmann 88] have conjectured that the effective action has a Hopf term with $\theta = 2\pi s$. This is a subtle business since Wu and Zee have shown that, in its CP^1 form, the

Hopf term is a total derivative which does not alter the equations of motion but changes the spin and statistics of the topological excitations [Wu 84].

Let us first derive the effective action, following the methods of Fradkin and Stone [Fradkin 88]. The result will be a sigma model *without* a topological term [Fradkin 88], [Haldane 88], [Wen 88], [Ioffe 88], [Dombre 88].

First, we need to integrate out the fast degrees of freedom. We write

$$\vec{n}(\vec{r}\,) = m(\vec{r}\,) + (-1)^{x_1+x_2} a_0 \vec{l}(\vec{r}\,). \tag{5.9.18}$$

Following the *same* procedure which in the one-dimensional case led to a sigma model *with* a topological term (see Eq. (5.5.11)), we find

$$\mathcal{L}^M_{\text{eff}}(\vec{m},\vec{l}) = -\frac{Js^2}{2}\left((\partial_i \vec{m})^2 + 8\vec{l}^{\,2}\right) + \frac{s}{a_0}\vec{l}\cdot(\vec{m}\times\partial_0\vec{m}). \tag{5.9.19}$$

If we now proceed to integrate out the fast modes, the $\vec{l}$ field, we find a non-linear sigma model *without* a topological term. The (bare) coupling constant and (bare) spin-wave velocity are given by (see Eqs. (5.5.14) and (5.5.15))

$$\begin{aligned} g &= \sqrt{2}a_0\frac{2}{s} \\ v_s &= \sqrt{2}2a_0 Js. \end{aligned} \tag{5.9.20}$$

The terms, which in the one-dimensional case gave rise to the toplogical term, now have cancelled each other out. The reason for this cancellation can be traced back to the staggered character of the Néel state. Naively, we expect that each row will make a contribution similar to the one-dimensional result. But neighboring rows are staggered in the opposite way. The result is that the terms originating from each neighboring rows, now, effectively cancel out. We are assuming a lattice with even rows and columns. In the case of an *odd* number of rows, we may get a non zero contribution from the last row. However, this is a boundary condition effect which, incidentally, was not needed in the case of the chains. But we do expect to see changes in the spectrum of elementary excitations if we change the boundary conditions.

The argument which led to the cancellation is a bit too naive and maybe dangerous. We know from the work of Wu and Zee that, at least in the CP^1 representation, the Hopf term is a total derivative. Thus, a local cancellation is not a sufficient argument for the study of a *global* effect. Slowly varying configurations may have an accumulated effect near the boundaries and yield a non-zero answer. We can dispel these fears by computing the alternating sum Φ

$$\Phi = s\sum_{\vec{r}}(-1)^{x_1+x_2}\mathcal{S}_{\text{wz}}[\vec{n}(\vec{r}\,)] \tag{5.9.21}$$

for a configuration which, in the continuum limit, has soliton number $\mathcal{Q} = 1$. If we let this soliton configuration rotate slowly around its center such that it turns exactly once during its history, the history of this configuration should have Hopf number or linking number $+1$. We should choose a lattice configuration which, in the limit of soliton radius r_s *large* compared with the lattice spacing a_0 should go smoothly over to the continuum soliton. Any

soliton profile should do the job. For instance, we can imagine a configuration obtained by a stereographic configuration (Fig. 5.19). The area swept by each spin is $a = 2\pi R^2(1-\cos\theta)$. Thus the sum $\Phi(N)$, for a system of size $2N \times 2N$, is given by

$$\Phi(N) = \sum_{n,m=-N+1}^{N} 2\pi s R^2 \left(1 - \cos\theta(n,m)\right)(-1)^{n+m}. \tag{5.9.22}$$

The sphere has radius R and its south pole, which has coordinates (α_1, α_2), is in the first unit cell. The radius of the soliton r_s is equal to the diameter $2R$ of the sphere, if we define the radius as the location in which the spins are orthogonal to the asymptotic configuration at spatial infinity. Hence we find

$$\Phi(N) = \sum_{n,m=-N+1}^{N} \frac{16\pi s R^4 e^{i\pi(n+m)}}{4R^2 + (n-\alpha_1)^2 + (m-\alpha_2)^2}. \tag{5.9.23}$$

In the thermodynamic limit, $N \to \infty$, and by making use of the Poisson Summation Formula

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{n=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{i2\pi kn} f(k), \tag{5.9.24}$$

we get for $\Phi \equiv \lim_{N\to\infty} \Phi(N)$

$$\Phi = \int \frac{d^2k}{(2\pi)^2} \sum_{\vec{n}} \frac{16\pi s R^4 e^{i2\pi(\vec{k}+\vec{\alpha})\cdot(\vec{n}+\vec{G})}}{4R^2 + \vec{k}^2} \tag{5.9.25}$$

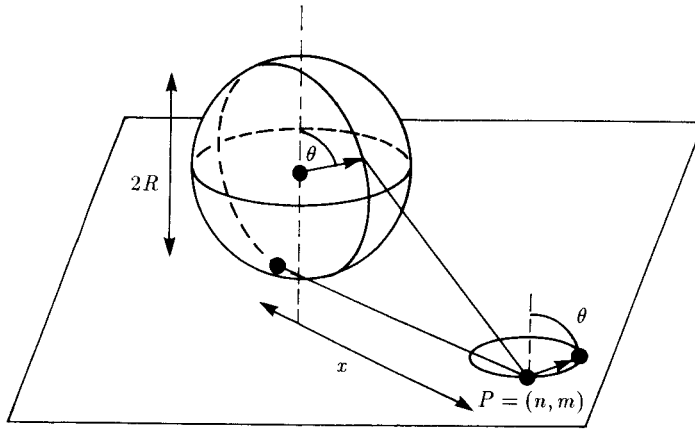


Figure 5.19 A soliton configuration can be generated using a stereographic projection. The spin is paralell-transported from the sphere to point $P = (n, m)$. Its history is pictured as an ellipse.

where $\vec{G} = (\frac{1}{2}, \frac{1}{2})$. In the limit $R \to \infty$ it is easy to see that Φ is exponentially small since we can write

$$\Phi = \sum_{n_1,n_2=-\infty}^{+\infty} 8sR^4 e^{i2\pi\vec{\alpha}\cdot(\vec{n}+\vec{G})} K_0(\pi r_s|\vec{n}+\vec{G}|) \tag{5.9.26}$$

where $K_0(x)$ is a modified Bessel function. Thus, for $r_s \gg 1$, we may keep just the leading terms:

$$\Phi \approx 4s \left(\frac{r_s}{\sqrt{2}}\right)^{5/2} e^{-\frac{2\pi}{\sqrt{2}}r_s + i2\pi(\alpha_1+\alpha_2)} \cos(\pi\alpha_1)\cos(\pi\alpha_2). \tag{5.9.27}$$

This expression vanishes exponentially fast for solitons with radii $r_s \gg \frac{1}{\pi\sqrt{2}}$. Notice that even fairly small solitons with radii $r_s \approx 1$ are large according to this criterion. We must conclude that if we expect to see Néel order (even if this was true only at short distances!) the effective theory at long distances is given by a non-linear sigma model with renormalized coupling constant and spin-wave velocity. Phenomenologically this is what the experiments in $La_2Cu_1O_4$ seem to indicate [Chakravarty 88]. In Section 5.7 we calculated the one-loop β-function for the non-linear sigma model in $(2+\epsilon)$ dimensions (here 2 means $(1+1)$). We found the result (see Eq. (5.7.21))

$$\beta(u) = -\epsilon u + \frac{u^2}{2\pi}. \tag{5.9.28}$$

For space-time dimensions $d > 2$, the fixed point at the origin is *stable*. This means that if the bare dimensionless coupling constant u is sufficiently small, the effective coupling flows toward the $u = 0$ fixed point and we have a Néel state with weakly coupled spin-waves. Equation (5.7.21) has another fixed point at $u^* \approx 2\pi\epsilon$, which is *infrared unstable*. This fixed point is the location of a second-order phase transition (in terms of the coupling constant). Beyond this fixed point, i.e. for $u > u^*$, the effective coupling flows toward the $u = \infty$ fixed point just as in the (1+1)-dimensional case. However, we do not have a topological term anymore. Thus, we must conclude that, for $u > u^*$ the system is disordered at distances longer than some correlation length $\xi \sim |u-u^*|^{-\nu}$ $\left(\nu = \frac{1}{d-2} + \mathcal{O}(d-2)\right)$ and Néel-like order at scales between the lattice constant and ξ. Such a state is a *zero-temperature quantum paramagnet* (QP), i.e. a paramagnetic state driven purely by quantum fluctuations and in the absence of thermal fluctuations. A finite correlation length without long-range order means that the ground state is unique and there is a gap $\Delta = v_s/\xi$ for the elementary excitations ("spin-waves").

The theory described here, which is based in the $(2+\epsilon)$ expansion, is too crude to reliably predict the value of u^*. Since we saw that our approximations were equivalent to (a resummation of) the $\frac{1}{s}$ expansion, we must also conclude that u^* cannot be calculated with confidence from the $\frac{1}{s}$ expansion either. Qualitatively, we should still expect a non-trivial fixed point for $\epsilon = 1$. The perturbative β-function predicts that for $\epsilon \approx 1$ even $s = \frac{1}{2}$ on a square lattice is on the Néel side of the phase transition although not far from it. This

result appears to be consistent with existent experimental data on quasi-two-dimensional systems believed to be reasonably well described by the $s = \frac{1}{2}$ quantum Heisenberg antiferromagnet such as $La_2Cu_1O_4$. Experimentally [Shirane 87] one sees a Néel state but with a magnetic moment about fifty percents of its classical value. The dynamical structure factor predicted by the σ-model [Chakravarty 88] is also confirmed by these experiments. Numerical calculations on two-dimensional quantum Heisenberg models also exhibit a similar behavior [Liang 88], [Liang 90].

References

[Affleck 85] I. Affleck. Phys. Rev. Lett. 54, 966 (1985).

[Affleck 87] I. Affleck and F.D.M. Haldane. Phys. Rev. B 36, 5291 (1987).

[Affleck 88] I.K. Affleck. In *Strings, Fields and Critical Phenomena*, Les Houches Summer School 1988, Session XLIX, edited by E. Brezin and J. Zin-Justin, North-Holland, 1990.

[Anderson 87] P. W. Anderson. Science 235, 1196 (1987).

[Babudjian 86] H.M. Babudjian and A.M. Tsvelick. Nucl. Phys. B 265, 24 (1986).

[Baym 74] G. Baym. *Lectures on Quantum Mechanics*, Benjamin, 1974.

[Bloch 30] F. Bloch. Z. Physik 61, 206 (1930).

[Chakravar 88] S. Chakravarty, B. Halperin and D. Nelson. Phys. Rev. Lett. 60, 1057 (1988).

[Dombre 88] D. Dombre and N. Read. Phys.Rev. B 38, 7181 (1988).

[Dzyaloshin 88] I. Dzyaloshinskii, A. Polyakov and P. Wiegmann. Phys. lett. 127 A, 112 (1988).

[Fradkin 78] E. Fradkin and L. Susskind. Phys. Rev. D 17, 2637 (1978).

[Fradkin 88] E. Fradkin and M. Stone. Phys. Rev. B 38, 7215 (1988).

[Georgi 82] H. Georgi. *Lie Algebras in Particle Physics*, Benjamin, 1982.

[Haldane 83] F.D.M. Haldane. Phys. Lett. 93A, 464 (1983) and Phys. Rev. Lett. 50, 1153 (1983).

[Haldane 85] F.D.M. Haldane. J. Appl. Phys. 57, 3359 (1985).

[Haldane 88] F.D.M. Haldane. Phys. Rev. Lett. 61, 1029 (1988)).

[Holstein 40] T. Holstein and H. Primakoff. Phys. Rev. 58, 1098 (1940).

[Ioffe 88] L. Ioffe and A. Larkin. Int. Jour. Mod. Phys B 2, 203 (1988).

[Klauder 79] J. Klauder. Phys. Rev. D 19, 2349 (1979).

[Kogut 79] J.B. Kogut. Rev. Mod. Phys. 51, 659 (1979).

[Landau 75] L. Landau and E.M. Lifshitz. *Statistical Physics*, Pergamon, 1975.

[Landau 75] L. Landau and E.M. Lifshitz. *Statistical Physics*, Pergamon, 1975.

[Liang 88] S.D. Liang, B. Doucot and P.W. Anderson. Phys. Rev. B 61, 365 (1988).

[Liang 90] S. Liang. Phys. Rev. Lett. 64, 1597 (1990).

[Moreo 87] A. Moreo. Phys. Rev. B 36, 8582 (1987).

[Perelomov 86] A. Perelomov. *Generalized Coherent States and Their Applications*, Springer, 1986.

[Polyakov 75] A.M. Polyakov. Phys. Lett. 59B, 79 (1975).

[Schulman 81] L.S. Schulman. *Techniques and Applications of Path Integration*, Wiley, 1981.

[Shenker 80] S. H. Shenker and J. Tobochnik. Phys. Rev. B 22, 4462 (1980).

[Shirane 87] G. Shirane, Y. Endoh, R. Birgeneau and M. Kastner. Phys. Rev. Lett. 59, 1613 (1987).

[Stone 86] M. Stone. Phys. Rev. D 33, 1191 (1986).

[Wen 88] X.G. Wen and A. Zee. Phys. Rev. Lett. 61, 1025 (1988).

[Wiegmann 88] P. Wiegmann. Phys. Rev. Lett. 60, 821 (1988).

[Wilson 74] K.G. Wilson and J.B. Kogut. Physics Reports C 12, 74 (1974).

[Witten 83] E. Witten. Nucl.Phys. B 223, 422 (1983).

[Witten 84] E. Witten. Comm. Math. Phys. 92, 455 (1984).

[Wu 84] Y.S. Wu and A. Zee. Phys. Lett. 147B, 325 (1984).

[Ziman 87] T. Ziman and H.J. Schultz. Phys. Rev. Lett. 59, 140 (1987).

CHAPTER 6

The Spin Liquid States

In the previous chapters we discussed mostly ordered Néel-like ground states of spin systems. The sole exception was the case of the spin-chains in which the ordered state is *always* destroyed by quantum fluctuations. In this chapter we begin a discussion of the ground states of quantum magnets which, as a result of strong fluctuations, loose the long range order of their spin degrees of freedom. The key driving force behind this quantum disorder is *frustration*.

6.1 Frustration and Disordered Spin States

It is possible to drive a Heisenberg model toward a disordered state. One way to do that is to add extra interactions which, if strong enough, may destroy the Néel behavior. A popular choice is to consider next-nearest neighbor interactions with strength K (Fig. 6.1). These interactions *frustrate* the system in the sense that, for J close to K, the *classical* Néel state becomes degenerate in energy with other classical configurations which differ from it by *local* spin flips. Quantum mechanically, one may expect a substantial increase of fluctuations which should further decrease the value of the moment.

By following the steps that led to the non-linear sigma model (see sections 5.5 and 5.9) and to the bare coupling constant g and spin-wave velocity v_s (Eq. (5.9.21)),we can compute the new values of g and v_s if we assume that at least the short-range order has the Néel structure of $K = 0$. Clearly, this assumption is correct only for small K and should break down for $K \approx J$. We find

$$\begin{aligned} g' &= \frac{\frac{2\sqrt{2}}{s}a_0}{\sqrt{1-\frac{2K}{J}}} = \frac{g}{\sqrt{1-\frac{2K}{J}}} \equiv u' a_0, \\ v_s' &= 2\sqrt{2}Ja_0 s\sqrt{1-\frac{2K}{J}} = v_s\sqrt{1-\frac{2K}{J}}. \end{aligned} \tag{6.1.1}$$

Thus, the main effects of frustrating interactions, in the neighborhood of a Néel ordered state, are the increase of the bare coupling g and the decrease of the spin-wave velocity v_s. It is also clear that for values of K sufficiently large, the bare dimensionless coupling constant u' will become larger than

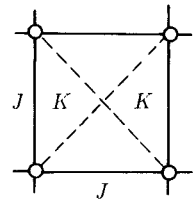


Figure 6.1 A square lattice with nearest neighbor (J) and next-nearest neighbor (K) interactions.

the critical value u^*. Consequently, there should be a critical value of the next-nearest-neighbor coupling strength K_c beyond which the long-range Néel order is destroyed. This theory would then predict that for $K \geq K_c$ the system becomes a quantum paramagnet.

It is also clear that if K gets to be large enough, a new form of long-range order should be found. Indeed, if $K \gg J$ a Néel-like state but with wave-vector $\vec{Q} = (\pi, 0)$ or $(0, \pi)$ is favored, instead of the usual $\vec{Q} = (\pi, \pi)$ ordered state (Fig. 6.2). This Néel-like state is antiferromagnetic along the x-axis but ferromagnetic along the y-axis. The low-energy effective action should then be a mixture of a sigma model which describes antiferromagnetism and a ferromagnetic Lagrangian of the form of Eq. (5.4.2). As a matter of fact, the Wess-Zumino terms of the individual spins do not cancel completely in this case. A term of the form

$$\gamma \int d^3x \vec{m} \cdot (\partial_0 \vec{m} \times \partial_1 \vec{m}) \tag{6.1.2}$$

is found, where γ is a parameter. However, this is *not* a topological (Hopf) term. It merely states that nearby chains exhibit the *same* antiferromagnetic order and the spins on one chain precess in the average field of its neighboring chains. This state should also become *unstable* for values lower than $K \approx J$. Thus, near the classically frustrated limit, $K = J$, new phases should appear.

There are several possibilities. One possible phase is a state without long-range magnetic order, with a gap for spin excitations and a unique ground

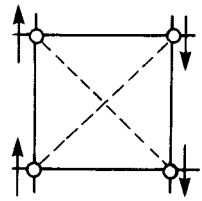


Figure 6.2 A $\vec{Q} = (\pi, 0)$ Néel state.

state. This is the usual paramagnetic state in the quantum zero-temperature limit (QP). We can think of other possible states by considering that when the spin-correlation length becomes very short (i.e. of the order of the lattice constant), the ground state is more naturally described in terms of pairs of spins forming $s = 0$ singlet states over fairly short distances. These states are dubbed *valence bond* states (VB). Various disordered states based on the VB picture have been proposed. They include VB crystals and Resonating Valence Bonds (RVB) states, in both its long [Anderson 87] and short range [Kivelson 87] varieties. Still, other proposals entertain the idea of ground states with *broken time reversal invariance*. Such is the case of the Kalmeyer-Laughlin (KL) state for the triangular lattice [Kalmeyer 88] , the chiral spin states for frustrated square lattices of Wen, Wilczeck and Zee (WWZ) [Wen 89] and unusual states with long range order such as the multisublattice Néel states including spirals of Shraiman and Siggia [Shraiman 89] and of Kane and all [Kane 89]. In this chapter we will deal with the disordered phases, whether chiral or not. Affleck and collaborators [AKLT 88] found a class of lattice models whose exact ground states are disordered.

6.2 Valence Bonds and Disordered Spin States

Imagine for the moment a microscopic spin system with interactions which are so strong that the Néel state is destroyed. If the local coupling between the spins is very strong, we should expect that a picture based on spin-waves, even massive ones, should not work very well. An alternative is to *pair up* the spins into singlet pairs or Valence Bonds [Anderson 73].

Our basic building block will be a singlet pairing (Valence Bond) of two spins at sites i and j of the lattice, not necessarily nearest neighbors. Let $|(ij)\rangle$ denote a Valence Bond pairing up sites i and j (Fig. 6.3). The state $|(ij)\rangle$ is the antisymmetric combination of up and down spins on sites i and j:

$$|(ij)\rangle = \frac{1}{\sqrt{2}}(|\uparrow_i\downarrow_j\rangle - |\downarrow_i\uparrow_j\rangle)\,. \tag{6.2.1}$$

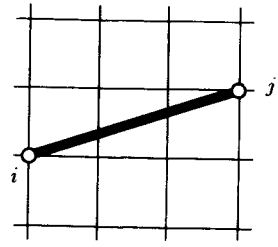


Figure 6.3 A Valence Bond $|(ij)\rangle$ on a 4×4 square lattice.

This is a spin-singlet state with respect to the total spin operators $\vec{S}^2$ and S_3

$$\begin{aligned} \vec{S}^2|(ij)\rangle &= 0 \\ S_3|(ij)\rangle &= 0 \end{aligned} \tag{6.2.2}$$

with

$$\vec{S} = \vec{S}_i + \vec{S}_j. \tag{6.2.3}$$

Next, we proceed to partition the set of sites of a lattice (with an even number of sites) into sets of all possible pairs of sites. If we assign a Valence Bond to each pair of a given partition, we can define a VB state for the partition as a tensor product of the Valence Bonds for each pair (Fig. 6.4)

$$|\mathrm{VB}\rangle = \prod_{\text{pairs}} |(i_k j_k)\rangle. \tag{6.2.4}$$

Since each Valence Bond is odd under the exchange of sites, the overall sign of the VB state is defined only up to a convention on how does one label the sites. I will assume that a *fixed* convention has been chosen. Since each pair is a spin singlet, the total spin of the system is necessarily equal to zero. However, zero total spin is not a good definition of a disordered spin state, as we will see below.

A priori we are tempted to consider an *arbitrary* spin singlet state as a linear superposition of VB states

$$|\Psi\rangle = \sum_P A(P) \prod_{\text{pairs}} |(i_k j_k)\rangle \tag{6.2.5}$$

which is a sum over all partitions $P = \{(i_k j_k)\}$ with amplitude $A(P)$. However, we run into a difficulty here. The VB states are not orthogonal and, what is more important, in general they cannot all be linearly independent at the same time. The set of VB states is, in general, an overcomplete set of states. Therefore, they are not good states for expanding a general wavefunction. On the other hand, if one is interested in just constructing a variational wavefunction, it may be convenient to write expressions of the type of Eq. (6.2.5) with variational parameters. One popular wavefunction has a *factorized* amplitude.

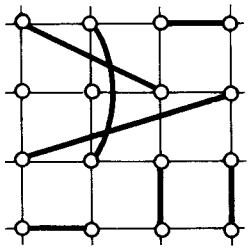


Figure 6.4 A VB state $|VB\rangle$ on a 4×4 square lattice is the product of 8 Valence Bonds $|(i,j)\rangle$.

In other words, $A(P)$ is written in the form

$$A(P) = \prod_{\text{pairs}} a(i_k, j_k) \tag{6.2.6}$$

and the total wavefunction looks like

$$|\Psi\rangle = \sum_P \prod_{\text{pairs}} a(i_k, j_k) |(i_k j_k)\rangle. \tag{6.2.7}$$

If we further assume that $a(i_k, j_k)$ is only a function of the distance between the paired sites i_k and j_k

$$a(i_k, j_k) = a(|i_k - j_k|) \tag{6.2.8}$$

we have a Resonating Valence Bond state (RVB) [Anderson 73]. This state has "resonances" in the sense that all Valence Bonds with sites at the same relative distance enter with the same amplitude. The optimal function $a(|\vec{x}|)$ can be determined by a variational calculation. The most extensive study of the Heisenberg model using states of this sort was carried out by Liang, Douçot and Anderson (LDA) [Liang 88].

The physical properties of a system depend on how fast does the function $a(|\vec{x}|)$ decay at infinity. For a power-law ansatz

$$a(|\vec{x}|) \sim \frac{\text{const}}{|\vec{x}|^\sigma} \qquad \text{for large} \quad |\vec{x}|, \tag{6.2.9}$$

LDA found that for $\sigma \geq 5$ there still is Néel long-range order, even though the wavefunction is a *global* spin singlet. Conversely, for $\sigma \leq 5$ LDA do not find Néel order beyond a scale ξ, the correlation length, which is finite.

An extreme case of an RVB state is the Short-Range RVB state which is defined as follows. Consider the VB states in which the paired sites are nearest neighbors to each other. There is a one-to-one correspondence between the underlying configurations of Valence Bonds and the configurations of classical *dimers* (Fig. 6.5) which occupy the bonds. The Short-Range RVB state, or Nearest-Neighbor RVB state (NNRVB), is simply the linear superposition of all such configurations with *equal* amplitude [Kivelson 87]. Thus, states which

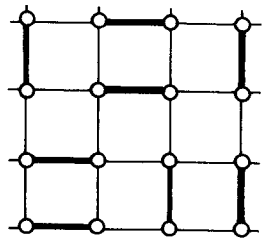


Figure 6.5 A Short-Range VB state on a 4×4 square lattice. The dark links ("dimers") are Valence Bonds or bonds.

differ by a local change in the dimer covering have exactly the same amplitude (resonance).

The NNRVB states have one important useful property: they are linearly independent. However, they are not orthogonal. To see this, consider two dimer coverings (a dimer covering is when every lattice site is connected to exactly one of its nearest-neighbors by a dimer) which differ only by a local rearrangement of a few nearby spins, such as the example of Fig. 6.6. I will pick the following convention for the signs of the VB states. I will only consider a bipartite lattice (say square). Since the lattice is bipartite, it can be partitioned into two interpenetrating sublattices called R (red) and B (black). A Valence Bond, or dimer, always joins a red site to a black site. The sign convention I pick, assigns a positive amplitude for every VB state provided that the red site appears (in the wavefunction) to the left of the black site. Equivalently, we can give an orientation to the Valence Bonds's: positive for red $\rightarrow$ black and negative for black $\rightarrow$ red [Kivelson 87]. We can picture this either by assigning an arrow to each VB or by coloring the sites, i.e. the endpoints of the bonds.

Once we have picked a sign convention, we can unambiguously compute overlaps. The overlap between the Short-Range VB states shown in Fig. 6.6 (a) and (b), call them $|a\rangle$ and $|b\rangle$, reduces to the overlap between the product of the two Valence Bonds which have been rearranged, since the other Valence Bonds have norm one by definition. Let 1 and 4 (2 and 3) belong to the red (black) sublattice. The overlap $\langle a|b\rangle$ is equal to

$$\langle a|b\rangle = \langle \vec{12}, \vec{43} | \vec{13}, \vec{42} \rangle \tag{6.2.10}$$

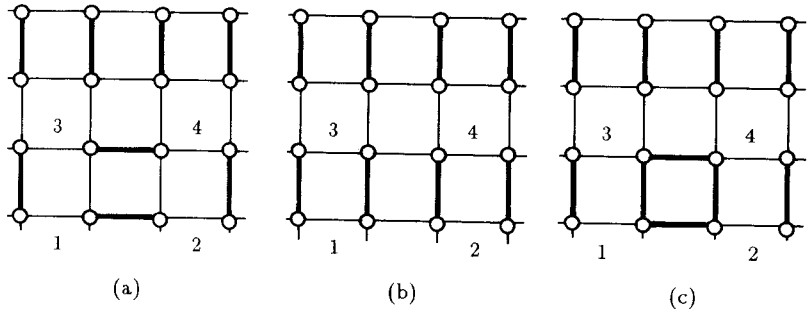


Figure 6.6 (a) and (b) are two configurations of dimers which differ only in the local arrangement of the dimers at the sites 1,2,3 and 4. (c) is the superposition of (a) and (b). The closed loop with non-vanishing area connects the sites 1,2,3 and 4 with 4 dimers, and represents the overlap of the non-orthogonal VB states $|a\rangle$ and $|b\rangle$ associated to the dimer covering in (a) and (b).

where $|1\vec{2}\rangle$, for instance, denotes the VB

$$|1\vec{2}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_1\downarrow_2\rangle - |\downarrow_1\uparrow_2\rangle). \tag{6.2.11}$$

Thus, $\langle a|b\rangle$ is simply given by

$$\begin{aligned}\langle a|b\rangle =& \frac{1}{4}(\langle\uparrow_1\downarrow_2\uparrow_4\downarrow_3 \,||\, \uparrow_1\downarrow_2\uparrow_4\downarrow_3\rangle + \langle\downarrow_1\uparrow_2\downarrow_4\uparrow_3 \,||\, \downarrow_1\uparrow_2\downarrow_4\uparrow_3\rangle)\\ =& \frac{1}{2}.\end{aligned} \tag{6.2.12}$$

More generally, overlaps between two arbitrary Short-Range VB states, say $|\Psi_a\rangle$ and $|\Psi_b\rangle$, will not be zero. These overlaps can be represented, and calculated, as a sum over all the closed loops on the square lattice obtained by superposing the dimer coverings associated to $|\Psi_a\rangle$ and $|\Psi_b\rangle$. The length of a loop Γ in units of the lattice spacing is $2L(\Gamma)$ where $L(\Gamma) = 1, 2, \ldots$. Its contribution to the overlap is equal to $2 \times 2^{-L(\Gamma)}$ (the factor $2^{-L(\Gamma)}$ comes from the choice of normalization, Eq. (6.2.1), while the factor 2 counts the number of ways to antiferromagnetically assign the spins on the sites of a loop) and therefore

$$\begin{aligned}\langle\Psi_a|\Psi_b\rangle =& \prod_{\Gamma} 2 \times 2^{-L(\Gamma)}\\ =& 2^{\sum_\Gamma} \times 2^{-\frac{1}{2}\sum_\Gamma 2L(\Gamma)}\\ =& 2^{P(a,b)} \times 2^{-\frac{N}{2}},\end{aligned} \tag{6.2.13}$$

where $P(a,b)$ ($P_{2L}(a,b)$) is the total number of loops (of loops of length $2L$) in the *loop covering* (a,b) and N the (even) number of sites. For example, the loop covering of the 4×4 square lattice shown in Fig. 6.6 (c) has seven loops: six of length two which, with our normalization, give factors of one and one of length four which gives a factor of one-half. Thus, the NNRVB state $|\Psi\rangle = \sum_a |\Psi_a\rangle$ has a wave-function normalization $\langle\Psi|\Psi\rangle$ which can be written as a sum of contributions from loops [Sutherland 88] of the form

$$\begin{aligned}\langle\Psi|\Psi\rangle =& \sum_{a,b}\langle\Psi_a|\Psi_b\rangle\\ =& 2^{-\frac{N}{2}}\sum_{a,b} 2^{P(a,b)} \times 2^{P(a,b)-P_2(a,b)}\\ \equiv& 2^{-\frac{N}{2}}\sum_{a,b} x^{P_2(a,b)} y^{P(a,b)-P_2(a,b)},\end{aligned} \tag{6.2.14}$$

with $x = 2$ and $y = 4$. Here, the factor $2^{P(a,b)-P_2(a,b)}$ accounts for the fact that there are two ways to have a loop of length $2L > 2$ with a given antiferromagnetic spin assignment on the sites of the loop.

Not only $\langle\Psi|\Psi\rangle$ can be written as a statistical sum such as Eq. (6.2.14), but the staggered spin-spin correlation function can also be written in a similar

form. Let $G(\vec{x})$ denote the staggered correlation function

$$G(\vec{x}) = 4(-1)^{x_1+x_2} \frac{\langle\Psi|\sigma_z(\vec{0})\sigma_z(\vec{x})|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{6.2.15}$$

For any loop covering (a, b), there are two possibilities [Kohmoto 88]: (i) The two points $\vec{0}$ and $\vec{x}$ are on the same loop, in which case due to the antiferromagnetic ordering on the loop, the contribution to the *staggered* correlation function is independent on their relative position. (ii) The two points belong to different loops and the loop covering does not contribute to the correlation function. In other words,

$$G(\vec{x}) = \frac{\sum_{a,b} \chi(\vec{x}) x^{P_2(a,b)} y^{P(a,b)-P_2(a,b)}}{\sum_{a,b} x^{P_2(a,b)} y^{P(a,b)-P_2(a,b)}}, \tag{6.2.16}$$

where

$$\chi(\vec{x}) = \begin{cases} 1 & \text{if } \vec{0} \text{ and } \vec{x} \text{ are on the same loop,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.17}$$

We can recast Eq. (6.2.16) in terms of sums over loops of non-vanishing area. If $L(a, b)$ is the total length of all loops with non-vanishing area for the loop covering (a, b), then $2P_2(a, b) + L(a, b) = N$. Now,

$$G(\vec{x}) = \frac{\sum \chi(\vec{x}) x^{\frac{-L}{2}} y^{P-P_2} d(P_2)}{\sum x^{\frac{-L}{2}} y^{P-P_2} d(P_2)} \tag{6.2.18}$$

where the summations are only for configurations of loops with non-vanishing areas and $d(P_2)$ is the number of configurations of loops of length 2. Thus the staggered correlation function gives us the probability for the two sites to belong to the same loop in a "gas" of loops. Since x and y are fairly small, the loop gas is reasonably dilute. A "quick-and-dirty" argument shows that the leading contribution to $G(\vec{x})$ should come from the smallest loop that contains both $\vec{0}$ and $\vec{x}$

$$\begin{aligned} G(\vec{x}) &= \frac{x^{\frac{1}{2}(N-2(\frac{|\vec{x}|}{a_0}+1))} y^1 + \cdots}{x^{\frac{N}{2}} y^0 + \ldots} \\ &\approx 2^{-(\frac{|\vec{x}|}{a_0}+1)} 4 \\ &\propto e^{-\frac{|\vec{x}|}{a_0} \ln 2}. \end{aligned} \tag{6.2.19}$$

Kohmoto and Shapiro [Kohmoto 88] have given a more refined argument which shows that $G(\vec{x})$ is bounded from above by an exponentially decreasing function with correlation length $\xi \approx a_0 e^{+1/\sqrt{2}}$. Thus, short-range RVB *wavefunctions* represent states with total spin equal to zero and exponentially decreasing correlation functions.

But are any of these RVB states, either short or long range, good approximations to the ground state wavefunction of a quantum Heisenberg model? The numerical evidence [Liang 88] indicates that for the unfrustrated model an RVB-like wave function with fairly long range is a good approximation to the ground state but it is a Néel state! The short-range RVB is not a good

approximation for this system. In the case of a *frustrated* system, such as the Heisenberg antiferromagnet on a *triangular* lattice, the situation is less clear. We will discuss this problem in chapter 7.

6.3 Spinons, Holons and Valence Bond States

We will now turn to other states which have been proposed. Since there is good evidence that the Heisenberg antiferromagnet may be in a Néel state, I will take the point of view that these phases may be realized by relatively small modifications of this Hamiltonian. Thus, I will carry out most of the discussion with the Heisenberg (or Hubbard) model in mind as a rather generic example.

At this point, it is convenient to go back to a representation of the spins either in terms of fermion operators or Bose operators. For the most part we have been using a fermion representation of the spins

$$\vec{S}(\vec{x}) = \frac{1}{2} c^\dagger_\alpha(\vec{x}) \vec{\sigma}_{\alpha\beta} c_\beta(\vec{x}). \tag{6.3.1}$$

The main motivation for this choice is that the fermion operators $c^\dagger_\alpha(\vec{x})$ are the fermion operators of the Hubbard model. Equation (6.3.1) reproduces the angular momentum algebra for spin $s = \frac{1}{2}$ only if the Hilbert space is restricted by the condition

$$n(\vec{x}) = c^\dagger_\alpha(\vec{x}) c_\alpha(\vec{x}) = 1 \tag{6.3.2}$$

which implies that each site is occupied by a *single* fermion with either up or down spin. Alternatively we may use bosons to represent spin. Let $a_\alpha(\vec{x})$ be a set of boson destruction operators. The boson bilinears

$$\vec{S}(\vec{x}) = \frac{1}{2} a^\dagger_\alpha(\vec{x}) \vec{\sigma}_{\alpha\beta} a_\beta(\vec{x}) \tag{6.3.3}$$

obey the angular momentum algebra for $s = \frac{1}{2}$ only if the bosons obey the hard core constraint

$$a^\dagger_\alpha(\vec{x}) a_\alpha(\vec{x}) = 1. \tag{6.3.4}$$

These formulas are reminiscent of the CP^1 representation of the non-linear sigma model of Section 5.9. Indeed, it is possible to derive the CP^1 model using bosons as a starting point. We will not do that here. This *boson* representation is closely related to standard spin-wave theory [Holstein 40].

Let us begin by looking for a representation of the Valence Bonds in terms of fermions. Let $|0\rangle$ represent the *empty* state. The Valence Bond on a pair of sites i and j is simply given by

$$|(ij)\rangle \equiv \epsilon_{\alpha\beta} c^\dagger_\alpha(i) c^\dagger_\beta(j) |0\rangle \equiv \left(c^\dagger_\uparrow(i) c^\dagger_\downarrow(j) - c^\dagger_\downarrow(i) c^\dagger_\uparrow(j) \right) |0\rangle. \tag{6.3.5}$$

We will be interested, for the moment, in the half-filled system. Thus the average number of particles per site is one and, because of the constraint, no

doubly occupied sites are allowed. For *finite* Hubbard U some doubly occupied sites, as well as empty sites, will occur. We may try to solve the constraint of no doubly occupied sites by using a "slave boson" construction [Coleman 84] [Read 83]. This leads to the RVB theories of Baskaran, Zou and Anderson (BZA), [Baskaran 87] and Ruckenstein, Hirschfeld and Appel [Ruckenstein 87]. In principle, there are several ways of implementing the slave boson approach. Let us consider the fermion operators normal ordered with respect to the half-filled state. In other words, we will assume that we are not too far from half filling. Let us now *define* a set of Bose and Fermi operators at each site, $b(\vec{x})$ and $f_\alpha(\vec{x})$ respectively, satisfying the constraint (at each site)

$$b^\dagger(\vec{x})b(\vec{x}) + f_\alpha^\dagger(\vec{x})f_\alpha(\vec{x}) = 1. \tag{6.3.6}$$

Let $|\bar{0}\rangle$ be the reference state for these operators and define the states $|h\rangle$, $|\uparrow\rangle$ and $|\downarrow\rangle$ representing a "hole" (or *holon*) with charge $+e$ and *spin zero* and a *spinon* $|\uparrow\rangle(|\downarrow\rangle)$ with *spin up (down)* and *no charge*:

$$\begin{aligned} |h\rangle &\equiv |e,0\rangle = b^\dagger|\bar{0}\rangle, \\ |\uparrow\rangle &\equiv |0,\uparrow\rangle = f_\uparrow^\dagger|\bar{0}\rangle, \\ |\downarrow\rangle &\equiv |0,\downarrow\rangle = f_\downarrow^\dagger|\bar{0}\rangle. \end{aligned} \tag{6.3.7}$$

Thus, the only possible states are a holon and a spinon of either orientation. More formally we can write the operator which creates a band fermion of charge e and spin σ at site $\vec{x}$, $c_\sigma^\dagger(\vec{x})$, in the form

$$c_\sigma^\dagger(\vec{x}) = b(\vec{x})f_\sigma^\dagger(\vec{x}). \tag{6.3.8}$$

Alternatively, we can also write

$$c_\sigma^\dagger(\vec{x}) = a(\vec{x})z_\sigma^\dagger(\vec{x}), \tag{6.3.9}$$

where a is a spinless charged fermion and the z_σ's are Schwinger bosons representation satisfying the constraint

$$z_\alpha^\dagger(\vec{x})z_\alpha(\vec{x}) = 1. \tag{6.3.10}$$

In this representation, the hole (or holon) is a fermion and the spinon is a boson. In either representation, at half-filling, there are no holons. Away from half-filling a number of holons will be present. In the bose-fermion version, the holons will superficially appear to undergo a condensation transition, which originally was mistakenly confused with "high-T_c".

6.4 Gauge Field Picture of the Disordered Spin States

I will consider now a particular form of Mean-Field theory for the Heisenberg antiferromagnet, first proposed by Affleck and Marston [Affleck 88] and by

Kotliar [Kotliar 88]. In this Mean-Field theory, one focuses on the valence bond operator of Eq. (6.3.5). The spin exchange term, i.e. $\vec{S}(\vec{x}) \cdot \vec{S}(\vec{y})$ can be written in the form

$$\vec{S}(\vec{x}) \cdot \vec{S}(\vec{y}) = \frac{1}{2} c_\alpha^\dagger(\vec{x}) c_\beta(\vec{x}) c_\beta^\dagger(\vec{y}) c_\alpha(\vec{y}) - \frac{1}{4} n(\vec{x}) n(\vec{y}). \tag{6.4.1}$$

Thus up to an additive constant, we have the fermion problem with the Hamiltonian

$$H = \frac{J}{2} \sum_{\vec{x}, j=1,2} c_\alpha^\dagger(\vec{x}) c_\beta(\vec{x}) c_\beta^\dagger(\vec{x} + \hat{e}_j) c_\alpha(\vec{x} + \hat{e}_j) \tag{6.4.2}$$

supplemented by the local constraint

$$n(\vec{x}) \equiv c_\alpha^\dagger(\vec{x}) c_\alpha(\vec{x}) = 1. \tag{6.4.3}$$

In Eq. (6.4.2), an underlying square lattice has been assumed and $j = 1, 2$ represents the x_1 and x_2 directions with $\hat{e}_1$ and $\hat{e}_2$ being the corresponding unit vectors. This approach can be generalized to other lattices as well.

The path-integral picture of this system involves the use of the Lagrangian

$$L = \sum_{\vec{x}} c_\alpha^\dagger(\vec{x}, t)(i\partial_t + \mu) c_\alpha(\vec{x}, t) + \sum_{\vec{x}} \varphi(\vec{x}, t)(c_\alpha^\dagger(\vec{x}, t) c_\alpha(\vec{x}, t) - 1) - H. \tag{6.4.4}$$

The second term in Eq. (6.4.4) contains the Lagrange multiplier field $\varphi(\vec{x}, t)$ which enforces the constraint of single occupancy, Eq. (6.4.3), at all times.

The Affleck-Marston Mean-Field Theory involves a Hubbard-Stratonovich factorization in terms of the link variables $\chi_j(\vec{x})$, which are complex Bose (c-number) fields. The Langrangian L'

$$\begin{aligned} L' = & \sum_{\vec{x}} c_\alpha^\dagger(x)(i\partial_t + \mu) c_\alpha(x) + \sum_x \varphi(x)(c_\alpha^\dagger(x) c_\alpha(x) - 1) + \\ & - \frac{2}{J} \sum_{\vec{x}, j} |\chi_j(x)|^2 + \\ & + \sum_{\vec{x}, j} \left(c_\alpha^\dagger(\vec{x}, t) \chi_j(\vec{x}, t) c_\alpha(\vec{x} + \hat{e}_j, t) + c_\alpha^\dagger(\vec{x} + \hat{e}_j, t) \chi_j^*(\vec{x}, t) c_\alpha(\vec{x}, t) \right) \end{aligned} \tag{6.4.5}$$

where $x \equiv (\vec{x}, t)$, is equivalent to L upon a Gaussian integration of the Hubbard-Stratonovich fields $\chi_j(x)$. Here, the link variables satisfy the relations $\chi_j(\vec{x}, t) = \chi_{-j}^*(\vec{x} + \hat{e}_j, t)$ since the current operator associated to an electron hoping from $\vec{x}$ to $\vec{x} + \hat{e}_j$ is Hermitean to that associated to the hoping from $\vec{x} + \hat{e}_j$ to $\vec{x}$.

The Mean-Field Theory (MFT) consists, as usual, in integrating out the fermions, *at a fixed density*, and treating the Bose (c-number) fields $\chi_j(x)$ within a saddle-point expansion. The fields $\chi_j(x)$, being complex, can be parametrized in terms of two real fields $\rho_j(x)$ and $\mathcal{A}_j(x)$ representing the amplitude and phase of $\chi_j(x)$ respectively. Before carrying out the MFT, it is important to consider the symmetries of this Lagrangian. Consider the *local,*

time-dependent, gauge transformations

$$\begin{aligned} \mathcal{A}_j(\vec{x},t) &= \mathcal{A}'_j(\vec{x},t) + \Delta_j\phi(\vec{x},t), \\ \varphi(\vec{x},t) &= \varphi'(\vec{x},t) + \partial_t\phi(\vec{x},t), \\ c_\alpha(x) &= e^{i\phi(x)}c'_\alpha(x). \end{aligned} \tag{6.4.6}$$

These transformations leave the Lagrangian unchanged up to the term $\sum_{\vec{x}} \partial_t\phi$. Thus, the Lagrange multiplier field, φ, transforms like the $\mathcal{A}_0$ component of a U(1) gauge field. We must conclude that this system has a "secret" gauge (local) symmetry. The effective Lagrangian Eq. (6.4.5) is reminiscent of Lagrangians of Lattice Gauge Theories [Kogut 82]. There are a few significant differences: (a) the amplitude field $|\chi_j(x)| = \rho_j(x)$ fluctuates; (b) there is no *explicit* kinetic energy term for the gauge fields $\mathcal{A}_\mu$ (i.e. an $F^2_{\mu\nu}$) and (c) there is an extra term in the Lagrangian proportional to φ, i.e. to $\mathcal{A}_0$. This last term may seem to break gauge invariance, since, according to Eq. (6.4.6), φ transforms like $\varphi \to \varphi' + \partial_t\phi$. However, we must keep in mind that what matters is not the Lagrangian but the *action*, $\mathcal{S}$

$$\mathcal{S} = \int dt L. \tag{6.4.7}$$

The extra term will transform the action by

$$\begin{aligned} \mathcal{S} \to & \mathcal{S} - \sum_{\vec{x}} \int dt \partial_t\phi(\vec{x},t) \\ = & \mathcal{S} - \sum_{\vec{x}} \left(\phi(\vec{x},t\to+\infty) - \phi(\vec{x},t\to-\infty)\right). \end{aligned} \tag{6.4.8}$$

If we impose periodic boundary conditions on the gauge fields, as we must when computing a *trace* over Bose fields, we must only allow for local gauge transformations which respect the boundary conditions. Thus, $\phi(\vec{x},t\to+\infty) = \phi(\vec{x},t\to-\infty)$ and the action is unchanged. We can relax this condition to a little extent. Let us notice that the "extra term" can be extracted from the action and written into the integrand in the form of a product of operators of the form

$$e^{-i\int dt \sum_{\vec{x}} \varphi(\vec{x},t)} \equiv \prod_{\vec{x}} e^{-i\int dt \varphi(\vec{x},t)}. \tag{6.4.9}$$

Since φ can be identified with $\mathcal{A}_0$, the time component of a vector potential $\mathcal{A}_\mu$ which obeys periodic boundary conditions, we can write the extra terms in the form of time-ordered exponentials of line integrals over loops $\Gamma(\vec{x})$ which close around the time direction (see Fig. 6.7). These operators are generally called *Wilson loops* :

$$e^{-i\sum_{\vec{x}}\int dt\varphi(\vec{x},t)} \equiv \prod_{\vec{x}} e^{-i\oint dt\mathcal{A}_0(\vec{x},t)} \equiv \prod_{\vec{x}} e^{-i\oint_{\Gamma(\vec{x})} dx_\mu \mathcal{A}^\mu}. \tag{6.4.10}$$

For the Wilson loops to be gauge invariant operators:

$$\oint_{\Gamma(\vec{x})} dx_\mu \mathcal{A}^\mu = \oint_{\Gamma(\vec{x})} dx_\mu \mathcal{A}'^\mu + \oint_{\Gamma(\vec{x})} dx_\mu \partial^\mu \phi = \oint_{\Gamma(\vec{x})} dx_\mu \mathcal{A}'^\mu, \qquad (6.4.11)$$

it is *sufficient* that $d\phi$ is exact, i.e. the gauge transformation is *non-singular everywhere*. Recall that these Wilson loops appeared in our problem since we had to enforce the *constraint* of single occupancy at *every* site and at *all* times.

Because of the gauge invariance, we only need to impose the constraint of single occupancy, Eq. (6.4.3), on the configuration space at some initial time surface, $t = t_0$. The local gauge invariance implies that the spin configurations at an arbitrary later time t must still obey the same constraint; i.e. they are smooth *deformations* of the initial configuration. For instance, we cannot try to fix the gauge $\mathcal{A}_0 = 0$ if only non-singular gauge transformations are allowed. This gauge is not consistent with the constraint of single occupancy since a configuration with $\mathcal{A}_0 = 0$ has $\oint \mathcal{A}_0 dt = 0$ and, because of gauge invariance, it cannot evolve into configurations with $\oint dt \mathcal{A}_0 \neq 0$. At best we can fix $\mathcal{A}_0(\vec{x}, t)$ to be a time-independent arbitrary function $\mathcal{A}_0(\vec{x})$ through

$$\oint dt \mathcal{A}_0(\vec{x}, t) \equiv T \mathcal{A}_0(\vec{x}) \equiv \bar{\mathcal{A}}_0(\vec{x}) \qquad (6.4.12)$$

where T is the time span. Alternatively, we may also choose the gauge

$$\mathcal{A}'_0(\vec{x}, t) \equiv \bar{\mathcal{A}}_0(\vec{x}) \delta(t - t_0) \qquad (6.4.13)$$

which yields the same value of the line integral. This choice means that, at $t = t_0$, we restrict the space of configurations to obey the constraint $n(\vec{x}) = 1$ at all points $\vec{x}$. Gauge-invariance then takes care of choosing only the time evolving configurations which satisfy this property.

However, it is worth noticing that from the point of view of quantum mechanics, what matters is not the invariance of the action $\mathcal{S}$ but the invariance of the amplitude $e^{i\mathcal{S}}$ assigned to a given history [Feynman 65]. Thus gauge transformations which change during the time span T by $\Delta\phi(\vec{x}) = 2\pi m(\vec{x})$ (an arbitrary integer modulo 2π at each point $\vec{x}$) *are allowed,* since they do not change the amplitude although they do change the action. These are the so-called *large* gauge transformations. These transformations change the

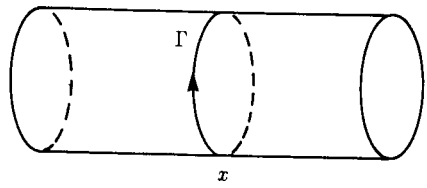


Figure 6.7 Wilson Loop along the closed curve $\Gamma(\vec{x})$ in the time direction.

time-like Wilson loops accordingly:

$$\oint dt\mathcal{A}_0 = \oint dt\mathcal{A}'_0 + 2\pi m(\vec{x}) \tag{6.4.14}$$

and thus, are *singular* gauge transformations. A correct description of these systems, particularly at non-zero temperatures, requires a careful treatment of these large gauge transformations.

We wish to evaluate the functional integral for a system with a Lagrangian of the form of Eq. (6.4.5). We will attempt a semi-classical treatment of this theory. One difficulty that we will encounter will be that there is no small parameter to organize this semi-classical expansion. Thus we should have every reason to suspect that the results may not be quite reliable. Indeed, using this approach it is quite hard to reproduce a Néel state. This is so because the approximations that we will make will be accurate for systems which can be described in terms of Valence Bonds. In this representation we deal with local spin singlets and the spins fluctuate very fast. Conversely, in a Néel state, the spins are *slow* variables but the VB's are *fast* ones. These are complementary descriptions.

An ingenious procedure has been devised in order to control the fluctuations. Affleck and Marston [Affleck 88] proposed to study a generalization of the Heisenberg model to a system with an SU(N) symmetry by attaching a "color" index $\alpha = 1, \ldots, N$ to each fermion. The spin one-half model was obtained by considering the $N = 2$ (SU(2)) case. The Affleck-Marston Lagrangian has, after an RVB decoupling by means of a link variable $\chi_j(\vec{x}, t)$, the same form as the Lagrangian of Eq. (6.4.5) except that (a) α ranges from 1 to N (not just 1 and 2, or $\uparrow$ and $\downarrow$) and (b) the local occupancy is not equal to one but to a function $n(\vec{x})$

$$\sum_{\alpha=1}^{N} c^{\dagger}_{\alpha}(\vec{x}) c_{\alpha}(\vec{x}) = n(\vec{x}) \tag{6.4.15}$$

which they proposed could take one of two forms on a system with two interpenetrating sublattices, A and B:

$$n(\vec{x}) = \begin{cases} 1 & \vec{x}\epsilon \text{ A}, \\ \\ N-1 & \vec{x}\epsilon \text{ B}, \end{cases} \tag{6.4.16}$$

or

$$n(\vec{x}) = \frac{N}{2} \qquad \vec{x}\epsilon \ A \text{ or } \vec{x}\epsilon \ B. \tag{6.4.17}$$

Read and Sachdev [Read 89] further generalized this model and considered an SU(N) "Heisenberg antiferromagnet" of the form

$$H = \frac{J}{N} \sum_{(\vec{x},\vec{x}')} \sum_{\alpha,\beta=1}^{N} \hat{S}^{\beta}_{\alpha}(\vec{x}) \hat{S}^{\alpha}_{\beta}(\vec{x}\,') \tag{6.4.18}$$

where $\vec{x}\epsilon A$ and $\vec{x}\,'\epsilon B$. The operators $\hat{S}^{\alpha}_{\beta}(\vec{x})$ are generators of the Lie group SU(N). If one chooses a representation of SU(N) determined by a Young

tableau with m rows and n_c columns ($0 < m < N$) on sublattice A and $N - m$ rows and n_c columns on sublattice B (which is the *conjugate* of the representation on sublattice A) (Fig. 6.8) we can write $\hat{S}^{\alpha}_{\beta}(\vec{x})$ in terms of fermions as follows

$$\hat{S}^{\beta}_{\alpha}(\vec{x}) = \sum_{a=1}^{n_c} c^{\dagger}_{\alpha a}(\vec{x}) c^{\beta a}(\vec{x}) - \delta^{\beta}_{\alpha}\frac{n_c}{2} \tag{6.4.19}$$

at the price of introducing an extra ("flavor") index $a = 1, \ldots, n_c$. The representation is fixed by the constraint [Read 89]

$$\sum_{\alpha=1}^{N} c^{\dagger}_{\alpha a}(\vec{x}) c^{\alpha b}(\vec{x}) = \begin{cases} \delta^{b}_{a} m & \vec{x} \epsilon \text{ A}, \\ \delta^{b}_{a}(N-m) & \vec{x} \epsilon \text{ B}. \end{cases} \tag{6.4.20}$$

Hence, there are mn_c fermions on sublattice A and $(N-m)n_c$ fermions on sublattice B. For example, for $N = 2$ (SU(2)), the only available value of m is 1 and n_c is arbitrary. It is easy to show that this representation has n_c spins one-half. The constraint means that the allowed states are *symmetric* under a permutation of the spins. This is the representation (or multiplet) with spin $s = \frac{n_c}{2}$. Thus, all the representations of SU(2) have been included. This is important since the limit $n_c \to \infty, N = 2$, is the spin-wave theory. The $\frac{1}{s}$ expansion discussed in Chapter 5 is simply the $\frac{1}{n_c}$ expansion here (since $s = \frac{n_c}{2}$).

From now on we will consider only the case of self-conjugate representations (i.e. the Young tableaus have the same number of rows $m = \frac{N}{2}$ for both sublattices). This is only possible for N even. We will only consider the fundamental representation, which has $n_c = 1$. The limit n_c large is more conveniently described in terms of bosons [Arovas 88] or in terms of a coherent-state path-integral [Read 89]. Both representations lead to a generalization of the non-linear sigma model of Chapter 5. We will not pursue this approach here.

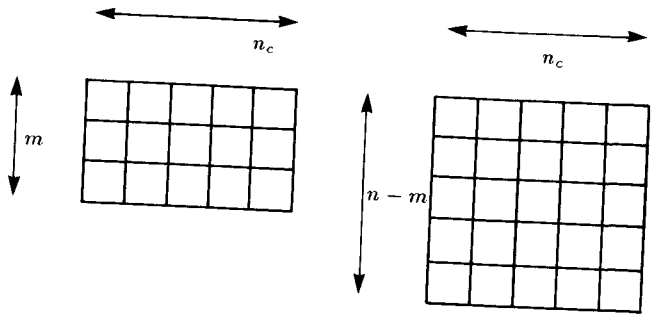


Figure 6.8 Conjugate Representations of SU(N) on a bipartite lattice. The case shown here is the $(m, n_c) = (3, 5)$ representation of SU(8) and its conjugate $(5, 5)$.

The Lagrangian density of Eq. (6.4.5) now has the form

$$\begin{aligned}\mathcal{L}' =& c^\dagger_{\alpha a}(\vec{x},t)(i\partial_t+\mu)c_{\alpha a}(\vec{x},t)+\varphi_{ab}(\vec{x},t)\left(c^\dagger_{\alpha a}(\vec{x},t)c_{\alpha b}(\vec{x},t)-\delta_{ab}\frac{N}{2}\right)+\\ &-\frac{N}{J}|\chi^{ab}_j(\vec{x},t)|^2+\\ &+c^\dagger_{\alpha a}(\vec{x},t)\chi^{ab}_j(\vec{x},t)c_{\alpha b}(\vec{x}+\hat{e}_j,t)+c^\dagger_{\alpha b}(\vec{x}+\hat{e}_j,t)\chi^{ab*}_j(\vec{x},t)c_{\alpha a}(\vec{x},t)\end{aligned} \tag{6.4.21}$$

where $\chi^{ab}_j(\vec{x},t)$ is an $n_c\times n_c$ complex matrix field satisfying

$$\chi^{ab}_j(\vec{x},t)=\chi^{ba}_{-j}{}^*(\vec{x}+\hat{e}_j,t). \tag{6.4.22}$$

The field $\chi^{ab}_j(\vec{x},t)$ is a generalization of $\chi_j(\vec{x},t)$ in Eq. (6.4.5). This Lagrangian density has a non-Abelian gauge invariance which is a generalization of Eq. (6.4.6). The functional integral is

$$Z=\int \mathcal{D}\chi\mathcal{D}\varphi\mathcal{D}c^\dagger\mathcal{D}c\; e^{iS}\prod_{\vec{x}} e^{-i\frac{N}{2}\oint dt\varphi_{aa}(\vec{x},t)}. \tag{6.4.23}$$

The action $\mathcal{S}$ is a *bilinear* form in fermions. Hence, once again, they can be integrated out at the expense of a determinant. The effective action $\mathcal{S}_{\text{eff}}$, resulting from integrating out the fermions, is

$$\mathcal{S}_{\text{eff}}[\varphi,\chi_j]=N\bar{\mathcal{S}}[\varphi,\chi_j] \tag{6.4.24}$$

where

$$\begin{aligned}\bar{\mathcal{S}}[\varphi,\chi_j]=-i\text{Tr}\;\ln[&(i(\partial_t+\mu)\delta_{ab}+\varphi_{ab}(\vec{x},t))\,\delta_{\vec{x},\vec{x}'}\delta_{t,t'}+\\ &+(\chi^{ab}_j(\vec{x},t)\delta_{\vec{x}',\vec{x}+\hat{e}_j}+\chi^{ba*}_j(\vec{x}-\hat{e}_j,t)\delta_{\vec{x}',\vec{x}-\hat{e}_j})\delta_{t,t'}]+\\ &-\int dt\sum_{\vec{x}}\frac{1}{J}|\chi^{ab}_j(\vec{x},t)|^2.\end{aligned} \tag{6.4.25}$$

We can also decompose $\chi^{ab}_j(\vec{x},t)$ into an amplitude and a phase

$$\chi^{ab}_j(\vec{x},t)=\rho^{ab}_j(\vec{x},t)e^{i\mathcal{A}^{ab}_j(\vec{x},t)} \tag{6.4.26}$$

where $\rho^{ab}_j(\vec{x},t)$ is a positive-definite real symmetric matrix and $\mathcal{A}^{ab}_j(\vec{x},t)$ is in the Lie algebra of SU(N) (i.e. $e^{i\mathcal{A}}$ is a group element). Clearly, $\varphi^{ab}(\vec{x},t)$ can be regarded as the time component $\mathcal{A}^{ab}_0(\vec{x},t)$ of the non-Abelian vector potential $\mathcal{A}^{ab}_\mu(\vec{x},t)$ while $\mathcal{A}^{ab}_j(\vec{x},t)$ are its space components. The saddle-point approximation is justified if we take the limit $N\to\infty$ keeping $n_c<\infty$. In the Bose representation, on the other hand, the limit one is forced to consider has $N<\infty$ and $n_c\to\infty$. Thus, although the theories should be equivalent, their saddle-point-approximations have quite different physics. The limit $n_c\to\infty$ means high representations and Néel-like behavior. The opposite limit, $N\to\infty$, n_c fixed, has VB states and flux phases but no Néel states.

6.5 Flux States and Valence Bond Crystals

For the most part I will consider only the case $n_c = 1$, which is simplest. However, there are some important new features which arise for $n_c > 1$ which I will mention in passing. For $n_c = 1$, the symmetry is Abelian. The saddle-point-approximation implies considering configurations of $\bar{\rho}_j(\vec{x}, t)$ and $\bar{\mathcal{A}}_\mu(\vec{x}, t)$ such that

$$\frac{\delta \mathcal{S}_{\text{tot}}}{\delta \bar{\rho}_j(\vec{x}, t)} = 0 \tag{6.5.1}$$

and

$$\frac{\delta \mathcal{S}_{\text{tot}}}{\delta \bar{\mathcal{A}}_\mu(\vec{x}, t)} = 0 \tag{6.5.2}$$

where $\mathcal{S}_{\text{tot}}$ is given from Eqs. (6.4.24) and (6.4.23) by

$$\mathcal{S}_{\text{tot}} = \mathcal{S}_{\text{eff}} - \frac{1}{2} \sum_{\vec{x}} \oint dt \mathcal{A}_0 = \mathcal{S}_{\text{eff}} - \sum_{\vec{x}} \oint dt J_\mu \mathcal{A}^\mu \tag{6.5.3}$$

with $J_\mu = \frac{1}{2}\delta_{\mu 0}$. Equation (6.5.1) determines the value (or configuration) of $\rho(\vec{x}, t)$ which extremizes the action. Similarly, Eq. (6.5.2) implies the absence of fermion currents j^F_μ in the ground state

$$\frac{\delta \mathcal{S}_{\text{tot}}}{\delta \bar{\mathcal{A}}_\mu(x)} = \frac{\delta \mathcal{S}_{\text{eff}}}{\delta \bar{\mathcal{A}}_\mu(x)} - J_\mu(x) \equiv j^F_\mu(x) - J_\mu(x) = 0. \tag{6.5.4}$$

In other words, the average fermion density is equal to one, as required by the constraint, and the average current vanishes. Two solutions have been proposed to solve the saddle-point equations: (i) flux state and (ii) VB crystal (or Peierls) state.

6.5.1 Flux States

We look for solutions of the saddle-point equations with maximal symmetry. For instance, we want solutions of Eq. (6.5.1) independent of $(\vec{x}, t)$ and of j:

$$\bar{\rho}_j(\vec{x}, t) = \bar{\rho}. \tag{6.5.5}$$

We may also ask for a possible solution with non-zero value of $\bar{\mathcal{A}}_j(\vec{x}, t)$ but with $\bar{\mathcal{A}}_0 = 0$. The value of $\bar{\mathcal{A}}_j(\vec{x}, t)$ may be chosen to be time-independent but not constant in space since, in that case, it would be gauge equivalent to zero. Thus we require that the *circulation* of $\bar{\mathcal{A}}_j(\vec{x}, t)$, or *flux* $\bar{\mathcal{B}}$, around *any* elementary plaquette be constant

$$\sum_{\text{plaquette}} \bar{\mathcal{A}}_j(\vec{x}, t) = \bar{\mathcal{B}}. \tag{6.5.6}$$

In general, a non-zero flux $\bar{\mathcal{B}}$ violates time reversal invariance since the time-reversal transformation maps $\bar{\mathcal{B}} \to -\bar{\mathcal{B}}$. But, this system is periodic in $\mathcal{A}_j$, i.e. $\mathcal{A}_j$ and $\mathcal{A}'_j = \mathcal{A}_j + 2\pi n_j$ (n_j an arbitrary integer) cannot be distinguished.

Thus $\bar{\mathcal{B}}$ is defined up to an integer multiple of 2π. There are two values of $\bar{\mathcal{B}}$ compatible with time reversal invariance: $\bar{\mathcal{B}} = 0, \pi$. Any other value of $\bar{\mathcal{B}}$ represents a state with broken time reversal symmetry. We will see below that these are the *chiral states* and are dubbed *generalized flux phases.*

6.5.2 Valence Bond States

There are two types of VB states: *chiral* and *non-chiral.* For both types of VB states, the field $\bar{\chi}_j(\vec{x}, t)$ has an amplitude $\bar{\rho}_j(\vec{x}, t)$ which takes non-zero values only on dimer configurations: $\bar{\rho}_j(\vec{x}, t) = \bar{\rho}$ on those links covered by dimers and zero elsewhere. The phases $\bar{\mathcal{A}}_j(\vec{x}, t)$ of $\bar{\chi}_j(\vec{x}, t)$ have circulations $\bar{\mathcal{B}}$ around elementary plaquettes which equal 0 or π for non-chiral states but take any other value in the chiral states.

Let us consider the saddle-point equations for $n_c = 1$ in more detail. We look for solutions which are time-independent and have $\bar{\mathcal{A}}_0 = 0$. Thus, $\bar{\rho}_j$ and $\bar{\mathcal{B}}$ are constant in time. From Eq. (6.4.21) we infer that the fermions, which are the *spinons* of this sytem, move with an effective Hamiltonian

$$H_{\rm MF} = -\sum_{\vec{x},j} \bar{\rho}_j(\vec{x}) \left(c^\dagger_\alpha(\vec{x}) e^{i\bar{\mathcal{A}}_j(\vec{x})} c_\alpha(\vec{x} + \hat{e}_j) + c^\dagger_\alpha(\vec{x} + \hat{e}_j) e^{-i\bar{\mathcal{A}}_j(\vec{x})} c_\alpha(\vec{x}) \right) + \\ + \frac{N}{J} \sum_{\vec{x},j} \bar{\rho}^2_j(\vec{x}), \tag{6.5.7}$$

in the background $\{\bar{\rho}_j(\vec{x}), \bar{\mathcal{B}}(\vec{x})\}$. Here, we have $\frac{1}{2}NL^2$ fermions in a system with the linear dimension L.

Let us consider first the uniform solutions which have $\bar{\rho}_j(\vec{x}) = \bar{\rho}$ (constant). We saw above that there are only two allowed values of $\bar{\mathcal{B}}$ consistent with time reversal invariance. For $\bar{\mathcal{B}} = 0$, the spinons have a square Fermi surface (see Eq. (2.4.10)). This is the state found by Baskaran, Zou and Anderson (BZA). The total energy of the BZA state is

$$E_{\rm BZA} = \frac{2NL^2}{J}\bar{\rho}^2 - \frac{8}{\pi^2} NL^2 \bar{\rho}. \tag{6.5.8}$$

The minimum is attained for $\bar{\rho} = \frac{2J}{\pi^2}$ and $E_{\rm BZA} = -\frac{8NL^2J}{\pi^4}$. Superficially, this state looks like a Fermi liquid of spinons. However, the fluctuations are likely to destroy this state. There are, naturally, amplitude fluctuations, $\tilde{\rho}_j(x) = \rho_j(x) - \bar{\rho}$. These fluctuations are essentially local in character and may trigger an instability towards a *Peierls* state in which $\bar{\rho}$ may have a periodic component in space. More importantly, the gauge fields are completely unconstrained. The result is a state in which the constraint of single occupancy is enforced and in which there is no current flow.

The state with $\bar{\rho}_j(\vec{x}) = \bar{\rho}$ constant, and $\bar{\mathcal{B}} = \pi$, everywhere, is usually called the flux phase. For the flux phase the spinons move according to the

Mean-Field Hamiltonian H_{flux} given from Eq. (6.5.7) by

$$H_{\text{flux}} = -\bar{\rho}\sum_{\vec{x},j}\left(c^{\dagger}_{\alpha}(\vec{x})e^{i\bar{\mathcal{A}}_j(\vec{x})}c_{\alpha}(\vec{x}+\hat{e}_j) + c^{\dagger}_{\alpha}(\vec{x}+\hat{e}_j)e^{-i\bar{\mathcal{A}}_j(\vec{x})}c_{\alpha}(\vec{x})\right) + \\ + \frac{2NL^2}{J}\bar{\rho}^2. \tag{6.5.9}$$

The vector potentials $\bar{\mathcal{A}}_j(\vec{x})$ should have circulation equal to π around every elementary plaquettes

$$\sum_{\text{plaquette}} \bar{\mathcal{A}}_j(\vec{x}) = \pi. \tag{6.5.10}$$

We can solve this requirement by the (gauge-dependent) *choice*

$$\bar{\mathcal{A}}_1(\vec{x}) = +\frac{\pi}{2}, \\ \bar{\mathcal{A}}_2(\vec{x}) = -\frac{\pi}{2}(-1)^{x_1}. \tag{6.5.11}$$

The Fermi fields $c_\alpha(\vec{x},t)$, the spinons, satisfy the equation of motions

$$i\dot{c}_\alpha(\vec{x},t) = [c_\alpha(\vec{x},t), H_{\text{flux}}] \\ = -\bar{\rho}\sum_{j=1,2}\left(e^{i\bar{\mathcal{A}}_j(\vec{x})}c_\alpha(\vec{x}+\hat{e}_j,t) + e^{-i\bar{\mathcal{A}}_j(\vec{x}-\hat{e}_j)}c_\alpha(\vec{x}-\hat{e}_j,t)\right). \tag{6.5.12}$$

It is convenient to introduce a separate amplitude for each of the four sublattices of Fig. 6.9:

$$\begin{aligned}
i\partial_0 f^{(1)}_\alpha(\vec{x}) &= -i\bar{\rho}\left[f^{(2)}_\alpha(\vec{x}+\hat{e}_1) - f^{(2)}_\alpha(\vec{x}-\hat{e}_1)\right] + \\
&\quad + i\bar{\rho}\left[f^{(3)}_\alpha(\vec{x}+\hat{e}_2) - f^{(3)}_\alpha(\vec{x}-\hat{e}_2)\right], \\
i\partial_0 f^{(2)}_\alpha(\vec{x}+\hat{e}_1) &= -i\bar{\rho}\left[f^{(1)}_\alpha(\vec{x}+2\hat{e}_1) - f^{(1)}_\alpha(\vec{x})\right] + \\
&\quad - i\bar{\rho}\left[f^{(4)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) - f^{(4)}_\alpha(\vec{x}+\hat{e}_1-\hat{e}_2)\right], \\
i\partial_0 f^{(3)}_\alpha(\vec{x}+\hat{e}_2) &= -i\bar{\rho}\left[f^{(4)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) - f^{(4)}_\alpha(\vec{x}-\hat{e}_1+\hat{e}_2)\right] + \\
&\quad + i\bar{\rho}\left[f^{(1)}_\alpha(\vec{x}+2\hat{e}_2) - f^{(1)}_\alpha(\vec{x})\right], \\
i\partial_0 f^{(4)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) &= -i\bar{\rho}\left[f^{(3)}_\alpha(\vec{x}+2\hat{e}_1+\hat{e}_2) - f^{(3)}_\alpha(\vec{x}+\hat{e}_2)\right] + \\
&\quad - i\bar{\rho}\left[f^{(2)}_\alpha(\vec{x}+\hat{e}_1+2\hat{e}_2) - f^{(2)}_\alpha(\vec{x}+\hat{e}_1)\right].
\end{aligned} \tag{6.5.13}$$

If we denote by $\Delta_j\phi(\vec{x},t)$ the finite symmetric difference

$$\Delta_j\phi(\vec{x},t) = \phi(\vec{x}+\hat{e}_j,t) - \phi(\vec{x}-\hat{e}_j,t), \tag{6.5.14}$$

we can write the equation of motion, Eq. (6.5.13), in the vector form ($a = 1,2,3,4$)

$$i\partial_t f^{(a)}_\alpha(\vec{x},t) = -i\bar{\rho}M^{ab}f^{(b)}_\alpha(\vec{x},t) \tag{6.5.15}$$

provided that $\vec{x}$ stands for an (even, even) site and the $f^{(1)}, f^{(2)}, f^{(3)}$ and $f^{(4)}$ components have the coordinates shown in Fig. 6.9. The matrix M^{ab} is given by

$$M^{ab} = \begin{pmatrix} 0 & \Delta_1 & -\Delta_2 & 0 \\ \Delta_1 & 0 & 0 & \Delta_2 \\ -\Delta_2 & 0 & 0 & \Delta_1 \\ 0 & \Delta_2 & \Delta_1 & 0 \end{pmatrix}. \tag{6.5.16}$$

Consider now the linear combinations

$$\begin{aligned} u_\alpha^{(1)}(\vec{x},t) &= f_\alpha^{(1)}(\vec{x},t) + f_\alpha^{(2)}(\vec{x}+\hat{e}_1,t), \\ u_\alpha^{(2)}(\vec{x},t) &= f_\alpha^{(3)}(\vec{x}+\hat{e}_2,t) - f_\alpha^{(4)}(\vec{x}+\hat{e}_1+\hat{e}_2,t), \end{aligned} \tag{6.5.17}$$

and

$$\begin{aligned} v_\alpha^{(1)}(\vec{x},t) &= f_\alpha^{(3)}(\vec{x}+\hat{e}_2,t) + f_\alpha^{(4)}(\vec{x}+\hat{e}_1+\hat{e}_2,t), \\ v_\alpha^{(2)}(\vec{x},t) &= f_\alpha^{(1)}(\vec{x},t) - f_\alpha^{(2)}(\vec{x}+\hat{e}_1,t). \end{aligned} \tag{6.5.18}$$

In terms of the spinons $u_\alpha^{(a)}$ and $v_\alpha^{(a)}$ $(a = 1, 2)$ we can write the equation of motion in the Dirac form

$$i\partial_0 u_\alpha^{(a)}(\vec{x},t) = -i\bar{\rho}(\sigma_3)_{ab}\Delta_1 u_\alpha^{(b)}(\vec{x},t) + i\bar{\rho}(\sigma_1)_{ab}\Delta_2 u_\alpha^{(b)}(\vec{x},t) \tag{6.5.19}$$

and the same equation for $v_\alpha^{(a)}(\vec{x},t)$. Let us define the 2×2 Dirac matrices γ_0, γ_1 and γ_2

$$\gamma_0 = -\sigma_2, \quad \gamma_1 = -i\sigma_1, \quad \gamma_2 = -i\sigma_3. \tag{6.5.20}$$

Equation (6.5.13), in this notation, has the form

$$\begin{aligned} i\left(\gamma_0\partial_0 - v_F\vec{\gamma}\cdot\vec{\nabla}\right)_{ab} u_\alpha^{(b)} &= 0, \\ i\left(\gamma_0\partial_0 - v_F\vec{\gamma}\cdot\vec{\nabla}\right)_{ab} v_\alpha^{(b)} &= 0, \end{aligned} \tag{6.5.21}$$

where I have taken the continuum limit and the Fermi velocity v_F is

$$v_F = 2a_0\bar{\rho}. \tag{6.5.22}$$

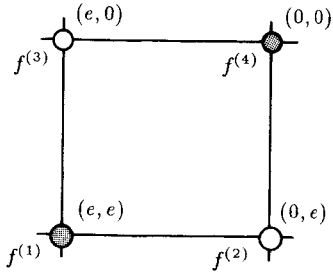


Figure 6.9 The four sublattices associated with a flux phase; (e,e) denotes a site with x_1 and x_2 even, and so on.

The eigenvalues are, in momentum space,

$$\epsilon(\vec{p}) = \pm 2\bar{\rho}\sqrt{\sin^2 p_1 + \sin^2 p_2} \tag{6.5.23}$$

with $|p_i| \leq \pi/2$. The dispersion curves form critical surfaces near $\vec{p} = 0$, characteristic of a continuum relativistic system (Fig. 6.10). The ground state energy in the flux phase is

$$\begin{aligned} E_{\text{flux}} =& \frac{2NL^2}{J}\bar{\rho}^2 - 2 \times 2NL^2\bar{\rho}\int_{|p_i|\leq\frac{\pi}{2}} \frac{d^2p}{(2\pi)^2}\sqrt{\sin^2 p_1 + \sin^2 p_2} \\ \equiv& \frac{2NL^2}{J}\bar{\rho}^2 - NL^2\alpha\bar{\rho} \end{aligned} \tag{6.5.24}$$

where the factor of 2 is due to the contribution of both u and v branches. The minimum is attained at $\bar{\rho} = \frac{1}{4}\alpha J$ and the total energy of the flux phase is

$$E_{\text{flux}} = -\frac{\alpha^2}{8}NL^2J \approx -0.115NL^2J \tag{6.5.25}$$

which is *lower* than that of the BZA state, $E_{\text{BZA}} = -\frac{8NL^2J}{\pi^4} \approx -0.082NL^2J$.

Both BZA and flux solutions have gapless excitations which carry the spin $\frac{1}{2}$ degree of freedom (for SU(2)) or, more generally, SU(N) color. While this spectrum appears to be stable at the level of Mean-Field theory, we will find problems once fluctuations are taken into account. First of all, we will find that a set of *dimer* states has lower energy than both the BZA and flux states. It is conceivable, however, that reasonable generalizations of this Hamiltonian do exist such that the flux state may be preferred. Affleck and Marson have indeed found such generalizations.

But what is more serious about these Mean-Field theories is the fact that they violate the *local* gauge invariance present in the full theory. In fact, we find spin non-singlet excitations which are not gauge invariant: the spinon

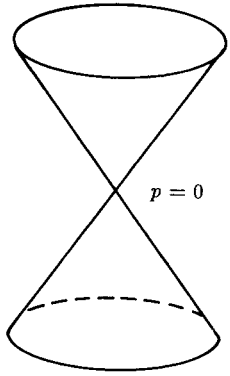


Figure 6.10 Dispersion law for spinons in a flux phase.

states. In Lattice Gauge Theories, there is a *theorem,* known as *Elitzur's theorem,* which states that in a theory with local interactions and with local gauge invariance, the only operators with a non-zero vacuum expectation value are locally gauge-invariant operators. In other words, the only states present in the spectrum are *local* gauge singlets. This result may appear to be puzzling at first glance. After all, even in theories with a global symmetry, such as the Ising model, the low-temperature magnetization is zero if the averages are computed over the entire configuration space. The procedure to remedy this problem is well known and it is crucial to a correct understanding of spontaneous breaking of *global* symmetries. First one considers a *finite* system of linear size L and the allowed space of configurations is reduced by either choosing a boundary condition (which fixes the asymptotic behavior of the spins at spatial infinity) or by turning on a weak external symmetry breaking field. Next, the thermodynamic limit $L \to \infty$ is considered in the presence of a fixed symmetry breaking procedure which is removed *after* the thermodynamic limit is taken. This procedure yields a non-zero magnetization because it takes an infinite order in the low-temperature expansion, i.e. the expansion around the state with broken symmetry, to mix the two degenerate classes of configurations. Hence, there is no mixing and the magnetization is non-zero if the expansion has a finite radius of convergence. However, if the symmetry is local, the situation is radically different. It always takes a finite order (of the order of the coordination number) in perturbation theory to mix states related through *local* gauge transformations. The behaviour of the system at the boundaries has little effect on the behavior near its center. The expressions for local expectation values are analytic functions of the coupling fields, *even* in the thermodynamic limit, $L \to \infty$. Thus, in the absence of external fields or gauge-fixing conditions, expectation values of locally gauge non-invariant operators must be zero. This is the content of Elitzur's theorem.

However, a gauge theory may be in a non-confined phase in which a gauge-invariant operator creates a *quark* (*spinon* in the terminology of magnetism) and antiquark (antispinon) at distances R, which can be separated all the way to infinity and still yield a non-zero amplitude. But for that to happen, the fluctuations of the gauge fields, or rather, of their field strengths, need to be controlled. This is *not* the case for the "RVB-type" Mean-Field theories. There is no term which controls the fluctuations of the gauge fields here. The gauge fields fluctuate so strongly that (a) they are able to enforce the local constraint and (b) they project out all current carrying states. The conclusion is the BZA and flux states need not only a Gutzwiller projection but also an additional procedure which eliminates all processes involving transport of spin over any significant distances.

The valence bond states, on the other hand, are manifestly local singlets. Thus gauge field fluctuations will play a rather small role in this case. We should expect states based on a VB description to be more stable. The problem of finding a "true" spin-liquid state, i.e. a state without broken symmetries *and* with spinon states in its spectrum, remains essentially an open issue.

Let us now turn our attention to a different set of solutions of the saddle-point equations which is based on valence bond states. Consider a configuration of $\bar{\rho}_j(\vec{x})$ which equals $\bar{\rho}$ on a set of links occupied by dimers such as in Fig. 6.5.

$$\bar{\rho}_j(\vec{x}) = \begin{cases} \bar{\rho} & \text{if the link } (\vec{x}, \vec{x}+\hat{e}_j) \text{ is occupied by a dimer,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.5.26}$$

The Mean-Field Hamiltonian, Eq. (6.5.7), with Eq. (6.5.26) describes a set of spinons *confined* inside the links, the VB states. Thus, we do not have spinon states propagating beyond the size of a dimer (one lattice spacing) in this dimer limit. Fluctuations will enable the effective size of a VB state to grow from the lattice spacing scale up to some finite ξ. This scale *is* the spin correlation length for this system in this phase. It is also clear that, at the level of Mean-Field theory, the average flux is not determined. This is simply reflecting the fact that the fluctuations of the gauge field are so strong that the average flux is wiped out. We will see later that, if holes are taken into account, a flux phase may develop. The energy of a VB state is

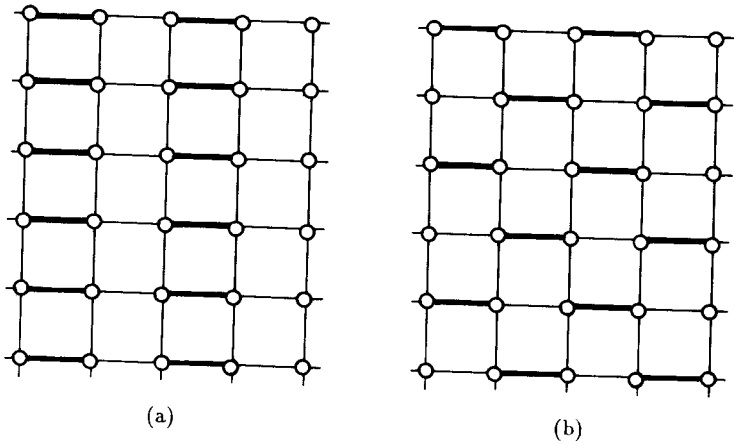
$$E_{\text{VB}} = \frac{2NL^2}{J}\bar{\rho}^2 - NL^2\bar{\rho} \tag{6.5.27}$$

which is minimized by the choice

$$\bar{\rho} = \frac{J}{4} \tag{6.5.28}$$

and has the ground state energy

$$E_{\text{dimer}} = -\frac{J}{8}NL^2 \tag{6.5.29}$$



(a) (b)

Figure 6.11 (a) One of the four column or Peierls states. (b) shows a staggered configuration.

for *any* dimer configuration. These states clearly have less energy ($-0.125JN$ per sites) than both BZA and flux states. However, we do not get a unique ground state at $N = \infty$. This degeneracy is lifted by fluctuations in the amplitude which appear at order $\frac{1}{N}$. Dombre and Kotliar [Dombre 89] as well as Read and Sachdev [Read 89] found that, for the case, $n_c = 1$, the *four* column or Peierls states are chosen (Fig. 6.11).

6.6 Fluctuatations of Valence Bonds: Quantum Dimer Models

The valence bond crystal of the past section has a spin-correlation length of the order of the lattice constant. It represents a quantum paramagnet. It is *not* a translationally invariant state, unlike the equal-amplitude short-range RVB state. It has crystalline order of its valence bonds and it is four-fold degenerate.

Alternatively we can imagine that the amplitude fluctuations, which represent transitions to states with broken bonds, are suppressed. The only way the system has to minimize its energy is to find a coherent rearrangement of bonds. If the amplitude fluctuations are frozen out, the system has states labeled by quantum numbers which describe the covering of the lattice by dimers. For the rest of our discussion we will ignore the SU(N) structure. Let $l_j(\vec{x})$ be an integer-valued variable associated with the bond $(\vec{x}, \vec{x} + \hat{e}_j)$. The Hilbert space is the space of states of the form $\{|\{l_j(\vec{x})\}\rangle\}$ where the integer l_j is either equal to zero (no dimer) or one (dimer). Every site has to belong to one and only one dimer. This requirement leads to the local constraint

$$l_1(\vec{x}) + l_2(\vec{x}) + l_1(\vec{x} - \hat{e}_1) + l_2(\vec{x} - \hat{e}_2) = 1. \qquad (6.6.1)$$

The "resonance" process of Fig. 6.12 is represented by an off-diagonal matrix element in which the integer degrees of freedom l_j for parallel bonds of a plaquette are raised from zero to one if the other two bonds are lowered from one to zero. This process can be described by a term in the effective

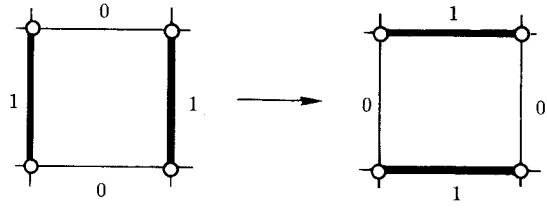


Figure 6.12 Resonance process. The integers $l = 0, 1$ represent the bond occupation by a dimer.

Hamiltonian of the form H_{reson}

$$H_{\text{reson}} = \bar{J} \sum_{\text{plaquette}} (|\,||\rangle\langle =|\,| + \text{h.c.}) \tag{6.6.2}$$

where $\bar{J}$ is an effective coupling constant $\bar{J} \propto J$. In addition there are diagonal matrix elements which give an energy V to a pair of neighboring parallel dimers, H_{diag}

$$H_{\text{diag}} = V\,(|\,=\rangle\langle =|\,| + |\,||\rangle\langle ||\,|)\,. \tag{6.6.3}$$

The full Hamiltonian of the Quantum Dimer Model (QDM) is

$$H_{\text{QDM}} = H_{\text{reson}} + H_{\text{diag}} \tag{6.6.4}$$

which is to be supplemented by the local constraint of Eq. (6.6.1).

We will now describe a solution of this problem. The result will be a VB crystal. Thus, even the QDM, originally proposed by Kivelson and Rokhsar [Rokhsar 88] as a model with a short-range RVB state as its ground state, has, in general, a crystalline ground state. For a particular value of $\frac{\bar{J}}{V}$ ($\frac{\bar{J}}{V} = -1, V > 0$) Kivelson and Rokhsar found that the short-range RVB wavefunction is the ground state wavefunction and it has zero energy. For $V > |\bar{J}|$, the staggered valence bond crystal of Fig. 6.13 is the exact ground state and it has zero energy. We saw before that the correlation length of the spins is short ranged [Kohmoto 88]. This does not imply that all other correlation functions must also be short ranged. For instance, the dimer-density correlation function is not short-ranged. This correlation functions, which measures the *amplitude* of finding two parallel dimers at some separation R in this state is equal to the probability of finding two parallel dimers in a random distribution of classical dimers covering the lattice. Fisher and Stephenson [Fisher 63] found that this correlation function decays like $\sim \frac{1}{R^2}$. Thus, the short-range RVB state is a dimer liquid while the VB crystals are dimer solids.

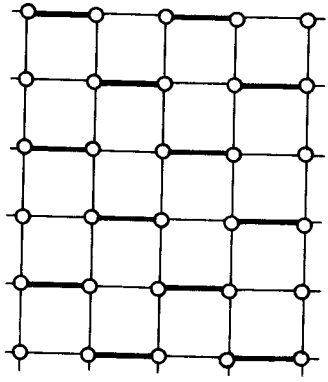


Figure 6.13 The staggered VB crystal.

6.7 Quantum Dimers and Gauge Theories

We wish to consider the full quantum dynamics of the QDM. We will find it most profitable to map this problem into a Lattice Gauge Theory [Kogut 79]. In a sense this mapping is suggested by the RVB-Mean-Field decoupling that we have been using all along. Baskaran and Anderson [Baskaran 88] first introduced a mapping of the static interactions of the RVB-Mean-Field theory, to a gauge theory. Here I am following the work by Kivelson and myself [Fradkin 88], [Fradkin 90].

Let us begin by defining an enlarged Hilbert space on the links of the lattice. Let $\{l_j(\vec{x})\}$ be a set of integer valued variables defined on the links $\{(\vec{x}, \vec{x}+\hat{e}_j)\}$ of the lattice. The states $|\{l_j(\vec{x})\}\rangle$ span the unrestricted Hilbert space. The angular-momentum operators $\hat{L}_j(\vec{x})$ have the integers $l_j(\vec{x})$ as their eigenvalues and $|\{l_j(\vec{x})\}\rangle$ as their eigenstates. If we wish to restrict this Hilbert space to the subspace in which $l_j = 0, 1$, we can do so by assigning an infinite energy to all unwanted states. Thus, let us define a dimer contribution, or kinetic energy term, which enforces the restriction and is nothing but a hard-core condition. We can write H_{dimer} in the form

$$H_{\text{dimer}} = \frac{1}{2k} \sum_{\vec{x},j} \left((\hat{L}_j(\vec{x}) - \frac{1}{2})^2 - \frac{1}{4} \right). \tag{6.7.1}$$

For any value of the coupling constant k, the configurations with $l_j = 0, 1$ have exactly zero energy while *any* other state will have energies growing like $\frac{1}{k}$ as $k \to 0$.

We need two terms: one for resonance and the other for the diagonal terms. In order to discuss resonance we need to introduce the variable $a_j(\vec{x})$ at each link which should be the eigenvalue of the operator $\hat{a}_j(\vec{x})$ canonically conjugate to $\hat{L}_j(\vec{x})$, i.e.

$$[\hat{a}_j(\vec{x}), \hat{L}_{j'}(\vec{x}\,')] = i\delta_{jj'}\delta_{\vec{x},\vec{x}'}. \tag{6.7.2}$$

Since the spectrum of $\hat{L}_j(\vec{x})$ is the integers $l_j(\vec{x}), a_j(\vec{x})$ should be an angle

$$0 \leq a_j(\vec{x}) < 2\pi, \tag{6.7.3}$$

and the Hilbert space is the space of the periodic functions of $a_j(\vec{x})$ with period 2π, independently at each link. Using the commutation relations Eq. (6.7.2), we see that the operator $e^{im_j\hat{a}_j}$ acts like a *ladder* operator with stepsize m_j, where m_j is an integer. Indeed, we can write for any site

$$\hat{L}_j e^{im_j\hat{a}_j}|l_j\rangle = e^{im_j\hat{a}_j}\left(e^{-im_j\hat{a}_j}\hat{L}_j e^{im_j\hat{a}_j}\right)|l_j\rangle. \tag{6.7.4}$$

The commutation relations tell us that the operator between brackets in Eq. (6.7.4) is the shifted operator

$$e^{-im_j\hat{a}_j}\hat{L}_j e^{im_j\hat{a}_j} = \hat{L}_j + m_j. \tag{6.7.5}$$

Thus, we get

$$\hat{L}_j e^{im_j\hat{a}_j}|l_j\rangle = e^{im_j\hat{a}_j}(l_j+m_j)|l_j\rangle = (l_j+m_j)e^{im_j\hat{a}_j}|l_j\rangle \tag{6.7.6}$$

and we can identify

$$e^{im_j\hat{a}_j}|l_j\rangle = |l_j+m_j\rangle. \tag{6.7.7}$$

The resonance term should remove from a plaquette two parallel dimers and replace them by another pair of parallel dimers but in an orthogonal direction (Fig. 6.11). We can accomplish this by writing, in terms of raising and lowering operators, the term

$$\begin{aligned} H_{\text{reson}} = \bar{J}\sum_{\vec{x}}\big(&e^{i[\hat{a}_1(\vec{x})+\hat{a}_1(\vec{x}+\hat{e}_2)-\hat{a}_2(\vec{x})-\hat{a}_2(\vec{x}+\hat{e}_1)]}+ \\ &+e^{i[\hat{a}_2(\vec{x})+\hat{a}_2(\vec{x}+\hat{e}_1)-\hat{a}_1(\vec{x})-\hat{a}_1(\vec{x}+\hat{e}_2)]}\big). \end{aligned} \tag{6.7.8}$$

The diagonal terms are

$$H_{\text{diag}} = V\sum_{\vec{x}}\left(\hat{L}_1(\vec{x})\hat{L}_1(\vec{x}+\hat{e}_2)+\hat{L}_2(\vec{x})\hat{L}_2(\vec{x}+\hat{e}_1)\right) \tag{6.7.9}$$

and the constraint is

$$\hat{Q}(\vec{x}) = \hat{L}_1(\vec{x})+\hat{L}_1(\vec{x}-\hat{e}_1)+\hat{L}_2(\vec{x})+\hat{L}_2(\vec{x}-\hat{e}_2) = 1. \tag{6.7.10}$$

This last equation looks peculiar since the left-hand side is an operator and the right-hand side is a number. The meaning of this equation is that the allowed states of the Hilbert states, which I will call $|\text{Phys}\rangle$, satisfy

$$\hat{Q}(\vec{x})|\text{Phys}\rangle = |\text{Phys}\rangle. \tag{6.7.11}$$

For this condition to be consistent, $\hat{Q}(\vec{x})$ should be diagonalizable simultaneously with the total Hamiltonian H, i.e.

$$[\hat{Q}(\vec{x}),\hat{H}] = 0 \tag{6.7.12}$$

where

$$H = H_{\text{dimer}}+H_{\text{reson}}+H_{\text{diag}}. \tag{6.7.13}$$

This is indeed the case, since $\hat{Q}(\vec{x})$ simply counts all the dimers touching a given site and this number is a constant of motion. The operator $\hat{Q}(\vec{x})$ generates a set of local time-independent transformations of the form

$$e^{i\sum_x \alpha(\vec{x})\hat{Q}(\vec{x})}|\text{Phys}\rangle \tag{6.7.14}$$

which leave $\hat{H}$ unchanged. Thus we discover that $\hat{H}$ has a local gauge symmetry and $\hat{Q}$ is the generator of local gauge transformations. The constraint equation is simply Gauss's law. This symmetry is simply the fact that we are free to change the phases of the Valence Bonds at each site independently. In this language, the wavefunctions which are being considered have the form

$$|\Psi\rangle = \sum_{\{c\}} A(c)e^{i\Phi(c)}|c\rangle \tag{6.7.15}$$

where $\{c\}$ is a set of (linearly independent) VB states (i.e. dimer coverings). $A(c)$ is a real amplitude and $\Phi(c)$ is the phase. The phase $\Phi(c)$ depends on the configuration and we have chosen to write $\Phi(c)$ in the form of a sum over links

$$\Phi(c) = \sum_{\vec{x},j} a_j(\vec{x}). \tag{6.7.16}$$

States of the form of Eq. (6.7.15) are coherent states parametrized by the variables $a_j(\vec{x})$.

We can write these formulas in a much more transparent and familiar way, by staggering the configuration $\{a_j(\vec{x})\}$. Clearly this can only be done consistently for a bipartite lattice. Define the staggered gauge field $\hat{A}_j(\vec{x})$ and "electric fields" $\hat{E}_j(\vec{x})$ by

$$\begin{aligned} \hat{A}_j(\vec{x}) &= e^{i\vec{Q}_0\cdot\vec{x}}\hat{a}_j(\vec{x}) \\ \hat{E}_j(\vec{x}) &= e^{i\vec{Q}_0\cdot\vec{x}}\hat{L}_j(\vec{x}) \end{aligned} \tag{6.7.17}$$

with $\vec{Q}_0 = (\pi, \pi)$. It should be stressed that these fields do *not* represent the electromagnetic fields. With these definitions, we can write the constraints of Eq. (6.7.11) in the form

$$[\Delta_j \hat{E}_j(\vec{x}) - \rho(\vec{x})]|\text{Phys}\rangle = 0 \tag{6.7.18}$$

where Δ_j is the lattice divergence

$$\Delta_j \hat{E}_j(\vec{x}) \equiv \hat{E}_1(\vec{x}) - \hat{E}_1(\vec{x} - \hat{e}_1) + \hat{E}_2(\vec{x}) - \hat{E}_2(\vec{x} - \hat{e}_2) \tag{6.7.19}$$

and the density $\rho(\vec{x})$ is

$$\rho(\vec{x}) = e^{i\vec{Q}_0\cdot\vec{x}}. \tag{6.7.20}$$

Equation (6.7.18) now has the standard form of Gauss's law. Note that $\rho(\vec{x})$ represents a background staggered charge density which equals $+1(-1)$ on red (black) sites and is enforcing the condition that each site should belong to one and only one dimer. In the presence of holes, $\rho(\vec{x})$ will vanish on sites occupied by holes. In this formulation the Hamiltonian reads

$$\begin{aligned} H =& \frac{1}{2k}\sum_{\vec{x},j}\left([\hat{E}_j(\vec{x}) - \alpha_j(\vec{x})]^2 - \alpha_j^2(\vec{x})\right) + 2\bar{J}\sum_{\vec{x}}\cos\left(\sum_{\text{plaquette}}\hat{A}_j(\vec{x})\right) + \\ &- V\sum_{\vec{x}}\left(\hat{E}_1(\vec{x})\hat{E}_1(\vec{x}+\hat{e}_2) + \hat{E}_2(\vec{x})\hat{E}_2(\vec{x}+\hat{e}_2)\right) \end{aligned} \tag{6.7.21}$$

where $\sum_{\text{plaquette}}\hat{A}_j(\vec{x})$ stands for the oriented sum of staggered vector potentials $\hat{A}_j(\vec{x})$ around the elementary plaquette labeled by $\vec{x}$ (its southwest

corner):

$$\begin{aligned}\sum_{\text{plaquette}} \hat{A}_j(\vec{x}) &\equiv e^{i\vec{Q}_0\cdot\vec{x}}\left(\hat{a}_1(\vec{x}) + \hat{a}_1(\vec{x}+\hat{e}_2) - \hat{a}_2(\vec{x}) - \hat{a}_2(\vec{x}+\hat{e}_1)\right) \\ &= \hat{A}_1(\vec{x}) - \hat{A}_1(\vec{x}+\hat{e}_2) - \hat{A}_2(\vec{x}) + \hat{A}_2(\vec{x}+\hat{e}_1) \\ &= \Delta_2\hat{A}_1(\vec{x}) - \Delta_1\hat{A}_2(\vec{x})\end{aligned} \tag{6.7.22}$$

and $\alpha_j(\vec{x})$ is

$$\alpha_j(\vec{x}) = \frac{1}{2}e^{i\vec{Q}_0\cdot\vec{x}}. \tag{6.7.23}$$

By expanding the square in the first term in Eq. (6.7.21), and using Eq. (6.7.23), we can write the first term in the form

$$\frac{1}{2k}\left(\sum_{\vec{x},j}\hat{E}_j^2(\vec{x}) - \frac{L^2}{2}\right) \tag{6.7.24}$$

where L is the linear size of the square lattice.

We can take all these considerations into account by writing the Hamiltonian in the form

$$\begin{aligned}H =& \frac{1}{2k}\left(\sum_{\vec{x},j}\hat{E}_j^2(\vec{x}) - \frac{L^2}{2}\right) + 2\bar{J}\sum_{\vec{x}}\cos\left(\sum_{\text{plaquette}}\hat{A}_j(\vec{x})\right) \\ &+ \frac{V}{2}\sum_{\vec{x}}\left((\Delta_1\hat{E}_2(\vec{x}))^2 + (\Delta_2\hat{E}_1(\vec{x}))^2\right) - \frac{V}{2}L^2\end{aligned} \tag{6.7.25}$$

and considering the limit $k \to 0$. The states are restricted by demanding that Gauss's law (Eq. (6.7.18)) be satisfied.

This Hamiltonian is closely related to a problem solved by Polyakov in 1975 [Polyakov 77]: compact quantum electrodynamics in (2+1) dimensions. It is compact in the sense that its degrees of freedom, the gauge fields $\hat{A}_j$, or rather the exponentials $e^{i\hat{A}_j(\vec{x})}$, are elements of the compact Lie group U(1). It differs from Polyakov's compact QED in that (a) $\bar{J}$ has the wrong sign and (b) the constraint selects a space of states which is not the usual vacuum ($\rho = 0$) but which has an array of sources, $\rho(\vec{x}) = \pm 1$. The first problem can be solved very easily (in the absence of holes) by shifting the gauge variables $\hat{A}_j = \hat{A}'_j + \delta\hat{A}_j$ in such a way that $\sum_{\text{plaquette}}\delta\hat{A}_j = \pi$. For instance, we can shift $\hat{A}_1$ by π on every other horizontal row. The second caveat, (b), is intrinsic and cannot be done away with by any redefinitions of variables. The shift $\delta\hat{A}_j$ says that Eq. (6.7.25) represents a system which likes to have flux π per plaquette, on average. This result is reminiscent of the flux phase. Thus, in terms of shifted variables, H has exactly the same form but with $\bar{J} \leftrightarrow -\bar{J}$.

The apparent analogy with QED may lead us to think that the ground state of this system (after shifting) has weakly fluctuating gauge fields. In such a case one may expect that the elementary excitation should have $\hat{A}_j$ to be small, slowly varying and gapless: a "photon". Recall that we are working with staggered variables. Hence this "photon" has wavevectors close to $\vec{Q}_0 = (\pi, \pi)$.

This is the *resonon* of Kivelson and Rokhsar who argued about its existence for $-\bar{J} = V$.

Away from $|\bar{J}| = V$, the resonon does not exist! This is so since, as Polyakov showed, compact QED is a *confining* theory. His results, which he derived for the case $\rho(\vec{x}) = 0$ (i.e. the usual vacuum sector), imply that (i) the ground state is unique and it is a gauge singlet, (ii) the spectrum has a gap, and (iii) only gauge invariant states are present (in particular, there is no "photon"). We will see now, by following Polyakov's ideas and using the methods of Banks, Kogut and Myerson [Banks 77], and Fradkin and Susskind [Fradkin 78], how these results are modified by the presence of a non-zero $\rho(\vec{x})$.

Since we expect, after Polyakov, that the physics of the ground state and low-lying excitations may not be accessible by means of a perturbative expansion around a state with some background classical field $\hat{A}_j$, it is useful to identify the topological excitations of this system. If we consider the Euclidean evolution of the system (i.e. imaginary time), the field configurations which disorder the long-range properties of the classical background state look like Dirac magnetic monopoles with integer charge. Polyakov's observation was, and this will also be crucial to our problem, that fluctuations around a background configuration with monopoles induce an interaction among them which is identical to that of a (neutral) Coulomb gas in three Euclidean dimensions. Since the Coulomb gas has the property of screening for all values of the coupling constant, the ground state is unique and has gap $\Delta \sim \frac{1}{\xi_s}$, where ξ_s is the screening length of the monopole-antimonopole plasma. Let us rederive these results and, at the same time, keep track of the sources $\rho(\vec{x})$.

6.8 Duality transformation and Ground State Degeneracy

The first step is a *dual transformation*. We will define this transformation in terms of the evolution of the constraint equation (Eq. (6.7.18)). Let $\vec{r}$ label the sites of the dual lattice, which is also a square lattice (Fig. 6.14). Let $\hat{S}(\vec{r})$ be an operator defined on *sites* of the dual lattice with a spectrum labelled by the integers $S(\vec{r})$. Similarly, $B_j(\vec{r})$ is a *classical* background real valued field which resides on the *links* of the dual lattice. I require that

$$\hat{E}_j(\vec{x}) = \epsilon_{jk} \left(\Delta_k \hat{S}(\vec{r}) + B_k(\vec{r}) \right), \tag{6.8.1}$$

where ϵ_{jk} is the Levi-Civita tensor and $i, j = 1, 2$. If we now substitute Eq. (6.8.1) into the constraint Eq. (6.7.18), then, in the subspace of physical states,

we get

$$\begin{aligned}\Delta_j \hat{E}_j(\vec{x}) &= \epsilon_{jk}\left(\Delta_j\Delta_k \hat{S}(\vec{r}\,) + \Delta_j B_k(\vec{r}\,)\right) \\ &= \epsilon_{jk}\Delta_j B_k(\vec{r}\,) \\ &= \rho(\vec{x})\end{aligned} \tag{6.8.2}$$

where I used the antisymmetry of the ϵ_{jk}. Thus, the background fields $B_k(\vec{r}\,)$ are determined by the conditions

$$\epsilon_{jk}\Delta_j B_k(\vec{r}\,) = \rho(\vec{x}) = (-1)^{x_1+x_2}. \tag{6.8.3}$$

Notice that the *electrostatic-like* constraint Eq. (6.7.18) (i.e Gauss's law) has become the *magnetostatic* constraint Eq. (6.8.3). This is the usual electric, magnetic duality.

The set of solutions of Eq.(6.8.3) is in one to one correspondance with the dimer configurations of the lattice since Eq.(6.8.3) is the dual version of the constraint, Eq.(6.7.11). Moreover, different solutions $B_k(\vec{r}\,)$ and $B'_k(\vec{r}\,)$ are related through a gauge transformation since their difference $\bar{B}_k(\vec{r}\,) \equiv B_k(\vec{r}\,) - B'_k(\vec{r}\,)$ must satisfy

$$\begin{aligned}\epsilon_{jk}\Delta_j \bar{B}_k(\vec{r}\,) &= (\epsilon_{jk}\Delta_j B_k(\vec{r}\,) - \epsilon_{jk}\Delta_j B'_k(\vec{r}\,)) \\ &= (\rho(\vec{x}) - \rho(\vec{x})) \\ &= 0.\end{aligned} \tag{6.8.4}$$

In other words, $\bar{B}_k(\vec{r}\,)$ is curl free. Hence, at least locally, $\bar{B}_k(\vec{r}\,)$ must be a pure gradient

$$\bar{B}_k(\vec{r}\,) \equiv \Delta_k \Gamma(\vec{r}\,), \tag{6.8.5}$$

where, without loss of generality, $\Gamma(\vec{r}\,)$ is an integer-valued function on the dual lattice.

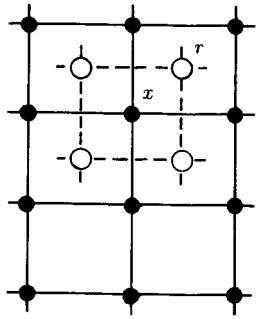


Figure 6.14 The sites of the direct lattice (filled circles) are labelled by $\vec{x}$ and the sites of the dual lattice (empty circles) are labelled by $\vec{r}$.

A local change in the gauge of $B_k(\vec{r}\,)$ can thus be absorbed in an appropriate redefinition of the operators $\hat{S}(\vec{r}\,)$

$$\hat{S}(\vec{r}\,) = \hat{S}'(\vec{r}\,) - \Gamma(\vec{r}\,). \tag{6.8.6}$$

There exists however a set of $\bar{B}_k(\vec{r}\,)$ which cannot be done away with by a suitable redefinition of the variables $\hat{S}(\vec{r}\,)$. They correspond to *large* gauge transformations, i.e. gauge transformations which change the value of the line integral (or sum) of $\bar{B}_k(\vec{r}\,)$ along a non-contractible loop around the torus (see Fig. 6.15). There are two generically non-contractible loops: one along the x_1 direction, $\gamma_1(\vec{r}\,)$, and the other along the x_2 direction, $\gamma_2(\vec{r}\,)$; where $\gamma_1(\vec{r}\,)$ and $\gamma_2(\vec{r}\,)$ go through the dual site $\vec{r}$ (Fig. 6.15). Thus the line integrals $I_{\gamma_1(\vec{r}\,)}[\vec{B}]$ and $I_{\gamma_2(\vec{r}\,)}[\vec{B}]$, defined by

$$\begin{aligned} I_{\gamma_1(\vec{r}\,)}[\vec{B}] &\equiv \sum_{\gamma_1(\vec{r}\,)} B_1(\vec{r}\,) \equiv \sum_{n_1=1}^{L} B_1(\vec{r} + n_1\hat{e}_1), \\ I_{\gamma_2(\vec{r}\,)}[\vec{B}] &\equiv \sum_{\gamma_2(\vec{r}\,)} B_2(\vec{r}\,) \equiv \sum_{n_2=1}^{L} B_2(\vec{r} + n_2\hat{e}_2), \end{aligned} \tag{6.8.7}$$

are invariant under ("small") gauge transformations (which satisfy periodic boundary conditions). However, ("large") gauge transformations, which do not respect the periodic boundary conditions, do change the values of $I_{\gamma_1(\vec{r})}[\vec{B}]$ and $I_{\gamma_2(\vec{r}\,)}[\vec{B}]$. The constraint (see Eq. (6.7.24))

$$\sum_{\vec{x},j} \hat{E}_j^2(\vec{x}) = \frac{L^2}{2} \tag{6.8.8}$$

requires that there should be no bond occupied by more than one dimer. These restrictions imply that the only allowed *large* gauge transformations have to

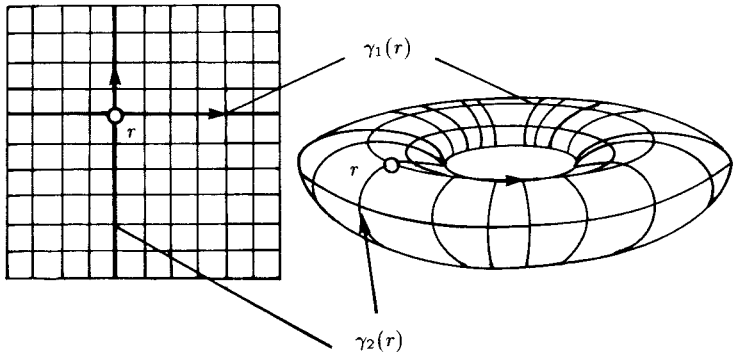


Figure 6.15 A square lattice with periodic boundary conditions is isomorphic to a torus. Two non-contractible loops $\gamma_1(\vec{r}\,)$ and $\gamma_2(\vec{r}\,)$.

satisfy a uniformity condition. For instance, a large gauge transformation which raises $I_{\gamma_1(\vec{r}\,)}[\vec{B}]$ by $+1$ everywhere, has the form (Fig. 6.16)

$$\bar{B}_k(\vec{r}\,) = \delta_{r_2,n_0}\delta_{k,1} \tag{6.8.9}$$

where n_0 is an integer $1 \leq n_0 \leq L$.

What is the meaning of these large gauge transformations? Recall that $\hat{E}_j(\vec{x})$ is given by

$$\hat{E}_j(\vec{x}) = \epsilon_{jk}(\Delta_k \hat{S}(\vec{r}\,) + B_k(\vec{r}\,)). \tag{6.8.1}$$

If we regard the operators $\hat{S}(\vec{r}\,)$ as the *quantum fluctuations* and $B_k(\vec{r}\,)$ as a *classical background*, we see that the configurations with $S(\vec{r}\,) = 0$ (or constant) have $E_j(\vec{x}) = \epsilon_{jk}B_k(\vec{r}\,)$. In other words, the classical background fields $B_k(\vec{r}\,)$ represent a set of classical dimer configurations which can be regarded as the *parent states* for the quantum evolution of the system. Indeed the line integral $I_{\gamma_i(\vec{r}\,)}[\vec{B}]$ is then from Eq. (6.7.17)

$$\begin{aligned} I_{\gamma_i(\vec{r}\,)}[\vec{B}] &= \sum_{\gamma_i(\vec{r}\,)} B_i(\vec{r}\,) \\ &= \sum_{\gamma_i(\vec{x})} \epsilon_{ji} E_j(\vec{x}) \\ &= \epsilon_{ji} \sum_{n_i=1}^{L} (-1)^{x_1+x_2+n_i} L_j(\vec{x} + n_i\hat{e}_i). \end{aligned} \tag{6.8.10}$$

Thus, $I_{\gamma_i(\vec{r}\,)}[\vec{B}]$ is the sum of the differences in the number of dimers occupying neighboring parallel links. This quantity is invariant under the dynamics

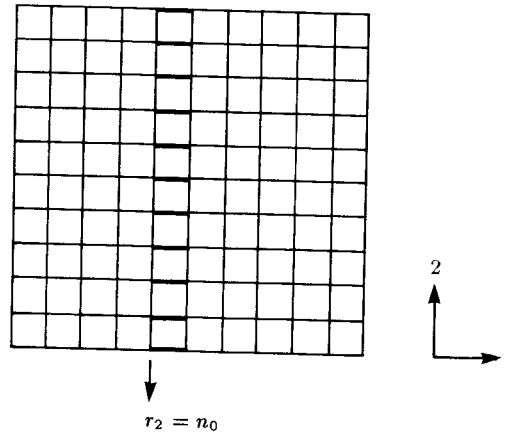


Figure 6.16 A large gauge transformation.

of the Quantum Dimer Model. Solutions which differ by local gauge transformations are equivalent to classical dimer configurations which differ by the "resonating" (or flipping) of a set (or sets) of plaquettes whose boundaries are contractable loops. Large gauge transformations correspond to processes in which a set of valence bonds circulate all the way around a non-contractable loop. Thus, the dimer configurations can be classified by the value of the circulation $\sum_{\gamma_i(\vec{r}\,)} B_i(\vec{r}\,)$ along a non-contractable loop. We can then identify $I_{\gamma_i(\vec{r}\,)}[\vec{B}]$ with the *winding number* introduced by Kivelson and Rokhsar [Rokhsar 88].

Consider for instance configurations which belong to the class with vanishing winding numbers $\sum_{\gamma_i(\vec{r}\,)} B_i(\vec{r}\,) = 0$. In the gauge $B_1(\vec{r}\,) = 0$, there are two possible solutions to Eq. (6.8.3)

$$\begin{aligned} B_1^{(1)}(\vec{r}\,) &= 0, \\ B_2^{(1)}(\vec{r}\,) &= -(\frac{1+(-1)^{r_1}}{2})(-1)^{r_2}, \end{aligned} \tag{6.8.11}$$

and

$$\begin{aligned} B_1^{(2)}(\vec{r}\,) &= 0, \\ B_2^{(2)}(\vec{r}\,) &= +(\frac{1-(-1)^{r_1}}{2})(-1)^{r_2}. \end{aligned} \tag{6.8.12}$$

In the gauge $B_2(\vec{r}\,) = 0$, there are also two (analogous) solutions. It is easy to see that these solutions are in the one-to-one correspondence with the four degenerate column or Peierls states (Fig. 6.17). It is clear that there should be a connection between the degeneracy of the ground state and its winding number. Indeed, the number of distinct solutions of Eq. (6.8.3) for a sector with a given winding number *is* equal to the degeneracy of the ground state in that sector. Since the line integrals do not change under the dynamics and the B_k's determine the subspace of states which are being considered, we expect that the winding number should determine the ground state degeneracy of the full quantum theory *unless* extra degeneracies occur, as a results of one or more modes becoming gapless. These arguments can be generalized to systems with valence bonds of finite but arbitrary length. In terms of the $\frac{1}{N}$ expansion, this means that this topological degeneracy is valid order by order in the $\frac{1}{N}$ expansion.

Now that we have solved the constraint Eq. (6.7.18), we can write the dual form of the Hamiltonian. I will assume that the constraint has been solved in a sector with winding number $I_{\gamma_i(\vec{r}\,)}[\vec{B}]$, $i = 1, 2$. We will have to find which sector yields the lowest ground state energy. The solution of the constraint, Eq. (6.7.18),

$$\hat{E}_j(\vec{x}) = \epsilon_{jk}[\Delta_k \hat{S}(\vec{r}\,) + B_k(\vec{r}\,)] \tag{6.8.1}$$

is one of the equations we need. We also need to define the *momentum* $\hat{P}(\vec{r}\,)$ canonically conjugate to $\hat{S}(\vec{r}\,)$

$$[\hat{P}(\vec{r}\,), \hat{S}(\vec{r}\,')] = i\delta_{\vec{r},\vec{r}'}. \tag{6.8.13}$$

Since the spectrum of $\hat{S}(\vec{r}\,)$ is the set of integers, $\hat{P}(\vec{r}\,)$ should have eigenvalues $P(\vec{r}\,)$ in the range $0 \leq P(\vec{r}\,) < 2\pi$. It is easy to see that the circulation $\sum_{\text{plaquette}} \hat{A}_j(\vec{x})$, around an elementary plaquette centered at dual site $\vec{r}$, has the same effect on its Hilbert space as $\hat{P}(\vec{r}\,)$ has on the integer $\hat{S}(\vec{r}\,)$. More specifically, according to Eqs. (6.7.7) and (6.7.17) the raising operator $e^{i\sum_{\text{plaquette}} \hat{A}_j(\vec{x})}$ shifts the eigenvalues of $\hat{E}_j(\vec{x})$ by $+1$ on the oriented path around the plaquette. This has exactly the effect of raising $\hat{S}(\vec{r}\,)$, on the dual lattice, by $+1$. Thus, we identify

$$\sum_{\text{plaquette}} \hat{A}_j(\vec{x}) \equiv \hat{P}(\vec{r}\,). \tag{6.8.14}$$

Alternatively, it is easy to check the consistency of this identification by an explicit calculation of the commutation relations.

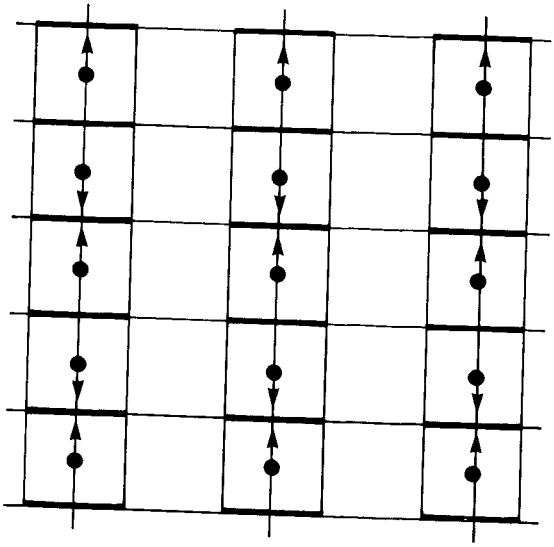


Figure 6.17 A column state and the background configuration of the B_k' s associated with it.

The Hamiltonian dual to that of Eq. (6.7.25) is

$$H = \frac{1}{2k}\left(\sum_{\vec{r},k}(\Delta_k \hat{S}(\vec{r}\,) + B_k(\vec{r}\,))^2 - \frac{L^2}{2}\right) - 2\bar{J}\sum_{\vec{r}}\cos\left(\hat{P}(\vec{r}\,)\right) + \\ + \frac{V}{2}\sum_{\vec{r}}\left(\left(\Delta_1(\Delta_1\hat{S}(\vec{r}\,) + B_1(\vec{r}\,))\right)^2 + \left(\Delta_2(\Delta_2\hat{S}(\vec{r}\,) + B_2(\vec{r}\,))\right)^2\right) + \\ - \frac{VL^2}{2}, \tag{6.8.15}$$

where the limit $k \to 0$ is always meant. Also, in principle, all topological terms have to be considered. We will keep the sector which minimizes the ground state energy. All the inequivalent solutions of Eq. (6.8.3) will represent degenerate states. The manifold of degenerate states is closed under the group of lattice translations and rotations by $\frac{\pi}{2}$. From now on we will work within a given topological sector.

By comparing the QDM Hamiltonian, Eq. (6.7.25), and its dual, Eq. (6.8.15), we notice several features: (i) kinetic and potential energy terms have been exchanged, (ii) the degrees of freedom in Eq. (6.7.25) are phases (i.e. elements of the group U(1)) whereas the degrees of freedom in Eq. (6.8.15) are integers and (iii) Hamiltonian of Eq. (6.8.15) has a *global* symmetry $\hat{S}(\vec{r}\,) \to \hat{S}'(\vec{r}\,) + n_0$ (with n_0 an arbitrary *integer*) whereas Eq. (6.7.25) has a local gauge symmetry. These features are present for all systems related through a duality transformation except (iii), which only holds in 2+1 dimensions [Fradkin 78].

A system with integer-valued degrees of freedom is usually referred to as a Discrete Gaussian (DGM) or Solid-on-Solid (SOS) Model. It was originally introduced by Onsager for the study of the statistical mechanics of *classical* interfaces. In that context $\hat{S}(\vec{r}\,)$ represents the height of a column of identical atoms standing atop the lattice site $\vec{r}$. The set of values of $\{S(\vec{r})\}$ can then be regarded as the surface (or interface) of a three-dimensional solid (Fig. 6.18). The constraint implied by the limit $k \to 0$ represents a *restriction* on this DGM model. The last term in Eq. (6.8.15), which represents a next-nearest-neighbor interaction between atoms, has the form of a *Laplacian* coupling. The second term is responsible for the quantum dynamics of the system. There is a very large body of literature on SOS and DG models. We will not discuss it here. Its most studied system has the classical Hamiltonian H_c

$$H_c = \frac{\gamma}{2}\sum_{\vec{r},k=1,2}(\Delta_k S(\vec{r} + \hat{e}_2))^2 \tag{6.8.16}$$

where γ is a constant. Most studies deal with this classical problem although the role of quantum fluctuations has also been considered.

Classically, systems such as the SOS model usually exhibit two distinct phases. At high temperatures $T > T_R$ the interface is *rough*, whereas at low temperatures $T < T_R$, the interface is *smooth*. The temperature T_R is the location of a critical point at which this *roughening* transition takes place. The

natural correlation functions of this problem are the height-height correlation function $G(\vec{r}-\vec{r}\,')$:

$$G(\vec{r}-\vec{r}\,') = \langle \hat{S}(\vec{r}\,) - \hat{S}(\vec{r}\,') \rangle \tag{6.8.17}$$

and the order parameter correlation function $g_\alpha(\vec{r}-\vec{r}\,')$:

$$g_\alpha(\vec{r}-\vec{r}\,') = \langle e^{i\alpha\hat{S}(\vec{r}\,)} e^{-i\alpha\hat{S}(\vec{r}\,')} \rangle, \tag{6.8.18}$$

where α is an arbitrary angle. For the *classical unrestricted* system, one finds the asymptotic behavior of $g_\alpha(R)$ where $R \equiv |\vec{r}-\vec{r}\,'| \gg a_0$ (a_0 the lattice constant), to be

$$g_\alpha(R) \approx \begin{cases} M^2 + \text{const} \times e^{\frac{-R}{\xi(T)}} & \text{for } T < T_R \text{ (smooth phase)}, \\ \text{const} \times R^{-\eta(\alpha,T)} & \text{for } T > T_R \text{ (rough phase)}, \end{cases} \tag{6.8.19}$$

where $\xi(T)$ is the correlation length, M^2 is the square of the order parameter and the exponent η is a function of α and the temperature. The corresponding behavior of $G(R)$ is

$$G(R) \approx \begin{cases} m^2 + \text{const} \times e^{\frac{-R}{\xi(T)}} & \text{for } T < T_R \text{ (smooth phase)}, \\ \text{const} \times \ln(\frac{R}{a_0}) & \text{for } T < T_R \text{ (rough phase)}, \end{cases} \tag{6.8.20}$$

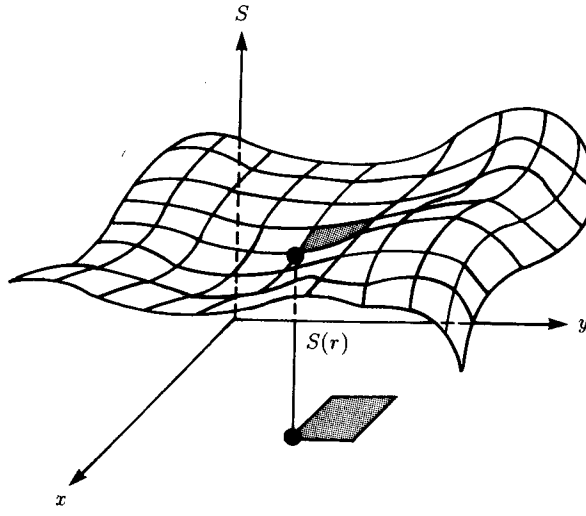


Figure 6.18 The configuration $\{S(\vec{r})\}$ parametrizes a surface in the Solid-on-Solid model.

where m^2 represents the square of the average height, $\langle S(\vec{r}\,)\rangle$.

The quantum fluctuations change this picture completely. If we ignore the restriction ($k \to 0$) and neglect the effects of *frustration* (introduced by the fields B_k), we arrive at the quantum DGM.

Let us introduce a path-integral for this system. It will be convenient for us to work in imaginary time so that we can also discuss thermal fluctuations. At a non-zero temperature T, the partition function of the quantum system is

$$Z = \mathrm{Tr}\ e^{-\beta H} \tag{6.8.21}$$

where $\beta = \frac{1}{T}$ and H is the Hamiltonian of Eq. (6.8.15). In order to derive a path-integral we proceed in the usual fashion [Feynman 65] . We first split up the imaginary time interval $0 \leq \tau \leq \beta$ into N_τ time-steps, each of size Δ_τ, such that

$$\Delta_\tau N_\tau = \beta. \tag{6.8.22}$$

The limit $\Delta_\tau \to 0$ and $N_\tau \to \infty$ is always implied. Next, we write

$$\begin{aligned} Z &= \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \mathrm{Tr}\ [e^{-(\Delta_\tau)H}]^{N_\tau} \\ &\equiv \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \mathrm{Tr}\ [e^{-(\Delta_\tau)H_{\mathrm{kin}}} e^{-(\Delta_\tau)H_{\mathrm{pot}}}]^{N_\tau} \end{aligned} \tag{6.8.23}$$

where we have split the Hamiltonian into a kinetic energy term (the second term of Eq. (6.8.15)) H_{kin} and a potential energy term H_{pot} (the rest). The next step is to introduce a resolution of the identity in terms of a complete set of eigenstates $|\{S(\vec{r},t)\}\rangle$ of the operators $\{\hat{S}(\vec{r},t)\}$ between neighboring factors of $e^{-\beta H}$:

$$Z = \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \sum_{\{S(\vec{r},j)\}=-\infty}^{+\infty} \prod_{j=1}^{N_\tau} \langle \{S(\vec{r},j)\}|e^{-\Delta_\tau H}|\{S(\vec{r},j+1)\}\rangle \tag{6.8.24}$$

with periodic boundary conditions in time, i.e.

$$|\{S(\vec{r},N_\tau+1)\}\rangle \equiv |\{S(\vec{r},1)\}\rangle. \tag{6.8.25}$$

In Eq. (6.8.24) the integer j represents the j^{th} time step and $\tau_j = \tau_0 + j\Delta_\tau$.

Let us compute the matrix elements

$$\begin{aligned} \langle \{S(\vec{r},j)\}|e^{-\Delta_\tau H}|\{S(\vec{r},j+1)\}\rangle \approx & \langle \{S(\vec{r},j)\}|e^{-\Delta_\tau H_{\mathrm{kin}}} e^{-\Delta_\tau H_{\mathrm{pot}}}|\{S(\vec{r},j+1)\}\rangle \\ = & \langle \{S(\vec{r},j)\}|e^{-\Delta_\tau H_{\mathrm{kin}}}|\{S(\vec{r},j+1)\}\rangle \\ & \times e^{-\Delta_\tau H_{\mathrm{pot}}(\{S(\vec{r},j+1)\})} \end{aligned} \tag{6.8.26}$$

where I used the facts that Δ_τ is small and that H_{pot} is *diagonal* in the basis $|\{S(\vec{r},j)\}\rangle$. In fact,

$$e^{-\Delta_\tau H_{\mathrm{pot}}}|\{S(\vec{r},j)\}\rangle = e^{-\Delta_\tau H_{\mathrm{pot}}(\{S(\vec{r},j+1)\})}|\{S(\vec{r},j)\}\rangle \tag{6.8.27}$$

with an eigenvalue $H_{\rm pot}(\{S(\vec{r},j)\})$ given by

$$H_{\rm pot}(\{S(\vec{r},j)\}) = \frac{1}{2k}\left(\sum_{\vec{r},k=1,2}\left(\Delta_k S(\vec{r},j)+B_k(\vec{r},j)\right)^2-\frac{L^2}{2}\right) + \frac{V}{2}\sum_{\vec{r},\ k=1,2}\left(\Delta_k^2 S(\vec{r},j)+\Delta_k B_k(\vec{r},j)\right)^2-\frac{VL^2}{2}. \tag{6.8.28}$$

The off-diagonal matrix elements

$$\langle\{S(\vec{r},j)\}|e^{-\Delta_\tau H_{\rm kin}}|\{S(\vec{r},j+1)\}\rangle = \langle\{S(\vec{r},j)\}|e^{2(\Delta_\tau)\bar{J}\sum_{\vec{r}}\cos(\hat{P}(\vec{r}\ ))}|\{S(\vec{r},j+1)\}\rangle \tag{6.8.29}$$

can be evaluated by repeated use of the expansion

$$e^{z\cos p}=\sum_{l=-\infty}^{\infty} I_l(z)e^{ilp}, \tag{6.8.30}$$

where $I_l(z)$ is the Bessel function of order l of imaginary argument. The matrix elements of Eq. (6.8.29) are products of matrix elements of the form

$$\langle S_j|e^{2(\Delta_\tau)\bar{J}\cos(\hat{P})}|S_{j+1}\rangle \tag{6.8.31}$$

which we can write in the form

$$\sum_{l=-\infty}^{+\infty}\langle S_j|e^{il\hat{P}}|S_{j+1}\rangle I_l(2\bar{J}\Delta_\tau)=I_{|S_{j+1}-S_j|}(2\bar{J}\Delta_\tau). \tag{6.8.32}$$

In this last equation we have used the orthogonality of the states $|S_j\rangle$. For convenience, and simplicity, we will use the approximate form of the Bessel function

$$I_l(z)\approx\frac{e^z}{\sqrt{2\pi}}e^{-\frac{l^2}{2z}}\left(1+\mathcal{O}(z^{-1})\right). \tag{6.8.33}$$

Putting it all together we can write the partition function in the suggestive form

$$Z=\lim_{\substack{\Delta_\tau\to 0\\ N_\tau\to\infty}}\sum_{\{S(\vec{r},j)\}}e^{-\mathcal{H}[S]}, \tag{6.8.34}$$

where the Euclidean (discretized) action $\mathcal{H}[S]$ is given by $(j=1,\ldots,N_\tau)$

$$\mathcal{H}[S]=\frac{1}{4\bar{J}\Delta_\tau}\sum_{\vec{r},j}\left[\Delta_0 S(\vec{r},j)\right]^2+ + \frac{\Delta_\tau}{2k}\left(\sum_{\substack{\vec{r},j\\ k=1,2}}\left(\Delta_k S(\vec{r},j)+B_k(\vec{r},j)\right)^2-\frac{L^2}{2}N_\tau\right)+ + \frac{V\Delta_\tau}{2}\sum_{\substack{\vec{r},j\\ k=1,2}}\left(\left(\Delta_k^2 S(\vec{r},j)+\Delta_k B_k(\vec{r},j)\right)^2\right). \tag{6.8.35}$$

I have also used the notation

$$\Delta_0 S(\vec{r}, j) \equiv S(\vec{r}, j) - S(\vec{r}, j-1). \tag{6.8.36}$$

Thus the quantum partition function of the dimer model is given by the classical partition function of a *Discrete Gaussian Model* in *three* Euclidean dimensions on a cubic lattice of size $L^2 N_\tau$. This system looks very similar to its two-dimensional classical counterpart Eq. (6.8.16) except for the fact that it is frustrated ($B_k \neq 0$), restricted ($k \to 0$) and it has second neighbor interactions ($V \neq 0$) in space.

If we work in the sector with zero-winding number, the configurations with $S(\vec{r}, j) = n_0$, a constant, represent the column states. Conversely, in the sector with maximal winding number, for instance $\sum_{\gamma_1(\vec{r}\,)} B_1(\vec{r}\,) = \frac{L}{2}$, the configuration $S(\vec{r}, j) = n_0$ is a staggered crystal. Which state dominates can only be discerned by solving the partition function Eq. (6.8.34). The action $\mathcal{H}[S]$, Eq. (6.8.35), is such that, for small Δ_τ, the V-term plays little role while the fluctuations of $S(\vec{r}, t)$ in time tend to be suppressed. The column states have a very large entropy ($\propto L^2$) whereas the staggered states, due to the constraints, have virtually no excitations. Numerical simulations indicate that, for V small and positive, the column state is stable. For large V the staggered state should win, at least at low temperatures. Hence, we expect that the QDGM should be in a smooth phase, albeit degenerate (see the discussion above).

6.9 Quantum Dimer Models and Monopole Gases

In Section 6.7 we used an intuitive argument which indicated that monopole configurations of the gauge fields play a fundamental role in this problem. We will now examine this issue more closely for the case $V = 0$.

The easiest way to relate the Quantum Dimer Model to a gas of monopoles is to apply the Poisson Summation Formula

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{m=-\infty}^{+\infty} \int d\phi e^{i2\pi m\phi} f(\phi) \tag{6.9.1}$$

to the *three-dimensional* Discrete Gaussian Model with action (6.8.35). This amounts to replace all the integer variables $\{S(\vec{r}, j)\}$ by continuous variable

$\{\phi(r)\}$ and another set of integers $\{m(r)\}$ where now $r = (r_0, r_1, r_2)$ are three-dimensional lattice vectors in Euclidean space-time:

$$
\begin{aligned}
Z &= \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \sum_{\{S\}} e^{-\mathcal{H}[S]} \\
&= \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \sum_{\{m(r)\}} \int \mathcal{D}\phi e^{i2\pi \sum_r m(r)\phi(r) - \mathcal{H}[\phi]} \\
&= \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} \sum_{\{m(r)\}} \int \mathcal{D}\phi e^{\frac{\Delta_\tau}{2k}\left(\frac{N_\tau L^2}{2} - \sum_{r,k=1,2} B_k^2(r)\right)} \times \\
& e^{-\frac{1}{2}\sum_r \left(\left(\frac{1}{2\bar{J}\Delta_\tau}(\Delta_0\phi(r))^2 + \frac{\Delta_\tau}{k}\sum_{k=1,2}(\Delta_k\phi(r))^2\right) + \phi(r)\left(i2\pi m(r) + \frac{\Delta_\tau}{k}\sum_{k=1,2}\Delta_k B_k(r)\right)\right)}
\end{aligned}
\tag{6.9.2}
$$

Since $\mathcal{H}[\phi]$ is *quadratic* in ϕ, the ϕ variables can be integrated out, yielding the result in terms of a partition function for a generalized Coulomb Gas, Z_{CG}, of the form

$$
\begin{aligned}
Z = \lim_{\substack{\Delta_\tau \to 0 \\ N_\tau \to \infty}} & e^{\frac{\Delta_\tau}{2k}\left(\frac{N_\tau L^2}{2} - \sum_{r,k=1,2} B_k^2(\vec{r}\,)\right)} \times \\
& (\mathrm{Det}\,\frac{M}{2\pi})^{-\frac{1}{2}} e^{+\frac{1}{2}(\frac{\Delta_\tau}{k})^2 \sum_{r,r'} \epsilon_{\alpha\mu\lambda}\Delta_\lambda^r B_\mu(r) G_0(r-r') \epsilon_{\alpha\nu\rho}\Delta_\nu^{r'} B_\rho(r')} Z_{\mathrm{CG}},
\end{aligned}
\tag{6.9.3}
$$

where

$$
\mathrm{Det}\, M = \mathrm{Det}\left(\frac{1}{2\bar{J}\Delta_\tau}\Delta_0^2 + \frac{\Delta_\tau}{k}\sum_{k=1,2}\Delta_k^2\right), \tag{6.9.4}
$$

while $G_0(r-r')$ is the three-dimensional anisotropic lattice Green function defined by

$$
-\left(\frac{1}{2\bar{J}\Delta_\tau}\Delta_0^2 + \frac{\Delta_\tau}{k}\sum_{k=1,2}\Delta_k^2\right) G_0(r-r') = \delta_{r,r'}, \tag{6.9.5}
$$

(the minus sign comes from a "partial integration") and Z_{CG} is given by

$$
Z_{\mathrm{CG}} = \sum_{\{m(r)\}} e^{-2\pi^2 \sum_{r,r'} m(r) G_0(r-r') m(r') + i2\pi \frac{\Delta_\tau}{k}\sum_r m(r)\Psi(r)}. \tag{6.9.6}
$$

The phase $\Psi(r)$ is found to be equal to

$$
\Psi(r) = \sum_{r'} G_0(r-r')\Delta_k^{r'} B_k(r'). \tag{6.9.7}
$$

In the thermodynamic limit ($L, N_\tau \to \infty$) [Banks 77] and at zero temperature $G_0(r-r')$ is

$$
G_0(r-r') = \int_{-\pi}^{\pi} \frac{d^3q}{(2\pi)^3} \frac{\frac{1}{4}e^{iq(r-r')}}{\frac{1}{2\bar{J}\Delta_\tau}\sin^2(\frac{q_0}{2}) + \sum_{k=1,2}\frac{\Delta_\tau}{k}\sin^2(\frac{q_k}{2})}. \tag{6.9.8}
$$

In the time-continuum limit we find ($\omega \equiv \Delta_\tau q_0$)

$$\lim_{\Delta_\tau \to 0} G_0(\vec{r}-\vec{r}\,', \tau-\tau') = \lim_{\Delta_\tau \to 0} \int_{\frac{-\pi}{\Delta_\tau}}^{\frac{\pi}{\Delta_\tau}} \frac{d\omega}{2\pi} \int_{-\pi}^{\pi} \frac{d^2q}{(2\pi)^2} \frac{e^{i\left(\omega(\tau-\tau')+\vec{q}\cdot(\vec{r}-\vec{r}\,')\right)}}{\frac{\omega^2}{2\bar{J}} + \frac{4\Delta_\tau}{k}\sum_{k=1,2}\sin^2\left(\frac{q_k}{2}\right)}. \tag{6.9.9}$$

At long distances ($R = |\vec{r} - \vec{r}\,'| \gg a_0$) and long (Euclidean) times ($\bar{\tau} = |\tau - \tau'| \gg \Delta_\tau$), $G_0(R, \bar{\tau})$ behaves like

$$G_0(R, \bar{\tau}) \approx \frac{k}{4\pi} \frac{1}{\sqrt{\bar{\tau}^2 + \frac{2\bar{J}\Delta_\tau}{k} R^2}}. \tag{6.9.10}$$

Except for the anisotropy ($2\bar{J} \neq k$) this is just the three-dimensional Coulomb interaction. Thus the monopoles behave like a gas of charged particles (of both signs) in three dimensions, with a pair effective interaction (at long distances) V_{eff} given by

$$V_{\text{eff}}(R, \bar{\tau}) = 2\pi^2 G_0(R, \bar{\tau}) \approx \frac{\pi k}{2} \frac{1}{\sqrt{\bar{\tau}^2 + \frac{2\bar{J}\Delta_\tau}{k} R^2}}. \tag{6.9.11}$$

The total partition function is

$$Z = e^{\frac{3}{16}\frac{L^2}{kT}} (\text{Det}\, \frac{M}{2\pi})^{-\frac{1}{2}} \sum_{\{m(r)\}} e^{-\frac{1}{2}\sum_{r,r'} m(r) V_{\text{eff}}(r,r') m(r') + i2\pi m(r)\Psi(r)}. \tag{6.9.12}$$

The phase $\theta(r) = 2\pi\Psi(r)$ (see Eq. (6.9.6)) takes one of four possible values, one per sublattice,

$$\theta(r) = \begin{cases} \frac{-\pi}{4} & \text{for } r_1 \text{ even and } r_2 \text{ even,} \\ \frac{+3\pi}{4} & \text{for } r_1 \text{ odd and } r_2 \text{ odd,} \\ \frac{+\pi}{4} & \text{for } r_1 \text{ odd and } r_2 \text{ even,} \\ \frac{-3\pi}{4} & \text{for } r_1 \text{ even and } r_2 \text{ odd.} \end{cases} \tag{6.9.13}$$

The conclusion is that in this case, very much like Polyakov's, the system is also equivalent to a three-dimensional Coulomb gas. Apart from some relatively minor issues, the main difference between Polyakov's case and the problem of our interest, is the presence of the phases $\theta(r)$. These phases can be thought of as *Berry phases* since they arise from non-trivial overlaps of the state of the system at nearby times. Indeed, Read and Sachdev have derived these phases, following Haldane's original suggestion, by means of an adiabatic calculation. It is remarkable that we find the same answer but started from a regime in which a non-linear sigma model cannot possibly work.

6.10 Chiral Spin States

So far we have considered solutions of the Mean-Field equations which respect time-reversal invariance. We will now consider states for which time-reversal invariance is broken. In terms of the Mean-Field theory of Section 6.4, we will consider situations in which the *phase* $\bar{\mathcal{A}}_j(\vec{x},t)$ of the link variable $\bar{\chi}_j(\vec{x},t)$ has a non-zero curl $\bar{\mathcal{B}}(\vec{x},t)$ around an elementary plaquette

$$\bar{\mathcal{B}}(\vec{x},t) = \sum_{\text{plaquette}} \bar{\mathcal{A}}_j(\vec{x},t) = \Delta_1 \bar{\mathcal{A}}_2(\vec{x},t) - \Delta_2 \bar{\mathcal{A}}_1(\vec{x},t). \tag{6.10.1}$$

In Section 6.5 we argued that such flux states violate time-reversal invariance unless $\bar{\mathcal{B}}(\vec{x},t) = 0, \pi$. A solution $\bar{\chi}_j(\vec{x},t)$ of the saddle-point equation applied to Eq. (6.4.5) satisfies

$$\frac{2}{J}\langle \bar{\chi}_j^*(\vec{x},t)\rangle = \langle c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}+\hat{e}_j,t)\rangle. \tag{6.10.2}$$

For a solution with $\bar{\rho}_j(\vec{x},t) = \bar{\rho}_j$ a constant and $\bar{\mathcal{A}}_j(\vec{x},t) \neq 0$, we get

$$\frac{2}{J}\bar{\rho}_j e^{-i\bar{\mathcal{A}}_j(\vec{x},t)} = \langle c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}+\hat{e}_j,t)\rangle. \tag{6.10.3}$$

Thus a flux phase implies that the *product of the band amplitudes* $\langle c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}+\hat{e}_j,t)\rangle$ around a closed loop γ of the lattice should have a *phase* determined by the *flux* going through the loop. Alternatively, we can consider not the product (around the loop) of expectation values $\langle c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}+\hat{e}_j,t)\rangle$, but the *expectation value of the path-ordered product*

$$W(\gamma) = \langle \prod_{(\vec{x},\vec{x}')\epsilon\gamma} \left(c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}\,')\right)\rangle \tag{6.10.4}$$

where $(\vec{x},\vec{x}\,')$ denotes a link of the lattice, with endpoints at $\vec{x}$ and $\vec{x}\,'$, which belongs to the *close* path γ. The expectation value $\langle c_\alpha^\dagger(\vec{x},t) c_\alpha(\vec{x}+\hat{e}_j,t)\rangle$ is not gauge invariant. Accordingly, Elitzur's theorem implies that this expectation value is actually equal to zero. As a matter of fact, the solutions of the saddle-point equations are not unique. All the configurations which can be reached by means of a local gauge transformation from a given solution are solutions too. The saddle-point-approximation violates this condition. The invariance is restored by fluctuations. The main effect of fluctuations is to rid the system of spurious states which violate gauge invariance. We will come back to this point shortly, when we discuss the spectrum of disordered spin states more generally.

How can we compute expectation values such as $W(\gamma)$ from a path-integral written in terms of $\chi_j(\vec{x},t)$ fields? Let us go back to the path-integral for this system with the effective Lagrangian density of Eq. (6.4.21). I will only discuss the simpler $n_c = 1$ case. Let us shift the $\mathcal{A}_0(\vec{x},t)$ and $\chi_j(\vec{x},t)$ variables by a fixed, but arbitrary, amount $\tilde{\mathcal{A}}_0(\vec{x},t)$ and $\tilde{\chi}_j(\vec{x},t)$. This is essentially a

mathematical device to compute expectation values involving Fermi field currents. We can regard the $\tilde{A}_0$ and $\tilde{\chi}_j$ as external sources in term of which the shifted Lagrangian density, $\mathcal{L}'$, reads (for $n_c = 1$)

$$\begin{aligned}\mathcal{L}' =& c^\dagger_\alpha(x)(i\partial_t + \mu)c_\alpha(x) + \left(A_0(x) + \tilde{A}_0(x)\right)\left(c^\dagger_\alpha(x)c_\alpha(x) - \frac{N}{2}\right) + \\ &- \frac{N}{J}|\chi_j(x)|^2 + \\ &+ c^\dagger_\alpha(\vec{x},t)\left(\chi_j(\vec{x},t) + \tilde{\chi}_j(\vec{x},t)\right)c_\alpha(\vec{x}+\hat{e}_j,t) + \text{h.c.}\,,\end{aligned} \tag{6.10.5}$$

where we recall that according to Eq. (6.4.22) $\tilde{\chi}_j(\vec{x},t) = \tilde{\chi}^*_{-j}(\vec{x}+\hat{e}_j,t)$. Since $\tilde{\chi}_j(\vec{x},t)$ couples to the hoping term from site $\vec{x}+\hat{e}_j$ to $\vec{x}$, it is clear that the functional differentiation of the action S by $\tilde{\chi}_j(\vec{x},t)$ yields

$$\frac{\delta S}{\delta\tilde{\chi}_j(\vec{x},t)} = \sum_{\alpha=1}^{N}\left(c^\dagger_\alpha(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t)\right), \tag{6.10.6}$$

while functional differentiation with respect to $\tilde{A}_0(\vec{x},t)$ gives

$$\frac{\delta S}{\delta\tilde{A}_0(\vec{x},t)} = \sum_{\alpha=1}^{N} c^\dagger_\alpha(\vec{x},t)c_\alpha(\vec{x},t) - \frac{N}{2} = 0 \tag{6.10.7}$$

as follows from the constraint of Eq. (6.4.20).

Thus, by computing functional derivatives we can compute the desired expectation values. For instance

$$\frac{\delta Z}{\delta\tilde{\chi}_j(\vec{x},t)} = \int \mathcal{D}\chi\mathcal{D}A_0\mathcal{D}c^\dagger\mathcal{D}c\; e^{iS}\; i\frac{\delta S}{\delta\tilde{\chi}_j(\vec{x},t)} \tag{6.10.8}$$

and

$$\begin{aligned}-\frac{i}{Z}\frac{\delta Z}{\delta\tilde{\chi}_j(\vec{x},t)} =& \langle\frac{\delta S}{\delta\tilde{\chi}_j(\vec{x},t)}\rangle \\ =& \langle\sum_{\alpha=1}^{N}\left(c^\dagger_\alpha(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t)\right)\rangle.\end{aligned} \tag{6.10.9}$$

In particular, the path-ordered product $W(\gamma)$ can also be computed. Let p label the p^{th} link on the path γ and $\tilde{\chi}(p)$ the corresponding $\chi_j(\vec{x},t)$, i.e. the link $(\vec{x},\vec{x}+\hat{e}_j,t)$ is the p^{th} link of the path starting at some arbitrary site $\vec{x}_0$ on the path. We can write, for a closed path γ with perimeter $L(\gamma)$

$$\begin{aligned}\frac{1}{Z}\frac{\delta^L Z}{\delta\tilde{\chi}(1)\ldots\delta\tilde{\chi}(L(\gamma))} =& i^{L(\gamma)}\langle\prod_{p=1}^{L(\gamma)}\left(\sum_{\alpha=1}^{N}c^\dagger_\alpha(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t)\right)\rangle \\ \equiv& i^{L(\gamma)}W(\gamma).\end{aligned} \tag{6.10.10}$$

On the other hand, the $\chi_j(\vec{x},t)$ degrees of freedom can be shifted without affecting the value of the partition function

$$\begin{aligned}A_0(\vec{x},t) &= A'_0(\vec{x},t) - \tilde{A}_0(\vec{x},t), \\ \chi_j(\vec{x},t) &= \chi'_j(\vec{x},t) - \tilde{\chi}_j(\vec{x},t).\end{aligned} \tag{6.10.11}$$

After this is done, all the information about the sources is in the quadratic term of $\mathcal{L}'$

$$\begin{aligned}\mathcal{L}' =& c_\alpha^\dagger(x)(i\partial_t + \mu)c_\alpha(x) + \mathcal{A}_0'(x)(c_\alpha^\dagger(x)c_\alpha(x) - \frac{N}{2}) + \\ &-\frac{N}{J}\left(\chi_j'(x) - \tilde{\chi}_j(x)\right)\left(\chi_j'^*(x) - \tilde{\chi}_j^*(x)\right) + \\ &+c_\alpha^\dagger(\vec{x},t)\chi_j'(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t) + c_\alpha^\dagger(\vec{x}+\hat{e}_j,t)\chi_j'^*(\vec{x},t)c_\alpha(\vec{x},t).\end{aligned} \tag{6.10.12}$$

Thus,

$$\begin{aligned}\langle\sum_{\alpha=1}^{N} c_\alpha^\dagger(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t)\rangle &= \langle\frac{\delta S}{\delta\tilde{\chi}_j(\vec{x},t)}\rangle|_{\tilde{\chi}_j=0} \\ &= \frac{2N}{J}\langle\chi_j'^*(\vec{x},t)\rangle.\end{aligned} \tag{6.10.13}$$

Similarly, $W(\gamma)$ is given by

$$W(\gamma) = \langle\prod_{p=1}^{L(\gamma)}\frac{2N}{J}\chi^*(p)\rangle. \tag{6.10.14}$$

Notice that there is no quadratic term in the action for $\mathcal{A}_0$. Thus, all functional derivatives of Z with respect to $\tilde{\mathcal{A}}_0$ are identically equal to zero

$$\frac{\delta Z}{\delta\tilde{\mathcal{A}}_0} = 0. \tag{6.10.15}$$

This only means that the constraint

$$\sum_{\alpha=1}^{N} c_\alpha^\dagger(\vec{x},t)c_\alpha(\vec{x},t) - \frac{N}{2} = 0 \tag{6.10.16}$$

is strictly enforced at all times and everywhere.

The quadratic terms in $\mathcal{L}'$ express the fluctuations of the amplitude $\rho_j(\vec{x},t)$ of $\chi_j(\vec{x},t)$ but not of its phase, the gauge field $\mathcal{A}_j(\vec{x},t)$. Thus if we imagine a state with $\bar{\rho}_j(\vec{x},t) = \bar{\rho}$, we will still have the fluctuations of the gauge fields $\mathcal{A}_j$ to deal with. The path-ordered product is, in this approximation, equal to

$$W(\gamma) \approx (\frac{2N}{J}\bar{\rho})^{L(\gamma)}\langle e^{i\sum_{l\in\gamma}\bar{\mathcal{A}}(l)}\rangle. \tag{6.10.17}$$

This last expectation value, $\langle e^{i\sum_{l\in\gamma}\mathcal{A}(l)}\rangle$, is known as the Wilson loop. It was introduced in the context of gauge theories of strong interactions (in particle physics) as a way to measure the interaction between quarks. In the present context, it measures the interactions between ideal *static* spinons which are carried around the loop γ. The interaction is mediated by the fluctuations of the field χ_j. The relevance of Wilson loops for flux spin states was first emphasized by Wiegmann [Wiegmann 88].

If the saddle-point-approximation was exact, the fluctuations of the gauge field $\mathcal{A}_j$ could be neglected. Thus $W(\gamma)$ would yield the result

$$W(\gamma) \approx (\frac{2N}{J}\bar{\rho})^{L(\gamma)} \langle e^{i\sum_{l\epsilon\gamma} \bar{\mathcal{A}}(l)} \rangle. \tag{6.10.18}$$

Let $a(\gamma)$ be the *area* of the lattice enclosed by the path γ. Using Stoke's theorem, we would then get

$$W(\gamma) \approx (\frac{2N}{J}\bar{\rho})^{L(\gamma)} \langle e^{ia(\gamma)\bar{\mathcal{B}}} \rangle \tag{6.10.19}$$

where $\bar{\mathcal{B}}$ is the flux per plaquette. If we denote by $\delta\mathcal{A}_j(\vec{x},t)$ the fluctuating part of the gauge field $\mathcal{A}_j(\vec{x},t)$, i.e. the deviation from the saddle-point configuration, we get for $W(\gamma)$

$$W(\gamma) \approx (\frac{2N}{J}\bar{\rho})^{L(\gamma)} \langle e^{ia(\gamma)\bar{\mathcal{B}}} e^{i\sum_{l\epsilon\gamma} \delta\mathcal{A}(l)} \rangle \tag{6.10.20}$$

where the expectation value involves only the fluctuating pieces. It has been argued that flux phases can generally be defined as phases in which ln $W(\gamma)$ has an imaginary part which scales like the area enclosed by the loop γ [Wiegmann 88] [Wen 89] . It is also constructive to consider the situation in which an extra fermion, i.e. a spinon, is added at some site $\vec{x}$ and other one is removed at site $\vec{x}\,'$. The constraints at $\vec{x}$ and $\vec{x}\,'$ are

$$\sum_{\alpha=1}^{N} c_\alpha^\dagger(\vec{y},t)c_\alpha(\vec{y},t) - \frac{N}{2} = \delta_{\vec{y},\vec{x}} - \delta_{\vec{y},\vec{x}'}. \tag{6.10.21}$$

This means that two extra factors enter into the partition function. They have the form $e^{\pm i\int dt\mathcal{A}_0(\vec{x},t)}$. We can close the paths both in the remote past and future (assuming an adiabatic switching on and off, i.e. a smooth path) and

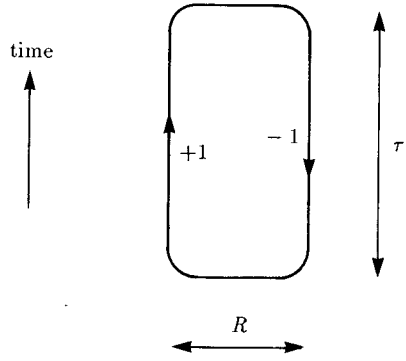


Figure 6.19 A space-time loop γ_t of size $R \times \tau$. A static spinon $(+1)$ and anti-spinon (-1) separated at a distance R.

write the extra contribution as an integral over a closed path γ_t

$$W(\gamma_t) \propto \langle e^{i\sum_{l\in\gamma_t} A(l)} \rangle|_{\gamma_t} \tag{6.10.22}$$

where γ_t stands for a space-time closed loops (see Fig. 6.19) of time span τ and special extent R. Thus $W(\gamma_t)$ measures the change of the ground state energy $\Delta E(\vec{x})$ of the system as a result of the presence of the static spinons

$$W(\gamma_t) = e^{i\tau\Delta E(\vec{x})}. \tag{6.10.23}$$

This expression is valid for $\tau \gg R$. Thus the effective interaction between static sources $V_{\text{eff}}(\vec{x})$ is

$$V_{\text{eff}}(\vec{x}) = \Delta E(\vec{x}) = \lim_{\tau\to\infty} -\frac{i}{\tau} \ln(W(\gamma_t)). \tag{6.10.24}$$

Notice that there is no classical flux associated with space-time loops γ. Thus $W(\gamma_t)$ does not necessarily exhibit the area law of Eq. (6.10.19), associated with the flux phase which we found for space loops. In fact, both $W(\gamma_t)$ and the fluctuating components of the space-like loops have a phase which decays like the *perimeter* of the loop, not its area. This is so because, unlike the case of confining gauge theories without dynamical matter fields, we only have gauge fields associated with a dynamical matter field. The gauge fields themselves do not have any other dynamics of their own. We will return to this important point in the next section.

There is an alternative way of understanding the products over closed loops. Consider the case of three spins, $\vec{S}(1)$, $\vec{S}(2)$ and $\vec{S}(3)$. Let us form the mixed product $\hat{E}_{123}$ which Wen, Wilczek and Zee call the chiral operator

$$\hat{E}_{123} \equiv \vec{S}(1) \cdot \left(\vec{S}(2) \times \vec{S}(3)\right). \tag{6.10.25}$$

Under time reversal $\hat{T}$ we have

$$\hat{T}^{-1}\vec{S}\hat{T} = -\vec{S}. \tag{6.10.26}$$

Thus $\hat{E}_{123}$ is *odd* under $\hat{T}$

$$\hat{T}^{-1}\hat{E}_{123}\hat{T} = -\hat{E}_{123}. \tag{6.10.27}$$

Similarly, under parity, $\hat{P}$, which in two space dimensions is the same as reflection through a link, we have

$$\begin{aligned} \hat{P}^{-1}\hat{E}_{123}\hat{P} &= \vec{S}(1) \cdot \left(\vec{S}(3) \times \vec{S}(2)\right) \\ &= +E_{132} \\ &= -\hat{E}_{123} \end{aligned} \tag{6.10.28}$$

where we have exchanged sites 2 and 3, keeping one fixed. Thus, for the three spins, parity implies turning an even permutation of the three spins into an odd one.

Let us now write $\hat{E}_{123}$ in terms of the link operators $\hat{\chi}(i,j) \equiv c_\alpha^\dagger(i)c_\alpha(j)$. Explicitly one finds [Wen 89]

$$\hat{E}_{123} = \frac{i}{4}\left(\hat{\chi}(1,2)\hat{\chi}(2,3)\hat{\chi}(3,1) - \hat{\chi}(1,3)\hat{\chi}(3,2)\hat{\chi}(2,1)\right). \tag{6.10.29}$$

If we consider now four spins, 1, 2, 3 and 4, we get

$$\begin{aligned}&\hat{\chi}(1,2)\hat{\chi}(2,3)\hat{\chi}(3,4)\hat{\chi}(4,1) - \hat{\chi}(1,4)\hat{\chi}(4,3)\hat{\chi}(3,2)\hat{\chi}(2,1)\\ &= 2i\left(-\hat{E}_{123} - \hat{E}_{134} - \hat{E}_{124} + \hat{E}_{234}\right).\end{aligned} \tag{6.10.30}$$

Thus if $\hat{E}_{123}$ acquires an expectation value, we should expect that the special Wilson loops implied by Eqs. (6.10.29) and (6.10.30) should exhibit a now trivial phase. At the level of the saddle-point-approximation, we expect

$$\begin{aligned}\langle\hat{E}_{123}\rangle =& \frac{i}{4}\langle\hat{\chi}(1,2)\hat{\chi}(2,3)\hat{\chi}(3,1) - \hat{\chi}(1,3)\hat{\chi}(3,2)\hat{\chi}(2,1)\rangle\\ \approx& \frac{i}{4}\left(\frac{2N\bar{\rho}}{J}\right)^3\left(e^{i\bar{\mathcal{B}}_\Delta} - e^{-i\bar{\mathcal{B}}_\Delta}\right)\\ =& -\frac{1}{2}\left(\frac{2N\bar{\rho}}{J}\right)^3\sin(\bar{\mathcal{B}}_\Delta)\end{aligned} \tag{6.10.31}$$

where $\bar{\mathcal{B}}_\Delta$ is the flux through the triangle with vertices at sites 1, 2 and 3. Thus, in a chiral phase, $\hat{E}_{123}$ should have a non-zero expectation value. Please notice that for the non-chiral flux phase, $\bar{\mathcal{B}}_\Delta = \pi$, $\langle\hat{E}_{123}\rangle$ is equal to zero.

For a system with just three spins one-half we can get a very simple interpretation of this statement. For three spins one-half, the Hilbert space is $2^3 = 8$ dimensional. The total spin is $\vec{S} = \vec{S}(1) + \vec{S}(2) + \vec{S}(3)$. The quadratic Casimir operator $\vec{S}^2$ and, say, S_3 commute with each other. What is important is that they also commute with $\hat{E}_{123}$. Thus, $\vec{S}^2$, S_3 and $\hat{E}_{123}$ can be diagonalized simultaneously. I will refer to the eigenvalues of $\hat{E}_{123}$ as the *chirality* χ of the state. The states of the three spins will thus be labeled accordingly by $|s, m; \chi\rangle$, where s is the spin quantum number, m is the z-projection, and χ is the chirality. The total spin s is either zero or $\frac{3}{2}$. The $\frac{3}{2}$ sector can be obtained trivially by applying the lowering operator S^- on the highest weight state $|\uparrow\uparrow\uparrow\rangle$:

$$\begin{aligned}|\uparrow\uparrow\uparrow\rangle &= |\frac{3}{2}, \frac{3}{2}; 0\rangle,\\ |\frac{3}{2}, \frac{3}{2} - m; 0\rangle &= (S^-)^m|\frac{3}{2}, \frac{3}{2}; 0\rangle.\end{aligned} \tag{6.10.32}$$

The state $|\uparrow\uparrow\uparrow\rangle$ has zero chirality since it is invariant under a permutation of any pair of spins. In terms of raising and lowering operators $S^\pm$ and S_3, $\hat{E}_{123}$ has the form

$$\begin{aligned}\hat{E}_{123} = \frac{i}{2}\big(&- S^-(1)S^+(2)S_3(3) + S^+(1)S^-(2)S_3(3) +\\ &+ S^-(1)S_3(2)S^+(3) - S^+(1)S_3(2)S^-(3) +\\ &- S_3(1)S^-(2)S^+(3) + S_3(1)S^+(2)S^-(3)\quad\big).\end{aligned} \tag{6.10.33}$$

Clearly

$$\hat{E}_{123}|\frac{3}{2},\frac{3}{2};\chi\rangle = \hat{E}_{123}|\uparrow\uparrow\uparrow\rangle = 0 \tag{6.10.34}$$

which proves that $\chi_{\uparrow\uparrow\uparrow} = 0$. From the form of $\hat{E}_{123}$ in Eq. (6.10.33), we see that all the states in the same multiplet defined by s and m have the same chirality.

There are two, orthogonal, sectors with $s = 0$, $m = 0$. They differ by their chirality χ. Consider the state $|+\rangle$, defined by the linear superposition

$$|+\rangle = \frac{1}{\sqrt{3}}\left(|\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\rangle e^{i\frac{2\pi}{3}} + |\downarrow\uparrow\uparrow\rangle e^{-i\frac{2\pi}{3}}\right). \tag{6.10.35}$$

This state $|+\rangle$ is an eigenstate of $\hat{E}_{123}$ with eigenvalue χ_+ given by

$$\hat{E}_{123}|+\rangle = -\frac{1}{2}\sin(\frac{2\pi}{3})|+\rangle. \tag{6.10.36}$$

Thus $\chi_+ = -\frac{1}{2}\sin(\frac{2\pi}{3})$. Similarly the state $|-\rangle$

$$|-\rangle = \frac{1}{\sqrt{3}}\left(|\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\rangle e^{-i\frac{2\pi}{3}} + |\downarrow\uparrow\uparrow\rangle e^{i\frac{2\pi}{3}}\right). \tag{6.10.37}$$

has eigenvalue $\chi_- = +\frac{1}{2}\sin(\frac{2\pi}{3})$. Both states, $|+\rangle$ and $|-\rangle$, have $S_3 = +\frac{1}{2}$. Thus we denote $|\pm\rangle$ as the states $|\frac{1}{2},\frac{1}{2};\pm\rangle$. Similarly the states with spin down can also have either chirality. These two remaining states are denoted by $|\frac{1}{2},-\frac{1}{2};\pm\rangle$.

The most singlet-like states, i.e. with smallest spins, can thus be arranged to have non-zero chirality. By inspection of Eqs. (6.10.35) and (6.10.36), we see that a state with non-zero chirality is a state in which a spin down moves around the triangle with a non-zero angular momentum $l = \pm 1$. Thus, a state with non-zero chirality is a state in which there is a non-zero spin current since a down spin is being transported, at a fixed rate, around the triangle.

For a macroscopic system, we can picture a situation in which $\langle\hat{E}_{123}\rangle$ is different from zero everywhere, as in a flux state, by saying that flux states are states in which there are non-vanishing *orbital* spin currents around every elementary plaquette. If we demand that the flux $\vec{\mathcal{B}}$ be uniform throughout the system, we are in fact requiring that the state should in fact exhibit an *orbital ferromagnetism* of some sorts [Volovik 88].

There is one interesting analogy here with the behavior of orbital angular momentum in the A phase of ^3He. As is well known, ^3He becomes a superfluid by forming bound states of two ^3He atoms. The bound state has total spin $S = 1$ (triplet) and orbital angular momentum $l = 1$ (p-wave) [Leggett 74]. In $^3He - A$, the orbital angular momentum vector $\vec{l}$ and the spin $\vec{S}$ of the state are orthogonal to each other. In a thin-film geometry, the orbital angular momenta $\vec{l}$ are all parallel to each other and perpendicular to the surface of the film. This superfluid has orbital ferromagnetism.

6.11 Mean-Field Theory of the Chiral Spin States

Let us consider the Mean-Field theory, or equivalently, the large-N limit, of frustrated Heisenberg antiferromagnet on a square lattice. We have two coupling constants: J (nearest neighbors) and J' (next-nearest neighbors). We have considered this problem in Section 6.1 in which we considered the effects of J' on the Néel state. The effective Lagrangian Eq. (6.4.21) can be easily generalized in order to include the effects of the J'. All we have to do is to decouple the next-nearest neighbor term using the same procedure we used for the nearest-neighbor term [Wen 89]. The Lagrangian density now is, including both J and J',

$$\begin{aligned}\mathcal{L}' =& c^\dagger_\alpha(x)(i\partial_t + \mu)c_\alpha(x) + \mathcal{A}_0(x)\left(c^\dagger_\alpha(x)c_\alpha(x) - \frac{N}{2}\right) + \\ &- \frac{N}{J}|\chi_j(x)|^2 - \frac{N}{J'}|\chi_{j'}(x)|^2 + \\ &+ c^\dagger_\alpha(\vec{x},t)\chi_j(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_j,t) + \text{h.c.} + \\ &+ c^\dagger_\alpha(\vec{x},t)\chi_{j'}(\vec{x},t)c_\alpha(\vec{x}+\hat{e}_1+j'\hat{e}_2,t) + \text{h.c.}\end{aligned} \tag{6.11.1}$$

where $j' = \pm$.

The saddle-point procedure can be carried out along very similar lines. At this level, we assume that the amplitudes $\bar{\chi}_j(\vec{x},t)$ and $\bar{\chi}_{j'}(\vec{x},t)$ are constant in time and as uniform as possible in space. If we choose the gauge of Eq. (6.5.11) (i.e. the gauge we chose in our earlier discussion of the flux phase) we

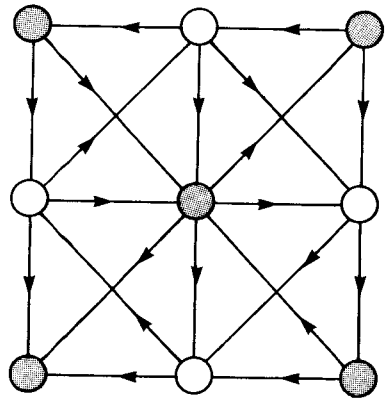


Figure 6.20 Chiral gauge on a frustrated square lattice. The left lower corner is an even-even site. The arrows on the links represent a phase $\frac{\pi}{2}$. The flux on a plaquette is π if the left lower corner is on an even column. Otherwise it is $-\pi$.

get [Wen 89]

$$\begin{aligned} \bar{\chi}_1(\vec{x}) &= \bar{\rho}e^{+i\frac{\pi}{2}}, & \bar{\chi}_2(\vec{x}) &= \bar{\rho}e^{-\sigma i\frac{\pi}{2}}, \\ \bar{\chi}_+(\vec{x}) &= \bar{\lambda}e^{+\sigma i\frac{\pi}{2}}, & \bar{\chi}_-(\vec{x}) &= \bar{\lambda}e^{-\sigma i\frac{\pi}{2}}, \end{aligned} \tag{6.11.2}$$

with $\sigma = (-1)^{x_1}$. Notice that the flux per plaquette $\mathcal{B}_{\text{plaquette}} = \pi$ but for the triangles is $\mathcal{B}_\Delta = +\frac{\pi}{2}$ for $\bar{\lambda} > 0$ and $\mathcal{B}_\Delta = -\frac{\pi}{2}$ for $\bar{\lambda} < 0$ (see Fig. 6.20). Thus, this state is chiral. At this level of approximation, the spinons behave like fermions moving on a frustrated lattice with the amplitudes listed in Eq. (6.11.2). Since the flux on the triangles is $\pm\pi/2$, some of the amplitudes must be complex no matter what gauge we choose. Thus, the effective one-particle Hamiltonian which controls the motion of spinons is complex (still Hermitean!). This means that time-reversal invariance (and parity) are broken. Since in this system we do not have any terms which explicitly break time-reversal invariance, what we are looking for is states with *spontaneously broken time-reversal invariance* and *parity*.

Using the notation of section 6.5 and the symbol Δ_i^+, $i = 1, 2$ for the lattice operator defined by

$$\Delta_i^+ f(\vec{x}) \equiv f(\vec{x} + \hat{e}_i) + f(\vec{x} - \hat{e}_i), \tag{6.11.3}$$

we can write down the equations of motion Eq. (6.5.13) including the effects of $\bar{\lambda}$. The new equations of motion are

$$\begin{aligned} i\dot{f}_\alpha^{(1)}(\vec{x}) =& -i\bar{\rho}\Delta_1 f_\alpha^{(2)}(\vec{x}) + i\bar{\rho}\Delta_2 f_\alpha^{(3)}(\vec{x}) + \\ & - i\bar{\lambda}\Delta_2^+ f_\alpha^{(4)}(\vec{x} + \hat{e}_1) - i\bar{\lambda}\Delta_2^+ f_\alpha^{(4)}(\vec{x} - \hat{e}_1), \\ i\dot{f}_\alpha^{(2)}(\vec{x} + \hat{e}_1) =& -i\bar{\rho}\Delta_1 f_\alpha^{(1)}(\vec{x} + \hat{e}_1) - i\bar{\rho}\Delta_2 f_\alpha^{(4)}(\vec{x} + \hat{e}_1) + \\ & + i\bar{\lambda}\Delta_2^+ f_\alpha^{(3)}(\vec{x} + 2\hat{e}_1) + i\bar{\lambda}\Delta_2^+ f_\alpha^{(3)}(\vec{x}), \\ i\dot{f}_\alpha^{(3)}(\vec{x} + \hat{e}_2) =& -i\bar{\rho}\Delta_1 f_\alpha^{(4)}(\vec{x} + \hat{e}_2) + i\bar{\rho}\Delta_2 f_\alpha^{(1)}(\vec{x} + \hat{e}_2) + \\ & - i\bar{\lambda}\Delta_2^+ f_\alpha^{(2)}(\vec{x} + \hat{e}_2 + \hat{e}_1) - i\bar{\lambda}\Delta_2^+ f_\alpha^{(2)}(\vec{x} + \hat{e}_2 - \hat{e}_1), \\ i\dot{f}_\alpha^{(4)}(\vec{x} + \hat{e}_1 + \hat{e}_2) =& -i\bar{\rho}\Delta_1 f_\alpha^{(3)}(\vec{x} + \hat{e}_1 + \hat{e}_2) - i\bar{\rho}\Delta_2 f_\alpha^{(2)}(\vec{x} + \hat{e}_1 + \hat{e}_2) + \\ & + i\bar{\lambda}\Delta_2^+ f_\alpha^{(1)}(\vec{x} + 2\hat{e}_1 + \hat{e}_2) + i\bar{\lambda}\Delta_2^+ f_\alpha^{(1)}(\vec{x} + \hat{e}_2). \end{aligned} \tag{6.11.4}$$

In Fourier space, Eq. (6.11.4) becomes

$$\begin{aligned} i\dot{f}_\alpha^{(1)}(\vec{p}) =& 2\bar{\rho}\sin p_1 f_\alpha^{(2)}(\vec{p}) - 2\bar{\rho}\sin p_2 f_\alpha^{(3)}(\vec{p}) - 4i\bar{\lambda}\cos p_1 \cos p_2 f^{(4)}(\vec{p}) \\ i\dot{f}_\alpha^{(2)}(\vec{p}) =& 2\bar{\rho}\sin p_1 f_\alpha^{(1)}(\vec{p}) + 2\bar{\rho}\sin p_2 f_\alpha^{(4)}(\vec{p}) + 4i\bar{\lambda}\cos p_1 \cos p_2 f^{(3)}(\vec{p}) \\ i\dot{f}_\alpha^{(3)}(\vec{p}) =& 2\bar{\rho}\sin p_1 f_\alpha^{(4)}(\vec{p}) - 2\bar{\rho}\sin p_2 f_\alpha^{(1)}(\vec{p}) - 4i\bar{\lambda}\cos p_1 \cos p_2 f^{(2)}(\vec{p}) \\ i\dot{f}_\alpha^{(4)}(\vec{p}) =& 2\bar{\rho}\sin p_1 f_\alpha^{(3)}(\vec{p}) + 2\bar{\rho}\sin p_2 f_\alpha^{(2)}(\vec{p}) + 4i\bar{\lambda}\cos p_1 \cos p_2 f^{(1)}(\vec{p}). \end{aligned} \tag{6.11.5}$$

As with Eqs. (6.5.17) and (6.5.18), we define the spinors $u_\alpha^{(a)}$ and $v_\alpha^{(a)}$ $(a = 1, 2)$

$$\begin{aligned} u_\alpha^{(1)}(\vec{p}) &= f_\alpha^{(1)}(\vec{p}) + f_\alpha^{(2)}(\vec{p}), \\ u_\alpha^{(2)}(\vec{p}) &= f_\alpha^{(3)}(\vec{p}) - f_\alpha^{(4)}(\vec{p}), \end{aligned} \tag{6.11.6}$$

and

$$\begin{aligned} v_\alpha^{(1)}(\vec{p}) &= f_\alpha^{(3)}(\vec{p}) + f_\alpha^{(4)}(\vec{p}), \\ v_\alpha^{(2)}(\vec{p}) &= f_\alpha^{(1)}(\vec{p}) - f_\alpha^{(2)}(\vec{p}). \end{aligned} \tag{6.11.7}$$

In matrix notation, we can now write (a,b=1,2)

$$i\dot{u}_\alpha^{(a)}(\vec{p}) = \left(2\bar{\rho}\sin p_1\sigma_3 - 2\bar{\rho}\sin p_2\sigma_1 - 4\bar{\lambda}\cos p_1\cos p_2\sigma_2\right)_{ab} u_\alpha^{(b)}(\vec{p}) \tag{6.11.8}$$

where σ_1, σ_2 and σ_3 are the Pauli matrices. The other spinor, $v_\alpha^{(a)}(\vec{p})$, obey the *same* equation. We can also write Eq. (6.11.9) in a Dirac form by defining the $\hat{\alpha}$ and γ-matrices through

$$\hat{\alpha}_1 \equiv \gamma_0\gamma_1 \equiv +\sigma_3, \quad \hat{\alpha}_2 \equiv \gamma_0\gamma_2 \equiv -\sigma_1, \quad \hat{\beta} \equiv \gamma_0 \equiv -\sigma_2. \tag{6.11.9}$$

In this notation, the equation of motion (6.11.9) takes the Dirac form

$$i\dot{u}_\alpha^{(a)}(\vec{p}) = \left(2\bar{\rho}\sin p_i\hat{\alpha}_i + 4\bar{\lambda}\cos p_1\cos p_2\hat{\beta}\right)_{ab} u_\alpha^{(b)}(\vec{p}) \tag{6.11.10}$$

Thus in the small momentum limit $|\vec{p}| \to 0$, we obtain the equation for *two* Dirac spinors, u_α and v_α, in the continuum with the *same* Fermi velocity $v_F = 2a_0\bar{\rho}$ and, what is more important, the *same effective mass* $m_c = \frac{\bar{\lambda}}{\bar{\rho}^2 a_0^2}$. Notice that *both* species u_α *and* v_α have the same *sign* of the effective mass m_c. This is a consequence of the fact that the one-particle Hamiltonian

$$H_{\text{chiral}}(\vec{p}) = 2\bar{\rho}\sin p_i\hat{\alpha}_i + 4\bar{\lambda}\cos p_1\cos p_2\hat{\beta} \tag{6.11.11}$$

is *complex* since all three Pauli matrices are present. This fact is, in turn, the result of the breaking of Time-Reversal invariance. We will see in chapter 7 that this result gives rise to a *Parity Anomaly* which greatly changes the behavior of the low-lying excitations. The eigenvalues of H_{chiral} are

$$\epsilon(\vec{p}) = \pm\sqrt{\bar{\rho}^4(\sin^2 k_1 + \sin^2 p_2) + 16\bar{\lambda}^2\cos^2 p_1\cos^2 p_2}. \tag{6.11.12}$$

This is what we found for the flux phase, Eq. (6.5.23), up to a mass term proportional to the next-nearest amplitude $\bar{\lambda}$. The two branches nearly touch at $(p_1, p_2) = (0, 0)$.

Thus far, we have not discussed energetics. Wen, Wilczek and Zee [Wen 89] have studied this problem with some detail. They found that as J' increases, the energy of the chiral state drops below that of the flux state and gets to be closed to the energy of the Dimer state. For the square lattice, even in the classically frustrated limit $J = J'$, it appears that the Dimer states are still preferred although not by much. Furthermore, at least in the large N-limit, the Néel states are not favored when $J \approx J'$. There is numerical evidence, based on the exact diagonalization of small clusters of up to 30 sites, that the Néel states are not favored for $J \approx J'$. In fact, at least for such small systems, the column states appear to be the ground states in this regime [Dagotto 89]. Thus, altough the chiral states are locally stable, they do not appear to be the global minimum of energy. But, it is quite conceivable to imagine slight modifications of the Hamiltonian which will drive the Mean-Field ground state energy of the chiral states down and make them a global minimum.

What appears to be more serious is the fact that the chiral Mean-Field theory has low-lying excited states, the spinons, which are not gauge invariant. The removal of gauge non-invariant states is likely to raise the energy of the ground state. We will come back to these issues in the next section.

Finally it is instructive to consider the effects of a Peirls gap, i.e. the gap which appears in the presence of a column state. This problem was studied by Dombre and Kotliar [Dombre 89]. Consider a column state of the type depicted in Fig. 6.7. There are four such states. With the choice of gauge, Eq. (6.11.2), the simplest case to consider has a column state with the "dimers" on the y-axis and the columns running along the x-axis. We can represent such a state by a *modulation* of the *amplitude* $\bar{\rho}_j(\vec{x})$ such that $\bar{\rho}_j(\vec{x})$ equals $\bar{\rho}+\delta\bar{\rho}$ if there is a dimer in the bond $(\vec{x},\vec{x}+\hat{e}_j)$ and equals $\bar{\rho}-\delta\bar{\rho}$ if there isn't a dimer in that bond. The next-nearest neighbor hopping terms have the same form as in Eq. (6.11.2). Thus, we can consider the competition between the Peierls state and the chiral state. We will see that, unlike the chiral state which breaks Parity and thus leads to a complex Hamiltonian, the Peierls state does not break Parity. If we assume that the selected Peierls state has the (vertical) dimers with their lower endpoints on *even* rows, the modified equations of motion are

$$\begin{aligned}
i\dot{f}^{(1)}_\alpha(\vec{x}) =& -i\bar{\rho}\Delta_1 f^{(2)}_\alpha(\vec{x}) + i\bar{\rho}\Delta_2 f^{(3)}_\alpha(\vec{x}) + \\
&+ i\delta\bar{\rho}\Delta_2^+ f^{(3)}_\alpha(\vec{x}) + \\
&- i\bar{\lambda}\Delta_2^+ f^{(4)}_\alpha(\vec{x}+\hat{e}_1) - i\bar{\lambda}\Delta_2^+ f^{(4)}_\alpha(\vec{x}-\hat{e}_1), \\
i\dot{f}^{(2)}_\alpha(\vec{x}+\hat{e}_1) =& -i\bar{\rho}\Delta_1 f^{(1)}_\alpha(\vec{x}+\hat{e}_1) - i\bar{\rho}\Delta_2 f^{(4)}_\alpha(\vec{x}+\hat{e}_1) + \\
&- i\delta\bar{\rho}\Delta_2^+ f^{(4)}_\alpha(\vec{x}+\hat{e}_1) + \\
&+ i\bar{\lambda}\Delta_2^+ f^{(3)}_\alpha(\vec{x}+2\hat{e}_1) + i\bar{\lambda}\Delta_2^+ f^{(3)}_\alpha(\vec{x}), \\
i\dot{f}^{(3)}_\alpha(\vec{x}+\hat{e}_2) =& -i\bar{\rho}\Delta_1 f^{(4)}_\alpha(\vec{x}+\hat{e}_2) + i\bar{\rho}\Delta_2 f^{(1)}_\alpha(\vec{x}+\hat{e}_2) + \\
&- i\delta\bar{\rho}\Delta_2^+ f^{(1)}_\alpha(\vec{x}+\hat{e}_2) + \\
&- i\bar{\lambda}\Delta_2^+ f^{(2)}_\alpha(\vec{x}+\hat{e}_2+\hat{e}_1) - i\bar{\lambda}\Delta_2^+ f^{(2)}_\alpha(\vec{x}+\hat{e}_2-\hat{e}_1), \\
i\dot{f}^{(4)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) =& -i\bar{\rho}\Delta_1 f^{(3)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) - i\bar{\rho}\Delta_2 f^{(2)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) + \\
&+ i\delta\bar{\rho}\Delta_2^+ f^{(2)}_\alpha(\vec{x}+\hat{e}_1+\hat{e}_2) + \\
&+ i\bar{\lambda}\Delta_2^+ f^{(1)}_\alpha(\vec{x}+2\hat{e}_1+\hat{e}_2) + i\bar{\lambda}\Delta_2^+ f^{(1)}_\alpha(\vec{x}+\hat{e}_2).
\end{aligned} \tag{6.11.13}$$

In terms of the spinors u_α and v_α of Eqs. (6.11.7) and (6.11.8), we get a modified mass term which is *different* for u_α and v_α. The one-particle Hamiltonian now is

$$H_{\text{Peierls}}(\vec{p}) = 2\bar{\rho}\sin p_i\hat{\alpha}_i + \left(4\bar{\lambda}\cos p_1\cos p_2 \pm 2\delta\bar{\rho}\cos p_2\right)\hat{\beta} \tag{6.11.14}$$

where the + (−) sign stands for the u_α (v_α) spionor. Thus, the low-energy spectrum still looks like two massive Dirac fermions which are propagating at the same speed but with different mass. What matters here is that the *sign* of the mass term depends on the relative strengths of $\delta\bar{\rho}$ and $\bar{\lambda}$. Indeed, for

$|\vec{p}| \to 0$, we find that $H_{\text{Peierls}}(\vec{p})$ takes the form

$$H_{\text{Peierls}}(\vec{p}) \approx 2\bar{\rho}\left(\hat{\alpha}_i p_i + (m_c \pm \delta m)\hat{\beta}\right) \tag{6.11.15}$$

where m_c is the chiral mass and δm is the splitting ($a_0 \equiv 1$)

$$m_c = \frac{\bar{\lambda}}{\bar{\rho}^2}, \qquad \delta m = \frac{\delta\bar{\rho}}{2\bar{\rho}^2}. \tag{6.11.16}$$

Hence, for $\delta m < m_c$, the u_α and v_α have different masses m_u and m_v ($m_u > m_v$, for $\delta\bar{\rho} > 0$) both with the *same* sign. Conversely, for $\delta m > m_c$, m_u and m_v not only are different but they have *opposite* sign. If $\bar{\lambda}$ is set to zero (i.e. no chiral state), there is a perfect symmetry. Thus, the *Peierls mass* does not lead to a complex Hamiltonian and consequently it does not break Parity. We will see in chapter 7 that the *relative sign* of the masses of the elementary excitations has very important consequences for the overall behaviour of the system in the generalized flux states. The eigenvalues of H_{Peierls} are

$$\epsilon(\vec{p}) = \pm\sqrt{4\bar{\rho}^2(\sin^2 p_1 + \sin^2 p_2) + (4\bar{\lambda}\cos p_1 \cos p_2 \pm 2\delta\bar{\rho}\cos p_2)^2}. \tag{6.11.17}$$

6.12 Fluctuations and Flux States

So far we have only considered flux states at the Mean-Field level and fluctuations have not been taken into account. We have already pointed out that this approach is not consistent since the fluctuations of the gauge fields, unlike the fluctuations of the amplitude, are completely out of control.

Consider first amplitude fluctuations around a flux phase with flux-π per plaquette. The Lagrangian density of Eq. (6.11.1) has degrees of freedom which, in addition to inducing both chiral and non-chiral mass terms in the low energy sector of the theory, can effectively drive the system into a highly anisotropic state, a *dimer state*. Since we are interested in understanding how do these different mechanisms compete with each other, it is convenient to parametrize the fluctuations of the bond lengths in such a way that these processes are most apparent. Thus, we are led to consider configurations in which the bond amplitudes vary slowly at the scale of the lattice spacing (I will refer to these processes as being uniform or unstaggered). In addition, there are fluctuations which vary rapidly at the scale of the lattice constant. These fast fluctuations induce scattering processes which mix different sublattices very strongly. We will refer to them as *staggered* amplitude fluctuations. Hence, the bond *amplitude* for the bond $(\vec{x}, \vec{x} + \hat{e}_j)$ has the form

$$\rho_j(\vec{x}) = \rho_j^u(\vec{x}) + \rho_j^s(\vec{x}) \tag{6.12.1}$$

where $\rho_j^u(\vec{x})$ is the unstaggered (or uniform) amplitude and $\rho_j^s(\vec{x})$ is the staggered amplitude. While $\rho_j^u(\vec{x})$ is slowly varying, $\rho_j^s(\vec{x})$ changes its sign from one bond to the next. Since we anticipate that the system may choose an

average uniform bond length $\bar{\rho}_j$ we write $\rho_j^u(\vec{x})$ and $\rho_j^s(\vec{x})$ in the form

$$\begin{aligned}\rho_j^u(\vec{x}) &= \bar{\rho}_j\left(1+\delta\rho_j^u(\vec{x})\right),\\ \rho_j^s(\vec{x}) &= \delta\rho_j^s(\vec{x}).\end{aligned} \tag{6.12.2}$$

Altough these amplitudes vary very slowly and over long wave lengths, they can be significantly different from each other. Thus the effective Dirac fermions may have different Fermi velocities along the x_1 and x_2 directions. What is more important, since these generalized Heisenberg models do not have any intrinsic length scale, apart from the lattice constant itself, there is an essential "softness" in the system which favors strong anisotropy. This can be clearly seen by writing down the spinon energy of such a state, which has the form for a non-chiral state (see Eq. (6.11.17))

$$\epsilon(\vec{p}) = \pm\sqrt{(2\rho_1^u \sin p_1)^2 + (2\rho_2^u \sin p_2)^2 + (2\delta\rho_2^s \cos p_2)^2}. \tag{6.12.3}$$

This energy can be made large and negative by setting

$$\delta\rho_1^u = -1, \quad |\delta\rho_2^s| = |\rho_2^u|, \tag{6.12.4}$$

which is the dimer limit. The symmetric amplitude $\delta\rho_j^u$ canr ot grow any larger than this without driving the total amplitude into negative values. Thus, this is the saturation limit. In this limit, the spin gap is infinitely large since all spinons are in singlet bond states one-lattice spacing long, the valence-bond states. The fluctuations of the gauge fields only cause dimer rearrangements as in our discussion of the Quantum-Dimer-Model. This phase does not break time reversal invariance.

The tendency to a collapse towards dimers can be suppressed by a suitable local modification of the Hamiltonian [Marston 89]. All that is needed is to have a scale $\bar{\rho}_0$ for the average bond amplitude around which they fluctuate. In the SU(N) model, this involves an interaction quartic in the spins. This possibility is not available for the case of interest, the nearest-neighbor spin-one-half Heisenberg model, but it may occur in further neighbor interactions. Let us assume for the moment that the dimer collapse has been avoided. Now the flux phase may be unstable against the development of both chiral and non-chiral mass terms. In turn, it is easy to write down an effective theory for the low energy modes. The effective Lagrangian density should include the (doubled) spinon modes (u_α and v_α). It should also contain both staggered amplitudes which, after normalization, can be denoted by two real Bose fields ϕ_1 and ϕ_2. The chiral modes are also bosonic and real and can be denoted by χ. The effective Lagrangian density should then have the form

$$\begin{aligned}\mathcal{L} = &(\bar{u}_\alpha i\gamma_\mu D_\mu u_\alpha + \bar{v}_\alpha i\gamma_\mu D_\mu v_\alpha) + \\ &- \phi_1(\bar{u}_\alpha v_\alpha + \bar{v}_\alpha u_\alpha) - \phi_2(\bar{u}_\alpha u_\alpha - \bar{v}_\alpha v_\alpha) - \frac{N}{J}U(\phi_1^2, \phi_2^2) + \\ &- \chi(\bar{u}_\alpha u_\alpha + \bar{v}_\alpha v_\alpha) - \frac{N}{J'}U'(\chi^2),\end{aligned} \tag{6.12.5}$$

where the potentials U and U' are *even* functions of ϕ_1, ϕ_2 and χ separately. The phases of the bond amplitudes, the gauge fields, have been included

through the covariant derivatives D_μ

$$D_\mu = \partial_\mu - i\mathcal{A}_\mu. \tag{6.12.6}$$

The potentials U and U' are assumed to have sharp minima at $\phi_1 = \phi_2 = \chi = 0$ and to diverge rapidly for large values of their arguments. This last condition is needed in order to avoid collapse towards a dimer state. The requirement that the potentials U and U' be even functions of their arguments, implies a four-fold degeneracy of the groundstate. In the absence of collapse , the symmetric amplitude modes, which represent local fluctuations of the length scale (i.e. Fermi velocity) and of anisotropy, do not change the qualitative physical properties of the system. The assumption that there is a well defined, and sharp, average bond amplitude $\bar{\rho}_0$ means that local dilatations and shears are strongly suppressed. When integrated out, these fluctuations only lead to effective interactions of the fermions which involve operators with many derivatives. In a renormalization group sense, such terms are irrelevant. This is equivalent to say, if the physics of the system is correctly described by the continuum model, operators with many derivatives may become important only if the fluctuations have large Fourier components at large values of the momentum. However, the main assumption of the continuum model is precisely that such Fourier components are small since only smooth configurations are correctely described by this model. Under these assumptions, the effective Lagrangian density of Eq. (6.12.5) is a good description of the physics of the system.

The fluctuations which are described in detail by Eq. (6.12.5) are the fluctuations of the gauge field $\mathcal{A}_\mu$ and of the amplitudes ϕ_i and χ. The fluctuations of the amplitudes ϕ_i and χ lead to a phase transition, in which one or several amplitudes have a non-zero expectation value, only if N is not too large. This can be checked by looking for a solutions of the saddle-point-equations. These equations, in the absence of a dimer solution, do not have a solution with $\langle\phi_0\rangle \neq 0$ (or $\langle\chi\rangle \neq 0$) unless N is smaller than some critical value N_c. The value of N_c depends on the details of the model, but it is typically large, $N_c \approx 40$. This regime is still described correctly by the $\frac{1}{N}$ expansion. Thus, unless one happens to be interested in unphysically large values of N, either $\langle\phi_1\rangle$, $\langle\phi_2\rangle$ or $\langle\chi\rangle$ will become non-zero. The fluctuations around this state are small and have very short correlation lengths.

From this discussion, we may conclude that, unless $N > N_c \gg 1$, there are spinons in the spectrum but they have a finite gap. For $N > N_c$, the spinons would be massless (i.e. no gap). Thus this model appears to predict the existence of electrically-neutral spin bearing excitations. However, this conclusion is not well founded since the fluctuations of the gauge field $\mathcal{A}_\mu$ have been ignored altogether.

What are the effects of he gauge fields $\mathcal{A}_\mu$? First of all, a simple inspection of the effective Lagrangian density, Eq. (6.12.5), shows that the gauge fields *only* appear in the kinetic energy term of the spinons, through the covariant derivatives. There is no separate term in this Lagrangian density which will control the fluctuations of the gauge field, such as $F_{\mu\nu}F^{\mu\nu}$ in electrodynamics.

Since the Lagrangian density is *linear* in $\mathcal{A}_\mu$, we can integrate the gauge fields out exactly. The integral over the $\mathcal{A}_\mu$'s yields

$$\int \mathcal{D}\mathcal{A}_\mu \; e^{iS[u,v,\phi_1,\phi_2,\chi,\; \mathcal{A}_\mu]} = e^{iS[u,v,\phi_1,\phi_2,\chi]} \int \mathcal{D}\mathcal{A}_\mu \; e^{\int d^3x \mathcal{A}_\mu J^\mu}, \qquad (6.12.7)$$

where J^μ is the total spinon gauge four-current density

$$J^\mu = \bar{u}_\alpha \gamma^\mu u_\alpha + \bar{v}_\alpha \gamma^\mu v_\alpha. \qquad (6.12.8)$$

The last factor in Eq. (6.12.4) shows that the integral over the $\mathcal{A}_\mu$'s is just a constraint

$$\int \mathcal{D}\mathcal{A}_\mu \; e^{\int d^3x \mathcal{A}_\mu J^\mu} = \prod_x \delta^3(J^\mu(\vec{x},t)). \qquad (6.12.9)$$

Hence, the only states allowed in the Hilbert space, let's call them $|\text{Phys}\rangle$, satisfy

$$J^\mu(\vec{x},t)|\text{Phys}\rangle = 0, \qquad (6.12.10)$$

which is a *local* condition. In components, this constraint is equivalent to the statement that the normal ordered spinon density $j_0(\vec{x},t)$

$$j_0(\vec{x},t) \equiv \rho(\vec{x},t) - \langle\rho(\vec{x},t)\rangle \qquad (6.12.11)$$

($\rho(\vec{x},t)$ being the electron density) and current $j_i(\vec{x},t)$ $i = 1,2$ annihilate the physical states. Thus, the condition of $\frac{N}{2}$ occupancy is exactly satisfied. However, this also means that the allowed states carry zero spinon current and that there are no states in the spectrum of this system carrying the spinon quantum numbers, i.e. spin one-half in the SU(2) case. As a result, these *spin-liquid phases* do not have *spinon* states in their spectra. This is not to say that the spinons do not have a role. Gauge-invariant spinon bound states do not have spinon quantum numbers and hence are allowed. In spin one-half language, these states are either spin singlets (valence bonds) or triplets. These bound states have large energy gaps with the singlets being the states of lowest energy.

6.13 SU(2) Gauge Invariance of Spin one-half Heisenberg Models

There is something peculiar in the way we have treated the spins so far. For the most part, the spin degrees of freedom are either "swallowed" by dimers or appear in an almost trivial factor as in the large-N limit (N being the number of spin degrees of freedom!). Nowhere in our discussion do we see not even a hint of the fact that the spins, say for $s = \frac{1}{2}$, have an SU(2) symmetry. The reason for this can be traced back to the way we decoupled the quartic interaction in terms of an *abelian* field χ_{ij} living on the links. In the past section, we showed that, for $N = 2$, there are two types of spinors, up and down, coupled to amplitudes and gauge fields. (In reality, there are four

because of the doubling.) It may seem that, if there are spinons in the excitation spectrum, then even without doubling there should be *four* elementary excitations bearing spin: spinon particles and holes of either spin orientation. The gauge fields, however, make sure that the constraint of single occupancy is strictly enforced. Thus, at each site, only two, not four, degrees of freedom are allowed, each allowed by the orientation of the spin. We must conclude that the particle and hole excitations of the spinons cannot possibly be independent degrees of freedom. We also know that, in the absence of holes, particle-hole symmetry is strictly respected. Hence, the natural conclusion is that the spinon hole with, say, spin down must be the same physical excitation as the spinon particle with spin up, and vice versa. It is clear, then, that a combination of particle-hole and spin symmetries is playing a fundamental role in these systems. The gauge symmetry must then be larger than the local U(1) symmetry implied by the $\frac{1}{N}$ expansion or, for that matter, by any RVB-like abelian decoupling of the Heisenberg interaction.

We will show now that a spin-$\frac{1}{2}$ Heisenberg antiferromagnet on any lattice and in any dimension, is equivalent to the strong coupling limit of an SU(2) gauge theory coupled to fermions [Zou 88] and [Dagotto 88].

Let $\vec{x}$ and $\vec{x}\,'$ be two sites of a lattice. The term of the Heisenberg Hamiltonian describing the antiferromagnetic coupling between spins at $\vec{x}$ and $\vec{x}\,'$ is

$$J\vec{S}(\vec{x})\cdot\vec{S}(\vec{x}\,'). \tag{6.13.1}$$

Once again, we will use a fermion description of the spins

$$\vec{S}(\vec{x}) = c^{\dagger}_{\alpha}(\vec{x})\vec{\tau}_{\alpha\beta}c_{\beta}(\vec{x}), \tag{6.13.2}$$

where $\vec{\tau}$ is the set of 2×2 Pauli matrices and we require single occupancy at $\vec{x}$ and $\vec{x}\,'$:

$$1 = c^{\dagger}_{\alpha}(\vec{x})c_{\alpha}(\vec{x}) = c^{\dagger}_{\alpha}(\vec{x}\,')c_{\alpha}(\vec{x}\,'). \tag{6.13.3}$$

Let us perform a particle-hole transformation at every sites so as to ensure that the reference state satisfies Eq. (6.13.3). We define new fermion operators $\psi_1(\vec{x})$ and $\psi_2(\vec{x})$ given by the relationships

$$\begin{aligned} c_{\uparrow}(\vec{x}) &= \psi_1(\vec{x}), \quad c^{\dagger}_{\uparrow}(\vec{x}) = \psi^{\dagger}_1(\vec{x}), \\ c_{\downarrow}(\vec{x}) &= \psi^{\dagger}_2(\vec{x}), \quad c^{\dagger}_{\downarrow}(\vec{x}) = \psi_2(\vec{x}). \end{aligned} \tag{6.13.4}$$

This canonical transformation amounts to an "exchange" of charge and spin operators since

$$\begin{aligned} c^{\dagger}_{\uparrow}(\vec{x})c_{\uparrow}(\vec{x}) + c^{\dagger}_{\downarrow}(\vec{x})c_{\downarrow}(\vec{x}) &= \psi^{\dagger}_1(\vec{x})\psi_1(\vec{x}) - \psi^{\dagger}_2(\vec{x})\psi_2(\vec{x}) + 1, \\ c^{\dagger}_{\uparrow}(\vec{x})c_{\uparrow}(\vec{x}) - c^{\dagger}_{\downarrow}(\vec{x})c_{\downarrow}(\vec{x}) &= \psi^{\dagger}_1(\vec{x})\psi_1(\vec{x}) + \psi^{\dagger}_2(\vec{x})\psi_2(\vec{x}) - 1. \end{aligned} \tag{6.13.5}$$

Hence, the constraint

$$c^{\dagger}_{\uparrow}(\vec{x})c_{\uparrow}(\vec{x}) + c^{\dagger}_{\downarrow}(\vec{x})c_{\downarrow}(\vec{x}) = 1 \tag{6.13.6}$$

is equivalent to

$$\psi_1^\dagger(\vec{x})\psi_1(\vec{x}) - \psi_2^\dagger(\vec{x})\psi_2(\vec{x}) = 0. \tag{6.13.7}$$

In other words, we are projecting onto the subspace with an equal number of quantum numbers 1 and 2 per sites. Such states are denoted by $|\text{Phys}\rangle$. The constraint, Eq. (6.13.7), has the equivalent form

$$\psi^\dagger(\vec{x})\tau_3\psi(\vec{x})|\text{Phys}\rangle = 0. \tag{6.13.8}$$

However, Eq. (6.13.7) implies that the following identities must also hold

$$\begin{aligned}\psi^\dagger(\vec{x})\tau_1\psi(\vec{x})|\text{Phys}\rangle &= \left(\psi_1^\dagger(\vec{x})\psi_2(\vec{x}) + \psi_2^\dagger(\vec{x})\psi_1(\vec{x})\right)|\text{Phys}\rangle = 0,\\ \psi^\dagger(\vec{x})\tau_2\psi(\vec{x})|\text{Phys}\rangle &= i\left(\psi_1^\dagger(\vec{x})\psi_2(\vec{x}) - \psi_2^\dagger(\vec{x})\psi_1(\vec{x})\right)|\text{Phys}\rangle = 0.\end{aligned} \tag{6.13.9}$$

Indeed, Eq. (6.13.9) is equivalent to the statements

$$\begin{aligned}\left(c_\uparrow^\dagger(\vec{x})c_\downarrow^\dagger(\vec{x}) + c_\downarrow(\vec{x})c_\uparrow(\vec{x})\right)|\text{Phys}\rangle = 0,\\ i\left(c_\uparrow^\dagger(\vec{x})c_\downarrow^\dagger(\vec{x}) - c_\downarrow(\vec{x})c_\uparrow(\vec{x})\right)|\text{Phys}\rangle = 0,\end{aligned} \tag{6.13.10}$$

which are true since the states $|\text{Phys}\rangle$ are singly occupied. Therefore, we have the *local* constraint on the space of *allowed* states

$$\psi^\dagger(\vec{x})\vec{\tau}\psi(\vec{x})|\text{Phys}\rangle = 0 \tag{6.13.11}$$

at each site of the lattice. Note, however that $\psi^\dagger(\vec{x})\vec{\tau}\psi(\vec{x})$ is *not* a spin operator. Rather, the spin operators $\hat{S}_a(\vec{x})$, $a = 1, 2, 3$, are given by

$$\begin{aligned}\hat{S}_1(\vec{x}) &\equiv c_\uparrow^\dagger(\vec{x})c_\downarrow(\vec{x}) + c_\downarrow^\dagger(\vec{x})c_\uparrow(\vec{x}) = \psi_1^\dagger(\vec{x})\psi_2^\dagger(\vec{x}) + \psi_2(\vec{x})\psi_1(\vec{x}),\\ \hat{S}_2(\vec{x}) &\equiv i\left(c_\uparrow^\dagger(\vec{x})c_\downarrow(\vec{x}) - c_\downarrow^\dagger(\vec{x})c_\uparrow(\vec{x})\right) = i\left(\psi_1^\dagger(\vec{x})\psi_2^\dagger(\vec{x}) - \psi_2(\vec{x})\psi_1(\vec{x})\right),\\ \hat{S}_3(\vec{x}) &\equiv c_\uparrow^\dagger(\vec{x})c_\uparrow(\vec{x}) - c_\downarrow^\dagger(\vec{x})c_\downarrow(\vec{x}) = \psi_1^\dagger(\vec{x})\psi_1(\vec{x}) + \psi_2^\dagger(\vec{x})\psi_2(\vec{x}) - 1.\end{aligned} \tag{6.13.12}$$

This set of operators have a remarkable local symmetry. Let $\psi'(\vec{x})$ be a new spinor related to $\psi(\vec{x})$ by means of an SU(2) transformation $U(\vec{x})$:

$$\psi'_\alpha(\vec{x}) = U_{\alpha\beta}(\vec{x})\psi_\beta(\vec{x}). \tag{6.13.13}$$

Clearly, under such a transformation, we have

$$\begin{aligned}\psi'^\dagger_\alpha(\vec{x})\tau^a_{\alpha\beta}\psi'_\beta(\vec{x}) &= \psi_\alpha^\dagger(\vec{x})\left(U^{-1}(\vec{x})\tau^a U(\vec{x})\right)_{\alpha\beta}\psi_\beta(\vec{x})\\ &\equiv R^{ab}(\vec{x})\psi_\alpha^\dagger(\vec{x})\tau^b_{\alpha\beta}\psi_\beta(\vec{x}),\end{aligned} \tag{6.13.14}$$

where $R(\vec{x})$ is the SO(3) rotation associated with the SU(2) transformation $U(\vec{x})$.

The spin operators $\hat{S}_a(\vec{x})$, $a = 1, 2, 3$, are *invariant* under this SU(2) transformation. Firstly, $\hat{S}_3(\vec{x})$ is clearly invariant:

$$\hat{S}_3(\vec{x}) = \psi_\alpha^\dagger(\vec{x})\psi_\alpha(\vec{x}) - 1 = \psi'^\dagger_\alpha(\vec{x})\psi'_\alpha(\vec{x}) - 1. \tag{6.13.15}$$

Secondly, the invariance of $\hat{S}_1(\vec{x})$ and $\hat{S}_2(\vec{x})$ follows from the fact that the operators $\chi(\vec{x})$ and $\chi^\dagger(\vec{x})$, defined by

$$\chi(\vec{x}) \equiv \frac{1}{2}\epsilon_{ij}\psi_i(\vec{x})\psi_j(\vec{x}) \tag{6.13.16}$$

are also inveriant under SU(2).

It is convenient to introduce the SU(2) invariant operator $M(\vec{x})$

$$M(\vec{x}) \equiv \psi^\dagger_\alpha(\vec{x})\psi_\alpha(\vec{x}). \tag{6.13.17}$$

It is easy to show now that the Heisenberg Hamiltonian on *any* lattice and in *any* dimension with a translationally invariant interaction $J(\vec{l})$ ($\vec{l}$ is the relative position vector of a pair of spins) is equivalent to the following Hamiltonian

$$H = -N_s\left(\sum_{\vec{l}} J(\vec{l})\right)(1+2m_3)+ \\ +\sum_{\vec{x},\vec{l}} J(\vec{l})\left(M(\vec{x})M(\vec{x}+\vec{l}) + 2\left(\chi^\dagger(\vec{x})\chi(\vec{x}+\vec{l}) + \chi^\dagger(\vec{x}+\vec{l})\chi(\vec{x})\right)\right), \tag{6.13.18}$$

where N_s is the total number of sites of on the lattice and m_3 is the *total* z polarization of the allowed Hilbert space

$$\frac{1}{N_s}\sum_{\vec{x}} \hat{S}_3(\vec{x})|\text{Phys}\rangle = m_3|\text{Phys}\rangle. \tag{6.13.19}$$

The Heisenberg Hamiltonian in the form given by Eq. (6.13.18) is manifestly invariant under the *local* SU(2) transformations of Eq. (6.13.13) since it is written in terms of $M(\vec{x})$, $\chi(\vec{x})$ and m_3 which are locally invariant.

It is important to stress that this local SU(2) symmetry, which involves *both* spin rotations *and* a particle-hole transformation, is unrelated to the *global* SU(2) invariance

$$c_\alpha(\vec{x}) \rightarrow c'_\alpha(\vec{x}) = V_{\alpha\beta}c_\beta(\vec{x}), \tag{6.13.20}$$

which induces *global* rotations of the spin polarization

$$\hat{S}_a(\vec{x}) \rightarrow \hat{S}'_a(\vec{x}) = R^{ab}\hat{S}_b(\vec{x}). \tag{6.13.21}$$

In section (2.3.1), we showed that the Heisenberg antiferromagnet is the $U \rightarrow \infty$ limit of a half-filled Hubbard model. I will now show that it is also the strong coupling limit of an SU(2) lattice gauge theory.

Consider a system of fermions, with creation and annihilation operators $\psi^\dagger_\alpha(\vec{x})$ and $\psi_\alpha(\vec{x})$ respectively, coupled to a set of SU(2) gauge degrees of freedom $U(\vec{x},\vec{x}\,')$ on the bonds $(\vec{x},\vec{x}\,')$ of a lattice. The Hilbert space of this system is a tensor product of fermionic states on the sites of the lattice multiplied by states on the links associated with gauge degrees of freedom. Let $A^a(\vec{x},\vec{x}\,')$ be an operator which transforms like a *vector* under SU(2), i.e. $\vec{A}(\vec{x},\vec{x}\,')\cdot\vec{\tau}$ is an element of the Lie algebra. Let us label the states on the

links by the (real) eigenvalues of $A^a(\vec{x},\vec{x}\,')$, e.g. $|\{A^a(\vec{x},\vec{x}\,')\}|$. The operators $U(\vec{x},\vec{x}\,')$ are 2×2 matrices defined by

$$U(\vec{x},\vec{x}\,') = e^{i\tau^a A^a(\vec{x},\vec{x}')}, \tag{6.13.22}$$

where the τ^a's are the generators of SU(2) in the fundamental (spinor) representation. Moreover, we demand

$$A^a(\vec{x},\vec{x}\,') = -A^a(\vec{x}\,',\vec{x}). \tag{6.13.23}$$

Equivalently, the $U(\vec{x},\vec{x}\,')$ operators must satisfy the condition

$$U(\vec{x},\vec{x}\,') = U^\dagger(\vec{x}\,',\vec{x}). \tag{6.13.24}$$

Let $E^a(\vec{x},\vec{x}\,')$ be a set of operators acting on this Hilbert space. We will require that these operators be canonically conjugate to the $A^a(\vec{x},\vec{x}\,')$ i.e.

$$[A^a(\vec{x},\vec{x}\,'), E^b(\vec{y},\vec{y}\,')] = i\delta^{ab}\cdot\delta_{\vec{x},\vec{y}}\delta_{\vec{x}',\vec{y}'}. \tag{6.13.25}$$

In addition, the operators $E^a(\vec{x},\vec{x}\,')$ satisfy the angular momentum algebra

$$[E^a(\vec{x},\vec{x}\,'), E^b(\vec{y},\vec{y}\,')] = i\epsilon^{abc}E^c(\vec{x},\vec{x}\,')\cdot\delta_{\vec{x},\vec{y}}\delta_{\vec{x}',\vec{y}'}. \tag{6.13.26}$$

In other words, the operators $E^a(\vec{x},\vec{x}\,')$ transform like group generators. Clearly, the operators $E^a(\vec{x},\vec{x}\,')$ and the SU(2) matrices $U(\vec{x},\vec{x}\,')$, satisfy the commutation relations

$$[E^a(\vec{x},\vec{x}\,'), U(\vec{y},\vec{y}\,')] = \tau^a U(\vec{x},\vec{x}\,')\cdot\delta_{\vec{x},\vec{y}}\delta_{\vec{x}',\vec{y}'}. \tag{6.13.27}$$

All the commutators so defined (Eqs. (6.13.25)-(6.13.27)) vanish if the operators act on the Hilbert spaces associated to different links.

Consider now the Hamiltonian $\tilde{H}$ acting on the Hilbert space of gauge invariant states:

$$\begin{aligned}\tilde{H} =& \frac{G}{2}\sum_{\substack{(\vec{x},\vec{x}')\\ a}} E^a(\vec{x},\vec{x}\,')E^a(\vec{x},\vec{x}\,')+\\ &+\frac{i}{2}\sum_{\substack{(\vec{x},\vec{x}')\\ \alpha\beta}}\left(\psi^\dagger_\alpha(\vec{x})U_{\alpha\beta}(\vec{x},\vec{x}\,')\psi_\beta(\vec{x}\,') - \text{ h.c.}\right),\end{aligned} \tag{6.13.28}$$

where G is a coupling constant and $(\vec{x},\vec{x}\,')$ are pairs of sites on an arbitrary lattice.

On a given lattice, the equivalence between the system described by the Hamiltonian $\tilde{H}$ and the Heisenberg model holds in the limit $G\to\infty$. The argument goes as follows. First we note that $\tilde{H}$ is invariant under time-independent local SU(2) gauge transformations

$$\begin{aligned}U_{\alpha\beta}(\vec{x},\vec{x}\,') =& W^{-1}_{\alpha\gamma}(\vec{x})U'_{\gamma\delta}(\vec{x},\vec{x}\,')W_{\delta\beta}(\vec{x}\,'),\\ \psi_\alpha(\vec{x}) =& W^{-1}_{\alpha\beta}(\vec{x})\psi'_\beta(\vec{x}).\end{aligned} \tag{6.13.29}$$

In the limit $G\to\infty$, the ground state of the system has a huge degeneracy. In fact, to leading order in an expansion in powers of $\frac{1}{G}$, the low lying states are the gauge singlets which satisfy

$$E^a(\vec{x},\vec{x}\,')E^a(\vec{x},\vec{x}\,')|\Psi\rangle = 0, \tag{6.13.30}$$

(all links) and obey the constraint

$$\mathcal{Q}^a(\vec{x})|\Psi\rangle \equiv \psi^\dagger_\alpha(\vec{x})\tau^a_{\alpha\beta}\psi_\beta(\vec{x})|\Psi\rangle = 0. \tag{6.13.31}$$

This last condition implies that at each site $\vec{x}$ either we have no fermion, $|0\rangle$ or a "baryon" state $|\chi\rangle = \chi^\dagger|0\rangle$. We now can carry out a degenerate perturbation theory exactly identical to the one we used to derive the Heisenberg model from the Hubbard model. The first excited state available, $|\Psi_{\rm exc}\rangle$, has a link excited to a state with angular momentum $\frac{1}{2}$

$$E^a(\vec{x},\vec{x}\,')E^a(\vec{x},\vec{x}\,')|\Psi_{\rm exc}\rangle = \frac{3}{4}|\Psi_{\rm exc}\rangle, \tag{6.13.32}$$

only on that link. The effective Heisenberg exchange interaction thus obtained is equal to $J = \frac{2}{3G}$.

What is the physical meaning of this symmetry? What we have actually shown is that the strong correlation limit of the Hubbard model **at half filling ** has an effective gauge invariance. This gauge invariance, which is a mixture of local particle-hole transformation and spin rotation, merely reflects the fact that in the strong correlation limit the only excitations left do not violate the local constraint. Hence no charge motion is possible and the system is an insulator. The charge-carrying states, either holes or doubly occupied sites, violate the constraint and pay a large energy penalty of order U, the Hubbard coupling constant. The remaining states are charge neutral states which may or may not carry spin. It is thus no surprise that the gauge theory not only satisfies the constraint

$$\mathcal{Q}^a(\vec{x})|{\rm Phys}\rangle = 0, \tag{6.13.33}$$

but also the related condition for the current

$$J^a(\vec{x})|{\rm Phys}\rangle = 0. \tag{6.13.34}$$

In other words, the current must also be zero. In a sense, we can think of the Heisenberg model as a "free particle" problem with its Hilbert space projected onto the states with zero current and zero charge, at the scale of the lattice spacing. The insulating phase of the Hubbard model, on the other hand, satisfies the same condition at length scales larger than the inverse of the charge gap. Thus, the low energy behaviour of the Hubbard insulator is also described by a system with a gauge symmetry. This property is clearly violated once one considers states with non-zero charge. Indeed the chemical potential, which couples to the charge density $c^\dagger_\alpha(\vec{x})c_\alpha(\vec{x})$ yields a term in the Hamiltonian $\tilde{H}$ of the form $\tilde{H}_{\rm charge}$

$$\tilde{H}_{\rm charge} = \mu\sum_{\vec{x}}\psi^\dagger(\vec{x})\tau_3\psi(\vec{x}), \tag{6.13.35}$$

which clearly violates the SU(2) symmetry. Similarly, the hopping term becomes

$$\tilde{H}_{\rm hop} = t\sum_{\langle\vec{x},\vec{x}'\rangle} c^\dagger_\sigma(\vec{x})c_\sigma(\vec{x}\,') = t\sum_{\langle\vec{x},\vec{x}'\rangle}\psi^\dagger_\alpha(\vec{x})\tau_3^{\alpha\beta}\psi_\beta(\vec{x}\,'), \tag{6.13.36}$$

which also violates the local SU(2) gauge invariance. We will come back to this issues later on. Let us point out now that this symmetry does imply that the spinon particle (hole) state with spin up is the same state as a spinor hole (particle) with spin down. Thus local SU(2) tells us that there are only two spinon states, as it should be.

References

[Affleck 88] I.K. Affleck and J.B. Marston. Phys. Rev. B 37, 3774 (1988).

[AKLT 88] I.K. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki. Commun. Math. Phys. 115, 477 (1988).

[Anderson 73] P. W. Anderson. Mater. Res. Bull. 8, 153 (1973).

[Anderson 87] P.W.Anderson. Science 235, 1196 (1987).

[Arovas 88] D. Arovas and A. Auerbach. Phys. Rev. B 38, 316 (1988).

[Banks 77] T. Banks, J. Kogut and R. Myerson. Nucl. Phys. B129, 423 (1977).

[Baskaran 87] Baskaran, Zou and Anderson . Solid State Comm. 63, 973 (1987).

[Baskaran 88] G. Baskaran and P.W. Anderson. Phys. Rev. B37, 580 (1988).

[Coleman 84] P. Coleman. Phys. Rev. B 29, 3055 (1984).

[Dagotto 88] E. Dagotto, E. Fradkin and A. Moreo. Phys. Rev. B 38, 2926 (1988).

[Dagotto 89] E. Dagotto and A. Moreo. Phys. Rev. Lett. 63, 2148 (1989).

[Dombre 89] T. Dombre and G. Kotliar. Phys. Rev. B 39, 855 (1989).

[Feynman 65] R.P. Feynman and A.R. Hibbs. *Quantum Mechanics and Path Integrals*, Mc-Graw Hill, 1965.

[Fisher 63] M. Fisher and J. Stephenson. Phy. Rev. 132, 1411 (1963).

[Fradkin 78] E. Fradkin and L. Susskind. Phys. Rev. D17 , 2637 (1978).

[Fradkin 88] E. Fradkin. In *Field theories in Condensed Matter Physics*, Procceedings of the Johns Hopkins Workshop on Field Theories in Condensed Matter Physics, Baltimore 1988, edited by Z. Tesanovic, Addison-Wesley, 1990.

[Fradkin 90] E. Fradkin and S. Kivelson. Mod. Phys. Lett. B4, 225 (1990).

[Holstein 40] T. Holstein and H. Primakoff. Phys. Rev. 58, 1098 (1940).

[Kalmeyer 88] V. Kalmeyer and R.B. Laughlin. Phys. Rev. Lett. 59, 2095 (1988).

[Kivelson 87] S. Kivelson, D. Rokhsar and J.P. Sethna. Phys. Rev. B 35, 865 (1987).

[Kane 89] C.L. Kane, P.A. Lee, T.K. Ng, B. Chakrabarty and N. Read. Phys. Rev. B 41, 2653 (1990).

[Kogut 79] J.B. Kogut. Rev. of Mod. Phys. 51, 659 (1979).

[Kogut 82] J. Kogut. In *Recent Advances in Field Theory and Statistical Mechanics*, Les Houches Summer School 1982, session XXXIX, edited by L.B. Zuber and R. Stora, North Holland, 1984.

[Kohmoto 88] M. Kohmoto and Y. Shapir. Phys. Rev. B 37, 9439 (1988).

[Kotliar 88] G.B. Kotliar. Phys. Rev. B 37, 3664 (1988).

[Leggett 74] A. Leggett. Rev. Mod. Phys. 47, 331 (1974).

[Liang 88] S. Liang, B. Doucot and P.W. Anderson. Phys. Rev. Lett. 61, 2585 (1988).

[Marston 89] J.B. Marston and I. Affleck. Phys. Rev. B 39, 11538 (1989).

[Polyakov 77] A.M. Polyakov. Nucl. Phys. B120, 429 (1977).

[Read 83] N. Read and D.M. Newns. J. Phys. C 16, 3273 (1983).

[Read 89] N. Read and S. Sachdev. Nucl. Phys. B 316, 609 (1989).

[Rokhsar 88] D. Rokhsar and S. Kivelson. Phys. Rev. Lett. 61, 2376 (1988).

[Ruckenstein 87] A. Ruckenstein, P. Hirschfeld and J. Appel. Phys. Rev. B 36, 857 (1987).

[Shraiman 89] B.I. Shraiman and E.D. Siggia. Phys. Rev. Lett. 62, 1564 (1989).

[Sutherland 88] B. Sutherland. Phys. Rev. B 37, 3786 (1988).

[Volovik 88] G. Volovik. Sov. Phys. JETP 67, 1804, (1988).

[Wen 89] X.G. Wen, F. Wilczeck and A. Zee. Phys. Rev B 39, 11413 (1989).

[Wiegmann 88] P.B. Wiegmann. Phys. Rev. Lett. 60, 821 (1988).

[Zou 88] I. Affleck, Z. Zou, T. Hsu and P.W. Anderson. Phys. Rev. B 38, 745 (1988).

CHAPTER 7

Chiral Spin States and Anyons

7.1 Fluctuations in Chiral Spin States and the Chern-Simons Gauge Theory

In chapter 6 we encountered a state, the Chiral Spin State (CSS), which spontaneously violates time reversal and parity invariance. We will see in this chapter that this feature of the CSS has far reaching novel consequences. There are other states of Condensed Matter in which time reversal invariance is broken. A ferromagnet has such a property. However, unlike the CSS, the ferromagnetic ground state does not violate parity, and its properties are quite different from what we will find in the CSS.

A system of electrons moving on a plane, in the presence of a perpendicular magnetic field, does not have time reversal invariance. It is explicitely broken by the magnetic field. If the electrons are spin polarized, in some sense parity is also broken due to the orbital nature of the coupling. The results are the fascinating properties of the Quantum Hall Effect (QHE), both in its integer and fractional forms. In this chapter and the coming ones, we will discuss the deep connections which exist between the Chiral Spin State and the Quantum Hall Effect. We will see that, as a result of the combined effects of violation of parity and time reversal invariance, both systems have an extra term, the so called Chern-Simons (C-S) term, in the effective Lagrangians for their low energy degrees of freedom. These Lagrangians also provide for a natural phenomenological description of the physics. In particular, both systems have low-energy excitations with fractional statistics or *anyons*. We will see in chapter 8 that, if the system is *compressible*, these excitations lead to a novel form of superconductivity called *anyon superconductivity*. Deep and far reaching connections between the CSS, the QHE, the mathematical theories of knots and, more generally, with Topological Field Theories will be described. We begin this chapter by going beyond the discussion of section 6.12 on the fluctuations around a Chiral Spin State.

Under what circumstances should we expect to get "free spinons", i.e.

Under what circumstances should we expect to get "free spinons", i.e. states with finite energy which carry spinon quantum numbers? The arguments at the end of the Chapter 6 show that this is not possible unless the fluctuations of the gauge fields somehow get suppressed. Terms of the $F_{\mu\nu}^2$ type do not efficiently suppress fluctuations. Gauge field mass terms are, on the other hand, very efficient in suppressing fluctuations. In $2+1$ space-time dimensions two gauge mass terms can arise. The simplest one, A_μ^2, explicitely breaks the gauge symmetry and can only arise if the system goes superconducting. This is a likely scenario at non-zero hole density but not possible at half-filling. In $2+1$ dimensions, there is another possible source of mass for the gauge fields: the topological or Chern-Simons mass terms [Schonfeld 81], [Jackiw 82]. Chern-Simons terms are locally gauge invariant but break parity P and time reversal T invariance. They have the form $\mathcal{L}_{\rm CS}$

$$\mathcal{L}_{\rm CS} = \frac{\theta}{4}\epsilon_{\mu\nu\lambda}A^\mu F^{\nu\lambda}. \tag{7.1.1}$$

The coupling constant θ is dimensionless and measures the strength of P and T (but not C) violations. We will see below that a term of this sort does arise from the fluctuations of the gauge field in a Chiral Spin State. Since the gauge fields now have a mass, one does expect to get spinon states in the spectrum. These states are massive, i.e. have a non-vanishing mass. We will also see in chapter 8 that, if holes are allowed, the system develops a novel form of superconductivity driven by excitations with fractional statistics called *anyons*.

In the abscence of mass terms for the fluctuations of the gauge fields the spinons disappear from the spectrum. The only low-lying excitations of the system are associated with the gauge field A_μ. It is then natural to ask for the effective Lagrangian which governs the dynamics of the gauge fields.

The $\frac{1}{N}$ expansion provides for a simple way to determine, not only the effective action of the A_μ's, but also that of the amplitudes ϕ_i and χ introduced in Eq. (6.12.5). This is done by first integrating out the spinon fields and later by expanding around one of the saddle-points of the resulting action. The effective action determined this way is

$$\begin{aligned} S_{\rm eff}[\phi_i,\chi,A_\mu] = \int d^3x \left(-\frac{N}{J}U(\phi_1^2,\phi_2^2) - \frac{N}{J'}U'(\chi^2)\right) + \\ - iN \ln\det\begin{pmatrix} (i\gamma_\mu D^\mu - \chi - \phi_2) & -\phi_1 \\ -\phi_1 & (i\gamma_\mu D^\mu - \chi + \phi_2) \end{pmatrix} \end{aligned} \tag{7.1.2}$$

where the 2×2 matrix in Eq. (7.1.2) occurs because of the spinon doubling in terms of u and v components of Eqs. (6.11.6) and (6.11.7). At the saddle point we have

$$\langle A_\mu\rangle = 0, \quad \langle\phi_i\rangle = \bar{\phi}_i, \quad \langle\chi\rangle = \bar{\chi}. \tag{7.1.3}$$

Let us now consider the effects of fluctuations around this state. Let $\tilde{\phi}_i$ and $\tilde{\chi}$ denote the fluctuation component of the amplitude fields. The vector potential A_μ has zero average, Eq. (7.1.3), and hence it represents a fluctuation. The fluctuations of the amplitude fields are massive and thus do not lead to any new physics provided, of course, that the saddle point represents a stable state. We will not consider the effects of such fluctuations here. Qualitatively, amplitude fluctuations are important in the dimer limit. We have already considered such effects in chapter 6.

The fluctuations of the vector potentials A_μ lead to interesting effects. Their effective action can be calculated by expanding S_{eff} of Eq. (7.1.2) in powers of A_μ. To second order, we get $S^{(2)}_{\text{gauge}}$ given by

$$S^{(2)}_{\text{gauge}} = \frac{1}{2}\int d^3x d^3y\, \Pi_{\mu\nu}(x,y)A^\mu(x)A^\nu(y) \tag{7.1.4}$$

where $\Pi_{\mu\nu}(x,y)$ is the one-particle irreducible fermion current-current correlation function (or Polarization)

$$\Pi_{\mu\nu} = \langle J_\mu(x)J_\nu(y)\rangle. \tag{7.1.5}$$

In momentum-space we can write $S^{(2)}_{\text{gauge}}$ in the form

$$S^{(2)}_{\text{gauge}} = i\frac{N}{2}\int \frac{d^3k}{(2\pi)^3}\int \frac{d^3q}{(2\pi)^3}\text{Tr}\left[\hat{S}(\frac{k}{2}+q)\gamma^\mu\hat{S}(-\frac{k}{2}+q)\gamma^\nu\right]A_\mu(k)A_\nu(-k) \tag{7.1.6}$$

in terms of the fermion propagator $\hat{S}(p)$

$$\hat{S}(p) = \frac{1}{p_\mu\gamma^\mu - \bar{\chi} - \bar{\phi}_i T_i} \tag{7.1.7}$$

where the 2×2 matrices T_1 and T_2 are given by the Pauli matrices σ_1 and σ_2 respectively.

Explicit computation of the operator $\Pi_{\mu\nu}$ (in momentum space) gives

$$\begin{aligned}\Pi^{\mu\nu}(k) &= \int \frac{d^3q}{(2\pi)^3}\, iN\text{Tr}\,[\hat{S}(q+\frac{k}{2})\gamma^\mu\hat{S}(q-\frac{k}{2})\gamma^\nu] \\ &= (k^2 g^{\mu\nu} - k^\mu k^\nu)\Pi_0(k^2) - i\epsilon^{\mu\nu\lambda}k_\lambda\Pi_A(k^2).\end{aligned} \tag{7.1.8}$$

The kernels $\Pi_0(k^2)$ and $\Pi_A(k^2)$ have the explicit form

$$\Pi_0(k^2) = \frac{N|m_+|}{4\pi k^2} + \\ + \frac{N}{8\pi\sqrt{k^2}}\left(\frac{4m_+^2}{k^2}+1\right)\sinh^{-1}\left[\frac{1}{\sqrt{\frac{4m_+^2}{k^2}-1}}\right] + (m_+ \leftrightarrow m_-),$$

$$\Pi_A(k^2) = -\frac{N}{2\pi\sqrt{k^2}}\sinh^{-1}\left[\frac{1}{\sqrt{\frac{4m_+^2}{k^2}-1}}\right] + (m_+ \leftrightarrow m_-), \tag{7.1.9}$$

where $\pm$ denotes the mass gaps (poles of Eq. (7.1.6)) for the two species of fermions

$$m_\pm = \bar{\chi} \pm \sqrt{\bar{\phi}_1^2 + \bar{\phi}_2^2} \tag{7.1.10}$$

including their *signs*.

These expressions can now be used to find the effective Lagrangian $\mathcal{L}^{(2)}_{\text{gauge}}$ which governs the dynamics of the "RVB" gauge field A_μ at low energies. By gauge invariance we know that only locally gauge invariant terms can possibly occur. To lowest order in a gradient expansion (i.e. in k^2, m^2) we expect a Maxwell-like term $F_{\mu\nu}F^{\mu\nu}$. However, in $2+1$ dimensions a Chern-Simons term (see Eq. (7.1.1)) is also possible. The C-S term although gauge-invariant, breaks parity (P) and time reversal (T). Thus, it may occur in a chiral spin state. Indeed, this is what actually does happen! By explicit calculation we find that the effective Lagrangian $\mathcal{L}^{(2)}_{\text{gauge}}$ does have the low-energy form

$$\mathcal{L}^{(2)}_{\text{gauge}} = -\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} + \frac{\theta}{4}\epsilon_{\mu\nu\lambda}A^\mu F^{\nu\lambda}. \tag{7.1.11}$$

The gauge coupling constant g^2 ("spinon charge") and C-S coupling θ are equal to

$$\frac{1}{g^2} = \frac{3N}{32\pi}\left(\frac{1}{|m_+|}+\frac{1}{|m_-|}\right) \tag{7.1.12}$$

and

$$\theta = \frac{N}{4\pi}\left(\text{sgn}(m_+)+\text{sgn}(m_-)\right). \tag{7.1.13}$$

Clearly, θ vanishes if $\text{sgn}(m_+) = -\text{sgn}(m_-)$. This is to be expected, since time reversal is not violated if the masses have opposite signs. This is the *non-chiral* spin liquid state. In the *chiral* state, $\text{sgn}(m_+) = \text{sgn}(m_-)$ and either sign, plus or minus do occur. Thus in a *chiral spin liquid* state we find that the C-S coupling is $\theta = \pm\frac{N}{2\pi}$ and it does not vanish.

We can gain some insight into the meaning of this result by considering the propagator of the gauge field. In particular, we want to know if there is a massless "photon" state in the spectrum. If such a state were to be present the

whole approach would be in doubt since in our problem the vector potentials A_μ would fluctuate wildly and, as we showed in chapter 6, the spinons would in fact be confined by the monopoles of the field A_μ. However, if the field A_μ were to become massive, the scenario would be completely different. Let us consider this question more closely. The propagator of the gauge fields $G_{\mu\nu}(x, x')$ is

$$G_{\mu\nu}(x, x') = \langle A_\mu(x) A_\nu(x') \rangle \tag{7.1.14}$$

and it is *not* gauge invariant. It only makes sense *after* a gauge is fixed. We do so by the standard procedure [Itzykson 80] of adding a gauge fixing term to the Lagrangian $\mathcal{L}^{(2)}_{\text{gauge}}$ of the form

$$\mathcal{L}^{(2)}_{\text{fixing}} = \frac{\alpha}{2g^2}(\partial_\mu A^\mu)^2. \tag{7.1.15}$$

In particular, I will work in the Lorentz gauge in which $\alpha \to \infty$ (i.e. $\partial^\mu A_\mu = 0$).

The propagator of the gauge fields, in the Lorentz gauge, is given by

$$G^{\mu\nu}(p) = \frac{g^2}{k^2 - g^4\theta^2}(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}) - \frac{g^4\theta i\epsilon^{\mu\nu\lambda}}{k^2(k^2 - g^4\theta^2)}k_\lambda. \tag{7.1.16}$$

This propagator has a *pole* at $k^2 - g^4\theta^2 = 0$. This "photon" state is massive and its mass m_γ is equal to $g^2|\theta|$. Hence a chiral state implies massive RVB gauge field. This mass term does not spoil gauge invariance and it does *not* imply the occurence of superconductivity. However, it is just as efficient in suppressing the fluctuations of the RVB gauge field. We have already discussed in sections 6.6 to 6.9 how the wild fluctuations of this gauge field, parametrized in terms of monopoles, are responsible for the confinement of excitations bearing the fundamental quantum number, the spin. Conversely, we are led to suspect that the presence of an induced Chern-Simons term may signal the liberation of the *spinons* by suppressing the monopoles. We saw that monopoles were responsible for disordering these loops, leading to confinement which, in the present context, means a valence bond crystal. However, the presence of the induced Chern-Simons term makes a significant difference.

Let us first discuss the fate of the monopoles. Consider a configuration $A^{(c)}_\mu$ which represents a set of monopoles (with their strings) and assume that they are well separated. A configuration of monopoles and anti-monopoles is generated by a set of sources in the form of infinitesimally thin solenoids joining each monopole to an anti-monopole. The issue here is the existence of a long range monopole field in the presence of the (induced) Chern-Simons term. But the C-S term causes the gauge field to be massive. In such a situation, an elementary study of the Euclidean equations of motion reveals that, for instance, in the case of a simple monopole-anti-monopole pair, the RVB magnetic field does not extend beyond a distance $r \sim \frac{1}{g^2\theta}$ away from the solenoid. Thus, the *dominant* contribution to the Euclidean action comes from this effective flux tube. If the linear size of the monopole-anti-monopole pair is R, the action S_{pair} grows linearly with R. Hence, monopoles and anti-monopoles

confined and their contribution to the expectation value of gauge-invariant operators is exponentatially small and can be neglected. More generally, the fermion path integral, in a chiral spin state, looses its gauge invariance in the presence of monopoles. The result is the supression of the monopoles and the deconfinement of the spinons. In consequence, in the CSS there is genuine separation of spin and charge [Fradkin 91].

7.2 The Statistics of the Spinons

What properties do the liberated spinons have? The best way to address this question is to look at how does the spinon propagate in this system. Consider the amplitude $W^{(1)}(\vec{x},\vec{x};T)$ for a spinon, of any type, created in the remote past at site $\vec{x}$, to propagate throughout the system to finally return to the same place $\vec{x}$, in the same state, in the remote future (i.e. $T \to \infty$). The (imaginary time) path-integral for this amplitude is

$$\begin{aligned}\lim_{T\to\infty} W^{(1)}(\vec{x},\vec{x};T) &= \lim_{T\to\infty} \mathrm{Tr}\; S_F(\vec{x},-\frac{T}{2};x,+\frac{T}{2}) \\ &= \int \mathcal{D}A_\mu \mathrm{Tr}\,\langle \vec{x},-\frac{T}{2} \mid \frac{1}{D_\mu\gamma^\mu+m} \mid \vec{x},T\rangle\; e^{-S_{\mathrm{eff}}(A)}\end{aligned} \tag{7.2.1}$$

where I have used the (imaginary-time) spinon-propagator in a fixed background configuration of RVB vector potentials. We can now use the Feynman picture of a sum over paths by first writing [Polyakov 87] the spinon propagator in the form

$$\begin{aligned}\mathrm{Tr}\,\langle \vec{x},-\frac{T}{2} \mid \frac{1}{D_\mu\gamma^\mu+m} \mid \vec{x},+\frac{T}{2}\rangle =& \mathrm{Tr}\,\langle \vec{x},-\frac{T}{2} \mid (-D_\mu\gamma^\mu+m) \mid Z\rangle \\ &\times \langle Z \mid \frac{1}{-D^2+m^2} \mid \vec{x},+\frac{T}{2}\rangle\end{aligned} \tag{7.2.2}$$

The proper-time representation of the propagator yields the following expression for the trace in Eq. (7.2.2)

$$\mathrm{Tr}\,\langle x,-\frac{T}{2} \mid (-D_\mu\gamma^\mu+m) \mid Z\rangle \int_0^\infty d\tau \langle Z \mid e^{+\tau D^2} \mid \vec{x},\frac{T}{2}\rangle e^{-\tau m^2}. \tag{7.2.3}$$

The operator D_μ is the Euclidean covariant derivative, $D_\mu = \nabla_\mu + iA_\mu$. The (Euclidean) Dirac matrices are only present in the prefactors.

Notice that by taking a trace we are effectively summing over all spinon polarizations. A standard path-integral argument now yields an expression for $W^{(1)}(\vec{x},\vec{x};T)$ in terms of sums over paths Γ of arbitrary length τ. The boundary conditions that we are using here imply that the sum over paths runs over contributions with paths which close on the imaginary time direction (i.e. run around the cylinder). The result is the path-integral

$$\langle Z \mid e^{+\tau D^2} \mid \vec{x},+\frac{T}{2}\rangle = \int \mathcal{D}x e^{-\int_0^\tau dt\left(\frac{1}{2}(\frac{d\vec{x}}{dt})^2+i\vec{A}\cdot\frac{d\vec{x}}{dt}\right)} \tag{7.2.4}$$

which is the sum over paths Γ of length τ satisfying the boundary condition $x(0) = Z$ and $x(\tau) = (\vec{x}, T)$. The amplitude $W^{(1)}(\vec{x}, \vec{x}; T)$ can now be written in the form

$$W^{(1)}(\vec{x}, \vec{x}; T) = \int \mathcal{D}A_\mu e^{-S_{\rm eff}(A_\mu)} \mathrm{Tr}\, \langle \vec{x}, -\frac{T}{2} \mid (-D_\mu \gamma^\mu + m) \mid Z \rangle \times \int_0^\infty d\tau e^{-\tau m^2} \langle Z \mid e^{+\tau D^2} \mid \vec{x}, +\frac{T}{2} \rangle. \tag{7.2.5}$$

Equivalently, $W^{(1)}(\vec{x}, \vec{x}; T)$ can be written in the form

$$\int_0^\infty d\tau e^{-\tau m^2} \int \mathcal{D}x_\mu \int \mathcal{D}A_\mu \ \mathrm{Tr}\, \langle \vec{x}, -\frac{T}{2} \mid (-D_\mu \gamma^\mu + m) \mid Z \rangle \times e^{-\int_0^\tau \mathrm{d}t \frac{1}{2} (\frac{d\vec{x}}{dt})^2} e^{i \oint_\Gamma A_\mu dx_\mu} \tag{7.2.6}$$

where I have used the fact that the paths Γ are closed and, consequently, the term $\int_0^\tau dt \vec{A} \cdot \frac{d\vec{x}}{dt}$ is simply the circulation of $\vec{A}$ around Γ. Notice that this quantity is a gauge invariant and it arises because we are considering paths which close around the cylinder. The path integral requires that this amplitude be averaged over all the configurations of the RVB gauge fields, for each path Γ. After we do that we get, using an obvious notation,

$$\langle W^{(1)} \rangle \sim \sum_\Gamma (\text{amplitude})_\Gamma \times \langle e^{i \oint_\Gamma dx_\mu A^\mu} \rangle \tag{7.2.7}$$

which involves the Wilson loop operator.

If we ignore the contribution of the monopoles, the amplitude $W^{(1)}$ can be estimated just by using the effective action of Eq. (7.1.10). The average in Eq. (7.2.7)

$$\langle e^{i \oint_\Gamma dx_\mu A^\mu} \rangle \tag{7.2.8}$$

can now be calculated quite easily. Let $J_\mu(\Gamma)$ be a current in $(2+1)$-dimensional Euclidean space defined by

$$J_\mu(\Gamma) = \begin{cases} S_\mu(x), & x \in \Gamma, \\ 0, & \text{otherwise}, \end{cases} \tag{7.2.9}$$

where $S_\mu(x)$ is the unit vector tangent to the path Γ at x. The average to be computed has the form

$$\langle e^{i \oint_\Gamma dx_\mu A^\mu} \rangle = \langle e^{i \int d^3x J_\mu(x) A^\mu(x)} \rangle \tag{7.2.10}$$

Since the effective action of Eq. (7.1.10) is quadratic in A_μ, the average, Eq.(7.2.8), is simply given by

$$\langle e^{i \oint_\Gamma dx_\mu A^\mu(x)} \rangle_{\rm CS} = e^{-\frac{1}{2} \int d^3x \int d^3x' J_\mu(x) G^{\mu\nu}(x-x') J_\nu(x')} \tag{7.2.11}$$

where the propagator $G_{\mu\nu}(x,x')$ has the Fourier transform given in Eq.(7.1.14). By direct substitution of Eq.(7.1.14), we get

$$\begin{aligned} G_{\mu\nu}(x,y) &= \int \frac{d^3p}{(2\pi)^3} G^{\mu\nu}(p) e^{ip\cdot(x-y)} \\ &= -g^2 \langle x \mid (\frac{1}{\partial^2 + g^4\theta^2})(g^{\mu\nu} - \frac{\partial^\mu \partial^\nu}{\partial^2}) \mid y\rangle + \\ &\quad + g^4\theta\epsilon^{\mu\nu\lambda}\langle x \mid \frac{1}{\partial^2(\partial^2+g^4\theta^2)}\partial_\lambda \mid y\rangle. \end{aligned} \tag{7.2.12}$$

Thus, the argument I of the exponential in the right hand side of Eq. (7.2.12) reads

$$\begin{aligned} I &\equiv -\frac{i}{2}\int d^3x \int d^3y J_\mu(x) G^{\mu\nu}(x-y) J_\nu(y) \\ &= -\frac{ig^2}{2}\int d^3x \int d^3y J_\mu(x) G_0(x,y;m^2) J_\mu(y) + \\ &\quad -\frac{i}{2}g^4\theta \int d^3x \int d^3y J_\mu(x)\epsilon^{\mu\nu\lambda}\langle x \mid \frac{1}{\partial^2(\partial^2+g^4\theta^2)}\partial_\lambda \mid y\rangle J_\nu(y). \end{aligned} \tag{7.2.13}$$

Here I used $G_0(x,y;m^2)$ as the propagator for a massive field with $m^2 = g^4\theta^2$, which obeys

$$(-\partial^2 - m^2) G_0(x,y;m^2) = \delta(x-y). \tag{7.2.14}$$

If we restrict our discussion to long loops only, we can make the long-distance approximation

$$G_0(x,y;m^2) \approx \frac{1}{m^2}\delta(x-y). \tag{7.2.15}$$

In this limit we get

$$I \approx -\frac{iL(\Gamma)}{2g^2\theta^2} + \frac{i}{2\theta}\int d^3x \int d^3y J_\mu(x)\epsilon^{\mu\nu\lambda} G_0(x,y;0)\partial_\lambda J_\nu(y) \tag{7.2.16}$$

where $L(\Gamma)$ is the perimeter of the loop Γ, $G_0(x,y;0)$ is the propagator of a *massless* scalar field and, in $2+1$ space-time dimensions, behaves like $G_0(x,y;0) \approx \mid x-y \mid^{-1}$. We will see below that this last non-local term plays a crucial role in this phase.

The first term in Eq. (7.2.16) embodies the quantum corrections to the propagation amplitude of the spinon. Hence it can be interpreted as a finite renormalization of its mass. The second term in Eq. (7.2.16) is more interesting. The quantity $R(\Gamma)$, given by

$$R(\Gamma) = \int d^3x \int d^3y J_\mu(x)\epsilon^{\mu\nu\lambda} G_0(x,y;0)\ \partial^y_\lambda J_\nu(y), \tag{7.2.17}$$

is in fact a *topological invariant*. After an integration by parts and using the difinition of the current $J^\mu(x)$, we can write $R(\Gamma)$ in the form

$$R(\Gamma) = \oint_\Gamma dx_\mu \oint_\Gamma dy_\nu \ \epsilon^{\mu\nu\lambda} \ \partial^x_\lambda G_0(x, y; 0). \tag{7.2.18}$$

We can make more sense of $R(\Gamma)$ by means of the following magnetostatic analogy. In order to make these ideas precise it is necessary to momentarily go to Euclidean space. Now $G_0(x, y; 0)$ is just the inverse Laplacian in three dimensions:

$$G_0(x, y; 0) \equiv \langle \vec{x} \mid \frac{1}{-\nabla^2} \mid \vec{y} \rangle. \tag{7.2.19}$$

Let us regard $J_\mu(x)$ as a current in three dimensions. This current establishes the static magnetostatic field $B_\mu(x)$ which satisfies

$$\vec{\nabla} \times \vec{B} = \vec{J}, \qquad \vec{\nabla} \cdot \vec{B} = 0, \tag{7.2.20}$$

i.e. Ampère's Law. Thus, we can solve for B_μ by means of the Green function $G_0(x, y; 0)$ in the form

$$B_\mu(x) = \int d^3y G_0(x, y; 0) \ \epsilon_{\mu\nu\lambda} \partial_\nu J_\lambda(y). \tag{7.2.21}$$

Thus, $R(\Gamma)$can be written in the more compact form

$$R(\Gamma) = \int d^3x J_\mu(x) B_\mu(x) \tag{7.2.22}$$

where B_μ is the field established by J_μ. This is a self-interaction effect. Now we can use the definition of J_μ and Stoke's theorem to get $R(\Gamma)$ in the form of a surface integral

$$R(\Gamma) = \oint_\Gamma dx_\mu B_\mu = \int_\Sigma d\sigma \ n_\mu \epsilon_{\mu\nu\lambda} \partial_\nu B_\lambda \tag{7.2.23}$$

where Σ is an open surface bounded by Γ. By substitution of Eq. (7.2.15) into Eq. (7.2.23) we get

$$R(\Gamma) = \int_\Sigma d\vec{\sigma} \cdot \vec{J} \tag{7.2.24}$$

i.e. $R(\Gamma)$ is the flux of $\vec{J}$, the current, through a surface bounded by itself (see Eq. (7.2.9)). Thus, at least qualitatively, $R(\Gamma)$ should be equal to the self-linking, or writhing, number of the path Γ which measures the number of times a vector normal to Γ winds as the loop is traced. Polyakov [Polyakov 88], who was the first to put these arguments forward, has argued that the *writhing* $R(\Gamma)$ of the path should be interpreted as an intrinsic *spin*. On the other hand, this spin only makes sense after a choice is made of a specific prescription for both measuring lengths along the path (i.e. choice of metric) and a short-distance regularization of the integrals involved in $R(\Gamma)$. In his seminal work relating the theory of knots and Chern-Simons gauge theories, Witten [Witten 89] has argued that these definitions depend on the choice

of regularization at short distances ("the framing of the knot"). In the problem that we are considering here, the Chern-Simons gauge theory (abelian in our case) appears as the effective theory at distances long compared with the inverse spinon gap of the mean-field-theory. It is unclear what regularization one should adopt in this case. It is conceivable that the anomalous spin predicted by Polyakov may or may not be present depending on the size of the spinon gap. This issue is still unresolved.

Let us consider the properties of spinons upon exchange processes. That is to say, we want to know what statistics they obey. Microscopically, we have *defined* the spinons to be fermions. The Chern-Simon term may change that. To see how that can happen, let us consider the propagation amplitude $W^{(2)}(\{\vec{x},\vec{y}\},\{\vec{x},\vec{y}\};T)$ for *two spinons*, which in the remote past where located at $\vec{x}$ and $\vec{y}$, to end up *either* at the same locations in the remote future $(T \to +\infty)$ *or* to exchange their positions. Once again, we will carry out the computation in the imaginary time formalism in which the time direction is periodic, i.e. the space-time has, at least, the topology of a cylinder. The two-particle amplitude will be represented as a sum over paths which close on the time direction. In principle, we will be dealing with two different paths Γ_1 and Γ_2 each representing the evolution of each spinon. These paths may or may not be linked. In other words, the paths are equivalent to *knots* or *braids*. We will see that the path integral can be written as a sum over classes of topologically inequivalent knots. Each class will be characterized by a phase factor. These phase factors can effectively alter the statistics of the spinons. The two-spinon amplitude $W^{(2)}$ has the form

$$W^{(2)} = \pm \sum_{\nu=0}^{\infty} W_\nu^{(2)} e^{i\phi_\nu} \tag{7.2.25}$$

where ν is the linking number of the paths, to be defined below. The $\pm$ sign represents the two possible processes, direct and exchange. We will primarely be interested in the computation of the phases ϕ_ν. The amplitudes $W_\nu^{(2)}$ are renomalizations of the spinon self energies, scattering amplitude, etc.

In terms of a sum over paths Γ, which are the union of the individual paths of the spinons, $W^{(2)}$ has the form

$$W^{(2)} = \pm \sum_{\Gamma \equiv \Gamma_1 \cup \Gamma_2} \mathcal{A}(\Gamma) \langle e^{i \oint_\Gamma dx_\mu A^\mu(x)} \rangle \tag{7.2.26}$$

where $\mathcal{A}(\Gamma)$ is the absolute value of the amplitude. After a little algebra we get

$$W^{(2)} = \pm \sum_{\Gamma} \mathcal{A}(\Gamma) e^{\frac{1}{2}\int d^3x \int d^3x' J_\mu(x) G_{\mu\nu}(x,x') J_\nu(x')} \tag{7.2.27}$$

where J_μ is the sum of the currents which define the paths Γ_1 and Γ_2, and $G_{\mu\nu}(x,x')$ is the analytic continuation to imaginary time of the propagator of the gauge fields. We will only be interested in the behavior of very large loops in the Euclidian space. The paths for direct and exchange processes become

closed by identifying their endpoints. Thus, exchange and direct processes have an extra *relative* linking number. It is this extra linking number which is responsible for the fractional statistics. It will be sufficient, for our purpose, to compute just this relative linking number. Thus, we can consider a simple direct process, in which the paths Γ_1 and Γ_2 are not linked, and simple exchange process in which the two paths are linked in such a way that they form a *single* path. Now, the *linking number* of a *single* path is its *writhing number* $R(\Gamma)$. However, there are no regularization ambiguities now, since the path winds around the cylinder exactly *once*.

The cylinder represents a topological obstruction and no redefinition of the metric on the path (for example by stretching it) can change this number. Thus exchange and direct process have a relative linking number of ± 1. The sign depends on the process by which we define exchange. If we define exchange by a counterclockwise (clockwise) rotation of one spinon around the other by an angle of π, followed by a translation equal to their relative separation, the sign is $+1(-1)$.

In the phase of the amplitude of the path integral the writhing number $R(\Gamma)$ enters multiplying a factor of $\frac{1}{2\theta}$. Hence, the *total* amplitude changes by a factor of $-e^{\pm\frac{i}{2\theta}}$ when two particles are exchanged, i.e.

$$W_d^{(2)} = -e^{\pm\frac{i}{2\theta}}\, W_e^{(2)}. \tag{7.2.28}$$

Eq. (7.2.28) implies that the spinons have *fractional statistics* with a *statistical angle* δ equal to

$$\delta = \frac{1}{2\theta} \equiv \frac{\pi}{N} \tag{7.2.29}$$

relative to the fermions. In particular, Eqs. (7.2.28) and (7.2.29) require that the two-spinon state should have a multivalued wave function [Wilczek 82]

$$\psi^{(2)}(1,2) = -e^{\pm i\delta}\psi(2,1). \tag{7.2.30}$$

For the case of physical interest, $N = 2$, the statistical angle $\delta = \frac{\pi}{2}$ and the wave function gets multiplied by $\pm i$ when two spinons are exchanged. Since this phase factor is exactly half-way between fermions ($-$sign) and bosons ($+$sign) these excitations have been dubbed *semions* [Laughlin 88]. In general, they are *anyons*, particles with fractional statistics [Wilczek 82].

We are only left to compute the phases ϕ_ν in Eq. (7.2.25) and (7.2.27). Let us write the phase ϕ_ν in the form

$$\phi_\nu = \frac{1}{2\theta} R(\Gamma_1, \Gamma_2). \tag{7.2.31}$$

Clearly, since $R(\Gamma_1, \Gamma_2)$ is bilinear in the currents, we can write $R(\Gamma_1, \Gamma_2)$ in terms of the writhing numbers of the individual paths and of the linking number $\nu = \bar{R}(\Gamma_1, \Gamma_2)$

$$R(\Gamma_1, \Gamma_2) = R(\Gamma_1) + R(\Gamma_2) + 2\bar{R}(\Gamma_1, \Gamma_2) \tag{7.2.32}$$

with

$$\bar{R}(\Gamma_1,\Gamma_2)=\frac{1}{2}\oint_{\Gamma_1}dx_\mu\oint_{\Gamma_2}dx'_\nu\ G_{\mu\nu}(x-x'). \tag{7.2.33}$$

We can now use the magnetostatic analogy once again. Let $J_\mu^{(1)}$ and $J_\mu^{(2)}$ be the two currents which establish the static fields $B_\mu^{(1)}$ and $B_\mu^{(2)}$ respectively. We get

$$\nu=\bar{R}(\Gamma_1,\Gamma_2)=\int d^3x J_\mu^{(1)}(x)B_\mu^{(2)}(x)\equiv\oint_{\Gamma_1}d\vec{x}\cdot\vec{B}^{(2)}(\vec{x}) \tag{7.2.34}$$

as a *circulation* of the field $\vec{B}^{(2)}$ (established by Γ_2) around Γ_1. Using now Stoke's theorem we write ν as the surface integral

$$\begin{aligned}\nu&=\oint_{\Gamma_1}d\vec{x}\cdot\vec{B}^{(2)}(\vec{x})\\&\equiv\int_{\Sigma_1}d\sigma\vec{n}\cdot\vec{\nabla}\times\vec{B}^{(2)}\\&=\int d\sigma\vec{n}\cdot\vec{J}\end{aligned} \tag{7.2.35}$$

where Σ_1 is an arbitrary surface bounded by Γ_1. Thus ν counts how many times does the loop Γ_2 wind around Γ_1. Putting it all together, we get a formula for the two spinons amplitude $W^{(2)}$, of the form

$$W^{(2)}=\sum_{\Gamma_1,\Gamma_2}\left(\mathcal{A}(\Gamma)\ e^{\frac{i}{2\theta}((R(\Gamma_1)+R(\Gamma_2))}\right)\ e^{\frac{i}{\theta}\nu(\Gamma_1,\Gamma_2)} \tag{7.2.36}$$

which, for an exchange process, picks up an additional factor $-e^{\pm i\delta}$. The quantity in bracket in Eq. (7.2.36) is a renormalized amplitude including possible an anomalous spin effect. It represents the total two-spinon amplitude in the topological sector with fixed linking number ν.

In the next sections we will find that the remarkable properties of the spinons in the CST are generically present for any system with anyons.

7.3 Flux States and the Fractional Hall Effect

In sections 6.5 we considered solutions to the saddle-point-equations, Eqs. (6.5.1) and (6.5.2), with a spontaneously generated *flux* of π per plaquette. The problem was shown to be equivalent, at the saddle point level, to a system of fermions moving in an *uniform average field* with a one-half flux quantum per plaquette. In section 6.10, we saw that a next-nearest-neighbor exchange coupling, which frustrates the system, effectively lowers the energy of the flux state. Furthermore, it drives the flux state into a chiral phase with spontaneously broken time reversal invariance. The flux phase has two bands that

become degenerate at four points of the Brillouin Zone. The chiral states have gaps at those points and the gaps grow larger as the frustrated regime $J \simeq J'$ is approached.

If the fluctuations around the mean field are ignored (at a first stage), a flux phase is then equivalent to two species (up and down spinons) of fermions moving in that flux. In the chiral phase we also have a gap which grows larger as frustration increases (i.e. for increasing J'/J). The one-particle spinon states can, in this limit, be approximated by the eigenstates of the lowest Landau level of a continuum problem in which the fermions move in a field with the same total flux. This approximation should be qualitatively correct provided that no level crossings occur. However, as we stressed previously, it is not possible to ignore the fluctuations around the mean field. Nevertheless, such an anology offers the possiblity of a new sort of spin liquid: a Laughlin state.

Laughlin states [Laughlin 83] represent condensed states of N fermions moving on a plane in the presence of an external magnetic field. These *incompressible* states, which have been shown to exhibit the *fractional Hall Effect*, represent a featureless liquid. It is tempting to speculate that the spin liquid states, which are also incompressible if there is a gap, may be described in terms of a Laughlin wave function, which we will discuss below.

Kalmeyer and Laughlin [Kalmeyer 87] have shown that, in the case of frustrated quantum spin systems, there is indeed a close analogy with the Hall Effect system except for the case that, here, we have bosons instead of fermions.

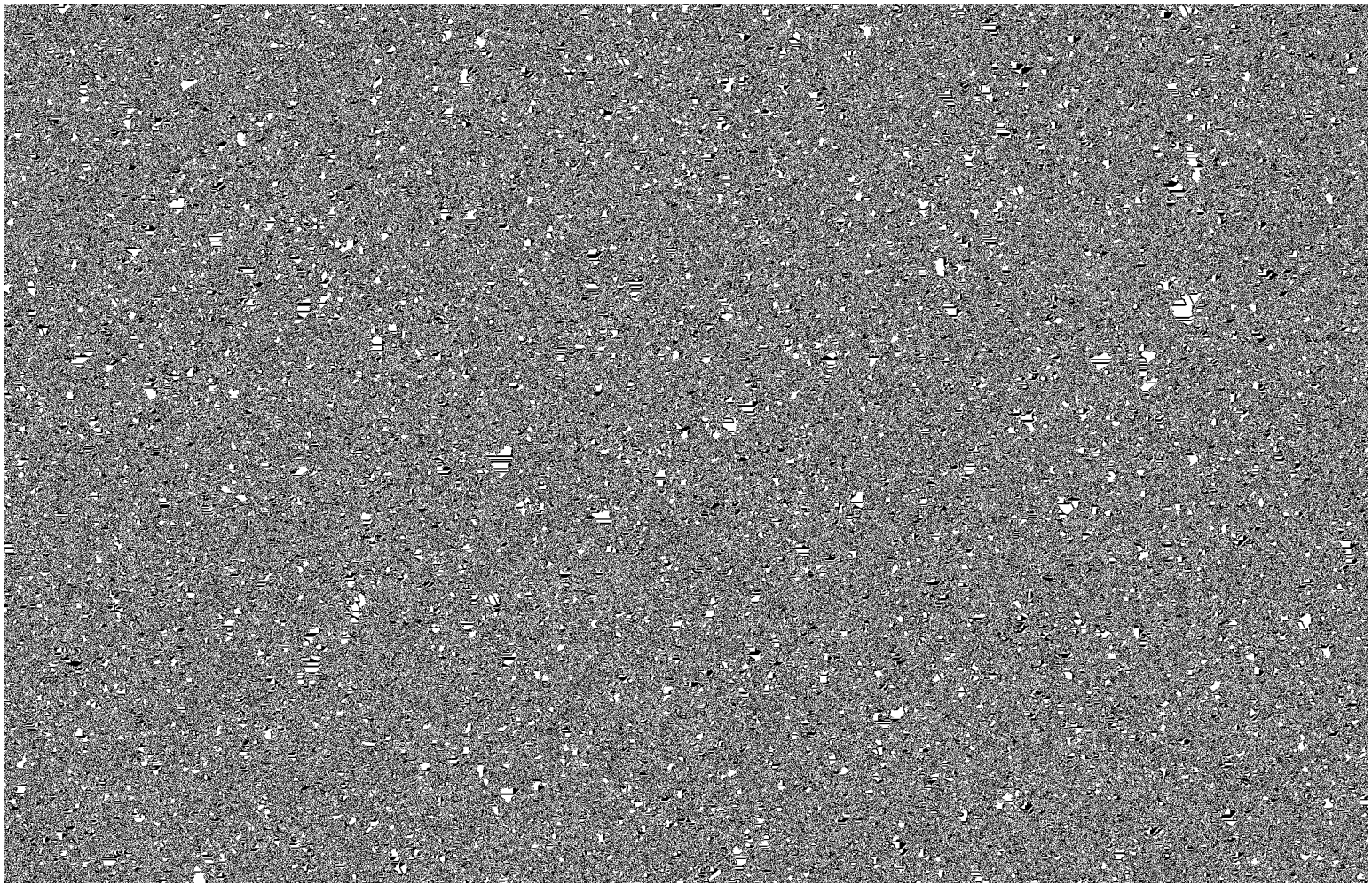
Let us discuss the Kalmeyer-Laughlin picture in more detail. Consider a frustrated quantum spin system, such as the square lattice with $J = J'$ or the triangular lattice. Let us assume that the Hamiltonian is still given by the usual Heisenberg exchange Hamiltonian. Instead of representing spins in terms of constituent bands of fermions, one can use *hard-core bosons* instead. This idea goes back to Holstein and Primakoff. Let $\mid F\rangle$ represent the ferromagnetic state which we will use as a *reference* state, not necessarily the ground state. Relative to $\mid F\rangle$, the raising operator $S^+(\vec{r})$ acts like a boson creation operator, a *spin-flip* being the boson. Since it is not possible to flip a spin twice, the bosons should have hard-cores: a site cannot be occupied by more than one boson. More formally we can write

$$\begin{aligned} S^+(\vec{r}) &= S_1(\vec{r}) + iS_2(\vec{r}) \equiv a^\dagger(\vec{r}) \\ S^-(\vec{r}) &= S_1(\vec{r}) - iS_2(\vec{r}) \equiv a(\vec{r}) \\ S_z(\vec{r}) &= \frac{1}{2} - a^\dagger(\vec{r})a(\vec{r}) \end{aligned} \tag{7.3.1}$$

where the operators a and $a^\dagger$ are bosons and, hence, satisfy the commutation relations

$$[a(\vec{r}), a^\dagger(\vec{r}\,')] = \delta_{\vec{r},\vec{r}'} \tag{7.3.2}$$

The Pauli spin algebra requires that these operators also satisfy a hard-core condition, $a^2 = (a^\dagger)^2 = 0$. Using these identities, it is now easy to write



(7.3.6)

per sign.
n the case of
π (mod 2π).
cture. More-
noving in an
per triangle.

s M moving
f a periodic
The bosons
he periodic
it, in which

alf of a flux
ie magnetic
oned above
quantum ϕ_0

nation that
e smoothly
he problem
short range
plane in the
Except for
be identical
of *fermions*
a magnetic
teractions.
sed a wave
es. It then
ntum Hall
aracterized
FQH case,
l magnetic
roken. We
blems have
a sense it

7.4 Fractional Statistics

One of the fundamental, and most cherished, axioms of local quantum field theory is the spin-statistics theorem. In the way that it is most commonly stated, it says that particles with integer (half-integer) spin are bosons (fermions) and that the corresponding second-quantized fields obey canonical equal time commutation (anticommutation) relations. At the root of this theorem is the need to preserve causality in a theory with local interactions as well as the requirement for the existence of a lowest energy state. Spin can only be integer or half-integer since the fields should transform like an irreducible representation of the Lorentz group in 3+1 dimensions: $SO(3,1)$. Even in a non-relativistic setting, the same requirements arise since the group of rotations $SO(3)$ is a subgroup of $SO(3,1)$. Furthermore, the many particle wavefunctions should be either symmetric or antisymmetric under the exchange of any pair of particles giving rise again to bosons and fermions. Thus, it may appear that these are the only possibilities.

The situation becomes radically different if the dimension of space-time is less than four. It has been known for a very long time [Jordan 28] that in one space dimension the statistics is essentially arbitrary. This is basically a kinematic effect. Fermions on a line cannot experience their statistics since they cannot get past each other and neither can bosons with hard cores. The Jordan-Wigner transformation, which we discussed in section 4.2, gives an explicit construction of a boson operator $a^\dagger(j)$ at the j^{th} site of a one dimensional lattice as a non-local function of fermion densities (see Eq. (4.2.23))

$$a^\dagger(j) = c^\dagger(j) e^{i\pi \sum_{m<j} c^\dagger(m)c(m)} \tag{7.4.1}$$

where the operators $c^\dagger(j)$ and $c(j)$ obey canonical anticommutation relations. In continuum quantum field theory, there exists an analogous construction known as *bosonization* (see section 4.3) , which yields a connection between a *canonical* Dirac Fermi field $\psi_\alpha(x)$ ($\alpha = 1,2$) and a *canonical* Bose field $\phi(x)$ in 1+1 dimensions given by the Mandelstam formula (see Eqs. (4.3.55-56))

$$\psi_\alpha(x) \approx e^{\frac{i}{\sqrt{\pi}} \int_{-\infty}^{x} dy \partial_0 \phi(y) \pm i\sqrt{\pi}\phi(x)} \tag{7.4.2}$$

with $\alpha = 1$ (2) for + (−).

Both constructions are based on the idea that in order to change the statistics one has to multiply an operator which creates a particle, such as $c^\dagger(j)$, times an operator which creates a *kink*, *i.e.*, a topological soliton. This idea, to some extent, can be generalized to higher dimensions. For instance, in 3+1 dimensions a *dyon*, a bound state of a charged bose particle and a Dirac magnetic monopole, behaves like a fermion. However, unlike the one-dimensional cases, all the examples in 3+1 dimensions are semi-classical in character. Furthermore, in one space dimensions, it is also possible to get *fractional* statistics (*i.e.*, intermediate between fermi and bose). A simple way

to do that is to change the exponent of the kink operator in the Jordan-Wigner formula by replacing π by an arbitrary angle δ. The resulting operators $a^\dagger(j)$ do not obey bose commutation relations but instead exhibit fractional statistics, *i.e.*,

$$a(j)a^\dagger(k) = \delta_{jk} - e^{i\delta}a^\dagger(k)a(j). \tag{7.4.3}$$

These operators, also known as parafermion operators, are generalizations of the fermion operators which are essential to the solution of the two-dimensional Ising model [Kadanoff 71]. They occur naturally in a number of quantum theories in 1+1 dimensions, such as the Gross-Neveu model, and in two-dimensional classical statistical mechanics. These operators have been found to play an important role in the critical behavior of the Clock Models in two dimensions, when studied using the methods of Conformal Field Theory [Dotsenko 84].

We will consider now the construction of anyon or parafermion operator more closely. This construction is due to Kadanoff and myself [Fradkin 80]. From the point of view of our discussion, the interest of this classical construction is that it has a natural generalization to 2+1 dimensions which has turned out to be quite useful. Consider a two-dimensional *classical statistical mechanics* model such as the $\mathcal{Z}_n$ Model on a square lattice. In the $\mathcal{Z}_n$ model, one defines an angle-like variable $\theta(\vec{r})$ residing at each site of a lattice. The angle $\theta(\vec{r})$ takes the discrete values $\theta = \frac{2\pi p}{n}$ at each site, where p and n are positive integers and $p = 1, \ldots, n$. The classical Hamiltonian H is chosen to be a local function of the angles $\theta(\vec{r})$ and invariant under *global* $\mathcal{Z}_n$ transformations $\theta(\vec{r}) \to \theta(\vec{r}) + \frac{2\pi m}{n}$, where m is a constant integer ($1 < m \leq n$). The partition function is

$$\mathcal{Z} = \sum_{\{\theta(\vec{r})\}} e^{\beta \sum_{\vec{r},\mu} \cos(\Delta_\mu \theta(\vec{r}))} \tag{7.4.4}$$

where β is the inverse temperature and $\mu = 1, 2$. In this particular case, the parafermion consists of an order operator $\mathcal{O}_m(\vec{r}) = \exp(i\frac{2\pi m}{n}\phi(\vec{r}))$. It measures the *order* at a site $\vec{r}$ of the lattice which is the endpoint of a *defect* or *domain wall* which flips the $\mathcal{Z}_n$ spins by a fixed angle $\frac{2\pi q}{m}$. This defect, which tries to create a *fractional vortex* of strength $\frac{2\pi q}{n}$, is most easily described by means of a gauge field $\mathcal{A}_j(\vec{x})$ defined on all the links of the square lattice. The $\mathcal{Z}_n$ spins and the gauge fields are minimally coupled through the covariant difference $\Delta_j\theta(\vec{r}) + \mathcal{A}_j(\vec{r})$. The vector potential can be chosen to have non-vanishing curl equal to $\frac{2\pi q}{n}$ on any arbitrary closed loop on the lattice which contains the site $\vec{R}$ on the dual lattice at the plaquette North-East of the site $\vec{r}$. A popular choice is to have $\mathcal{A}_j = 0$ except on a path on the dual lattice ending at $\vec{R}$ (a Dirac string). From this construction it is apparent that the fractional statistics of these operators results from a mechanism closely related to the Aharonov-Bohm effect.

It is now easy to check that the *correlation functions* of these operators are multivalued. Consider for instance the two-point function $\mathcal{G}_{pq}(\vec{r}, \vec{r}\,')$ which measures the correlations between operators $\mathcal{O}_p$,in the presence of defects of

strengths $\frac{\pm 2\pi q}{n}$, at sites $\vec{r}$ and $\vec{r}\,'$ respectively. Let $\mathcal{K}_q$ be the operator which creates a defect of strength $\frac{2\pi q}{n}$. Imagine carrying the site $\vec{r}\,'$ around site $\vec{r}$ on a closed loop Γ. After a full round trip, the spin operators have returned to their original locations but the Dirac strings are now misplaced: if in the original situation the spin at $\vec{r}$ was North of the string, now it is located South of it. The string can be returned to its original position by means of a gauge transformation. However the spin operator is not invariant under this operation. As a result, the correlation function picks up a phase of $\frac{4\pi}{n}pq$. Hence the *composite operator* $\Psi_{pq} = \mathcal{O}_p\mathcal{K}_q$ creates an excitation which is an anyon with statistical angle $\delta = \frac{2\pi}{n}pq$.

From the discussion outlined above it is apparent that any statistics is possible in one space dimension. Furthermore, the states created by operators which obey fractional statistics are, up to a boundary condition, completely determined by the coordinates of the particles on the line. In three dimensions, on the other hand, there does not seem to be room for *particles* with exotic statistics. However, a number of years ago, t'Hooft [t'Hooft 78] showed that there can be string-like states in four dimensional gauge theories which obey commutation relations with fractional statistics.

In two dimensions, however, one finds a very interesting situation. The Lorentz group for a two-dimensional system is $SO(2,1)$. The rotation group, which is crucial to both relativistic and non-relativistic systems, is $SO(2)$. This group has only one generator L_z, the generator of infinitesimal rotations in the plane, and hence it is abelian. Thus, all of its representations are one dimensional and labelled by the angular momentum quantum number ℓ. If the wave functions of the excitations are *required* to be single valued, the angular momentum ℓ can only be an integer. However *fractional* shifts of ℓ are also compatible with the algebra of $SO(2)$. States with fractional angular momentum have multivalued wavefunctions. In the Hilbert space which represents particles that move on the plane but are not allowed to sit on top of each other (a "punctured" plane) such wave functions are indeed allowed [Leinaas 77]. The plane becomes isomorphic to a Riemann surface punctured at the locations of the particles and different points are identified up to a phase determined by the fractional angular momentum. This framework provides for a natural construction of wave-functions which obey fractional statistics.

Wilczek [Wilczek 82] proposed the first fully quantum mechanical prescription to make such particles. He dubbed them *anyons*. Wilczek's model makes use of the Aharonov-Bohm effect experienced by a particle of charge q moving on the plane in the presence of a magnetic solenoid with flux ϕ perpendicular to the plane. More precisely, he assummed that each particle is *rigidly* bound to a solenoid which moves along with it. Consider now two such bound states. Let us make the *gedanken* experiment of adiabatically carrying one bound state around the other along some closed curve Γ. Because of the Aharonov-Bohm effect, the wavefunction Ψ of the bound state changes by an overall phase factor

$$\Psi \rightarrow e^{i\frac{qe}{\hbar c}\oint_\Gamma \vec{A}(\vec{x})\cdot\vec{x}}\ \Psi \tag{7.4.5}$$

where $\vec{A}$ is the vector potential associated with the magnetic flux of the solenoid. The angular momentum ℓ of the state is then equal to $\frac{qe}{hc}\phi$. If we denote with ϕ_0 the flux quantum, $\phi_0 = \frac{hc}{e}$, we can write the angular momentum ℓ in the form $\ell = q\alpha$, where $\alpha = \frac{\phi}{\phi_0}$. The angular momentum is not an integer if the Dirac quantization condition is not satisfied. The *statistics* obeyed by the bound states can be computed by considering an *exchange* process in which one bound state goes *half* its way around the other and, afterwards, both objects are shifted rigidly in such a way that they now have exchanged their initial positions. In this process their joint wave-function has picked up a phase factor exactly equal to *half* of what it is for a full round trip around the other particle, *i.e.*, $e^{i\pi\alpha}$. This definition is peculiar in the sense that the statistics of a state is determined by an adiabatic transport of the bound states in such a way that they never get on top of each other. Clockwise and counterclockwise processes yield complex conjugate phase factors. These wavefunctions are not representations of the permutation group. These states form representations of the Braid Group. These states are not defined in terms of the coordinates of the bound states alone. We have seen in section 7.2 that the amplitudes for the propagation of a pair of spinons in a CSL has precisely these properties. In that case, the fractional statistics was a consequence of the presence of an induced Chern-Simons term in the effective action for the low-energy degrees of freedom. We will see below that the Chern-Simons term is the most general local gauge invariant Lagrangian which binds particles and fluxes together. In chapter 10 we will see that the quasiholes of the Laughlin ground state for the Fractional Quantum Hall Effect have very similar properties.

7.5 Chern-Simons Gauge Theory: A Field Theory of Anyons

In order to make further progress we need a theory that will bind particles and fluxes together. Fluxes are most simply described as curls of a gauge field, which is usually called the statistical gauge field. Also, we want the particles to feel the fluxes through an Aharonov-Bohm mechanism. This means that the particles have to be minimally coupled to the statistical gauge fields through the covariant derivative. There is a problem with this approach. In most cases, a fluxoid which is electromagnetically coupled to a charged particle is not usually bound to it. The Aharonov-Bohm effect is not a bound-state problem. Rather, the amplitudes for the propagation of the particle get modified, by a phase factor, in the presence of flux. Thus, in the usual case, particles and fluxoids move quite independently from each other. In the problem that we are discussing, we want to *force* particles and fluxes to move *together*, as if they were the constituents of a bound-state. There is a theory which does all of that in a simple an straightforward way: the Chern-Simons gauge theory.

Let us imagine that we have a set of N particles. In a path-integral picture, the motion of the particles is described in terms of a set of trajectories $\Gamma = \Gamma_1 + \ldots + \Gamma_N$ with specified initial and final conditions. Quantum mechanics tells us that we have to sum over all possible trajectories weighing each history by the usual phase factor $\exp \frac{i}{\hbar} S(\gamma)$ in terms of the classical action of that particular history. If the particles have mass m, the classical action S_m of the particles is

$$S_m(\gamma) = \int_{t_i}^{t_f} dt \sum_{j=1}^{N} \left(\frac{1}{2} m (\frac{d\vec{x}_j}{dt})^2 + \frac{d\vec{x}_j}{dt} \cdot \vec{\mathcal{A}}(\vec{x}_j, t) - \mathcal{A}_0(\vec{x}_j, t) \right). \qquad (7.5.1)$$

The second term implies that the particle trajectories can also be regarded as a set of currents (and densities) $J_\mu = (J_0, \vec{J})$ $(\mu = 0, 1, 2)$ which are different from zero only on the trajectories of the particles and carry the unit of charge.

What should be the action for the statistical gauge fields? It cannot have the standard Maxwellian form since purely electrodynamical processes do not yield bound states of particles and fluxes. What is needed is a *constraint* which will *rigidly* bind particles and fluxes. There is only one *gauge invariant, local* expression which does the job: the Chern-Simons term

$$S_{\rm cs} = \int d^3x \frac{\theta}{4} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \mathcal{F}^{\nu\lambda}. \qquad (7.5.2)$$

The binding of particles to fluxes follows from the observation that the time component of the statistical vector potential $\mathcal{A}_0$ plays the role of a Lagrange multiplier field which enforces the local constraint

$$J_0 = \frac{\theta}{2} \epsilon_{ij} \mathcal{F}^{ij} \equiv \theta \mathcal{B}. \qquad (7.5.3)$$

This constraint simply means that a statistical flux of strength $\frac{1}{\theta}$ is present wherever there is a particle. In section 7.2 we saw that the presence of a Chern-Simons term modifies the two particle amplitudes in such a way that they exhibit fractional statistics. However, the Chern-Simons term does more than attaching particles to fluxes. It also determines the canonical structure of this system.

7.6 Anyons at Finite Density

In this section we consider a simple model which describes a gas of anyons at finite density. Since we are interested in systems in their thermodynamic limit, this theory is necessarily a field theory of anyons. The model that we will discuss is a system of "free" anyons on a square lattice (in space) with the topology of a torus. We choose to work on a spacial lattice both in order to avoid regularization problems and with an eye on applications to theories of high T_c superconductors. The time variable will remain continuous. This choice simplifies the formalism without any significant loss of generality. The

model can also be defined rigorously on a space-time lattice [Fröhlich 88]. The results have much wider applications than our derivation may suggest. For instance, as a byproduct, we will derive a Jordan-Wigner Transformation for systems in two space dimensions. This transformation is of great use for the study of two-dimensional quantum magnets. The theory can also be considered in the continuum, although some care has to be exercised at short distances. Chen, Wilczek, Witten and Halperin [Chen 89] have considered the continuum non-relativistic theory in great detail. In this section, I discuss the problem on a spacial lattice [Fradkin 89].

In the model that we consider, the anyons are free in the sense that the Hamiltonian only contains a nearest-neighbor hopping term. However, the anyons will be assumed to have *hard cores*. This last requirement is essential to the whole construction since otherwise the anyon worldlines can cross and the notion of braids falls apart.

Let us now show that the problem of a gas of N_a anyons with hard cores on a square lattice is equivalent to a gas of $N_f = N_a$ fermions, on the square lattice, coupled to a Chern-Simons gauge field defined on the links of that lattice. To be more precise, let $\hat{a}^\dagger(\vec{x})$ and $\hat{a}(\vec{x})$ be a set of anyon creation and annihilation operators defined on the sites $\{\vec{x}\}$ of the square lattice which satisfy the generalized equal-time commutation relations

$$\hat{a}(\vec{x})\hat{a}^\dagger(\vec{y}) = \delta_{\vec{x},\vec{y}} - e^{i\delta}\hat{a}^\dagger(\vec{y})\hat{a}(\vec{x}). \tag{7.6.1}$$

The angle δ indicates that we are dealing with fractional statistics. The choice of sign is such that for $\delta = 0$ we have fermions while for $\delta = \pi$ we have bosons. The hard core condition implies that, when acting on physical states, these operators obey

$$\hat{a}^\dagger(\vec{x})\hat{a}^\dagger(\vec{x}) = \hat{a}(\vec{x})\hat{a}(\vec{x}) = 0. \tag{7.6.2}$$

The second quantized Hamiltonian is simply given by

$$H = \sum_{\langle \vec{x},\vec{y} \rangle} \hat{a}^\dagger(\vec{x})\hat{a}(\vec{y}) + \text{h. c.} \tag{7.6.3}$$

where $\langle \vec{x}, \vec{y} \rangle$ are nearest neighboring sites on the square lattice.

Consider now a set of *fermion* creation and annihilation operators $\hat{c}^\dagger(\vec{x})$ and $\hat{c}(\vec{x})$ on the same square lattice. Let $\hat{A}_j(\vec{x})$ be a set of *boson* operators defined on the links of the lattice $\{(\vec{x}, \vec{x} + \hat{e}_j)\}$(with $j = 1, 2$) representing statistical gauge fields which satisfy the equal-time commutation relations

$$\left[\hat{A}_1(\vec{x}), \hat{A}_2(\vec{y})\right] = \frac{i}{\theta}\delta_{\vec{x},\vec{y}}. \tag{7.6.4}$$

Notice that, at every point $\vec{x}$ of the lattice, the component of the vector potential along the direction x_1 is the canonical pair of the component along the direction x_2. This choice may appear to be strange at first sight. In quantum mechanics, the canonical pairs are usually a coordinate and a momentum. The same is true in most field theories as well. However, there are a set of problems, both in field theory and in quantum mechanics, in which two "coordinates" form a canonical pair. The most common example is the case of

a charged particle moving on a plane in the presence of an external magnetic field. In this problem the guiding center coordinates for the position of a particle in the lowest Landau level satisfy an analogous set of commutation relations [Klauder 79].

The dynamics of the system is governed by the Hamiltonian

$$H_f = \sum_{\vec{x},j} \hat{c}^\dagger(\vec{x})\, e^{i\hat{A}_j(\vec{x})}\, \hat{c}(\vec{x}+\hat{e}_j) + \text{h.c.} \tag{7.6.5}$$

and the physical constraint states $\{|\text{Phys}\rangle\}$ are required to satisfy a local constraint ("Gauss' Law") between the fermion density $\rho(\vec{x})$ and the local magnetic flux $\hat{B}(\vec{x})$ of the statistical gauge fields

$$\left(\rho(\vec{x}) - \theta\hat{B}(\vec{x})\right)|\text{Phys}\rangle = 0. \tag{7.6.6}$$

This constraint implies that a fluxoid of strength $\frac{1}{\theta}$ is attached to each particle at the level of the lattice scale. The local statistical flux $\hat{B}(\vec{x})$ is given by the usual formula

$$\hat{B}(\vec{x}) = \Delta_1\hat{A}_2(\vec{x}) - \Delta_2\hat{A}_1(\vec{x}) \tag{7.6.7}$$

where Δ_j is the finite difference operator in the direction j. The flux thus defined effectively exists only on the dual lattice. This formulation has the additional advantage that the particles are not allowed to get "inside" the flux. The Hamiltonian H, together with the constraint and the commutation relations, follow from canonical quantization, in the gauge $A_0 = 0$, of the Lagrangian density $\mathcal{L}$

$$\mathcal{L} = c^\dagger(x)\,(i\partial_0 + A_0 + \mu)\,c(x) - \mathcal{H}(c^\dagger, c, \vec{A}) - \mathcal{L}_{\text{CS}}. \tag{7.6.8}$$

Here $\mathcal{H}$ is the Hamiltonian per site, μ the chemical potential, $x = (\vec{x}, t)$ and $\mathcal{L}_{\text{CS}}$ is the Chern-Simons Lagrangian density which, in terms of the vector potential A_μ and the field strength tensor $F_{\mu\nu}$ suitably defined on a lattice, has the form

$$\mathcal{L}_{\text{CS}} = \frac{\theta}{4}\epsilon^{\mu\nu\lambda}A_\mu F_{\nu\lambda}. \tag{7.6.9}$$

This Lagrangian density is explicitly invariant under local, time-dependent, gauge transformations, provided that the system does not have boundaries. The Chern-Simons Lagrangian density is not invariant at the boundary. The problem can be put on a lattice with boundaries by adding a potential energy term $V(\vec{r})$ to the Hamiltonian which rises very rapidly at the physical location of the boundaries. No particles would go into the forbidden region and, since the flux is rigidly tied to the particle density, the statistical flux would be zero in that region.

Let us apply the canonical quantization procedure to this system. As is the case in all gauge theories, the component A_0 of the Chern-Simons gauge field does not have a canonical pair since $\mathcal{L}$ does not depend on $\partial_0 A_0$. Thus, A_0 is a local Lagrange multiplier which is just enforcing the constraint, Eq.(7.6.6). The combination of fields which appears multiplying A_0 defines the generator $\hat{Q}$ for local time-independent gauge tranformations. It is easy to check that,

for this problem, $\hat{Q} = \rho(\vec{x}) - \theta\hat{B}(\vec{x})$. Thus, the constraint amounts to the requirement that the physical states $|\text{Phys}\rangle$ be gauge invariant. It is then customary to set $A_0 = 0$ and to demand that the constraint be satisfied locally.

The space components of the gauge field do have canonical pairs. The canonical momentum $\Pi_1(\vec{x})$ conjugate to $A_1(\vec{x})$ is given by

$$\Pi_1(\vec{x}) \equiv \frac{\delta\mathcal{L}}{\delta\partial_0 A_1(\vec{x})} = \theta A_2(\vec{x}). \tag{7.6.10}$$

Thus, the canonical commutation relations

$$[A_1(\vec{x}), \Pi_1(\vec{y})] = i\delta_{\vec{x},\vec{y}} \tag{7.6.11}$$

are equivalent to the commutation relations among the space components of the gauge field

$$\left[\hat{A}_1(\vec{x}), \hat{A}_2(\vec{y})\right] = \frac{i}{\theta}\delta_{\vec{x},\vec{y}}. \tag{7.6.12}$$

The equivalence between the anyon Hamiltonian and the Chern-Simons gauge theory coupled to fermions is established by solving the constraint, Eq. (7.6.6), which relates the local flux to the local density. This can be accomplished by fixing the remaining invariance under local time-independent gauge transformations. We will choose the *Coulomb* or *anyon* gauge $\vec{\nabla}\cdot\vec{A}(\vec{x}) = 0$. The statistical vector potential $\vec{A}(\vec{x})$ which is the solution of the constraint in this gauge is an explicit function only of the local particle density. Thus it may appear that there are no gauge degrees of freedom left. This however is not generally the case. Whether or not there are any gauge degrees of freedom left depends on the boundary conditions. On a torus, there are global gauge degrees of freedom which are not affected by the local fixing of the gauge.
We now have to solve the constraint, Eq. (7.6.6), for a lattice with the topology of a torus. Let L_1 and L_2 be the linear dimensions of the lattice along directions 1 and 2 respectively. It is impossible to eliminate all the gauge degrees of freedom by solving the constraint equation no matter what gauge is chosen unless *large* gauge transformations, which wrap around the torus along directions 1 or 2, are included. Consider the circulation of the statistical vector potential on a *non-contractible* closed loop wrapping around the torus along one of its large circles $\mathcal{C}_j$ $(j = 1, 2)$. Any *local* time-independent gauge transformation shifts the spacial components of the vector potential A_k by the gradient of a smooth function Λ of the coordinates $A_k(\vec{x},t) \to A_k(\vec{x},t) + \Delta_k\Lambda(\vec{x})$. Thus, the circulation Γ_j, with $\Gamma_j = \oint_{\mathcal{C}_j} d\vec{x}\cdot\vec{A}(\vec{x})$, is unchanged since Λ is a smooth and single-valued function of $\vec{x}$. Notice that this is the case even in the abscence of fermions! Thus, the circulations Γ_j, or *non-integrable phases*, are *global* degrees of freedom of the gauge field. A consistent treatment of this problem must take into account their dynamics.

There is a simple way to take care of both global and local gauge degrees of freedom. The local gauge degrees of freedom are non-local functions of the local particle density $\rho(\vec{x},t)$ given by the solution of the local constraint equation in some particular gauge. The global degrees of freedom are the

non-integrable phases Γ_j. To make any further progress it is necessary to fix the gauge. At the level of the functional integral, we first observe that the component A_0 of the statistical gauge field can always be integrated out giving rise to the local constraint at all times. We next write the spacial components of the statistical vector potential A_j in the form

$$A_j(x) = \mathcal{A}_j(x) + \bar{A}_j(x) \tag{7.6.13}$$

where $\mathcal{A}_j$ is a particular solution of the constraint equation and $\bar{A}_j$ generates the non-integrable phases which are solutions to the homogeneous constraint equation (*i.e.*, without fermions). We can completely determine all of these fields by choosing a particular gauge.

Let us consider first the local gauge degrees of freedom. In the Coulomb gauge, the inhomogeneous solution for the constraint equation is given in terms of the scalar field $\Phi(\vec{r})$:

$$\mathcal{A}_j(\vec{x}) = \epsilon_{jk}\Delta_k\Phi(\vec{r}) \tag{7.6.14}$$

where $\vec{r}$ are the sites of the dual lattice [Fradkin 89]. Here the scalar field Φ is the solution to the equation (see Eqs. (7.6.6-7))

$$\Delta^2\Phi(\vec{r}) = -\frac{1}{\theta}\rho(\vec{x}) \tag{7.6.15}$$

where $\vec{r}$ is the site on the dual lattice located northeast of the site $\vec{x}$ on the direct lattice and Δ^2 is the lattice Laplacian.

In this approach, fluxes are on the dual lattice while particles are on the direct lattice. Particles and fluxes never sit on top of each other and we have no ambiguities. On the other hand we could have chosen to put the flux southeast of the particle, or some other similar prescription. These different prescriptions are related to the possible existence of a self-linking number and an anomalous spin. We will not explore these issues any further. Let us simply note that this lattice regularization provides a natural way to separate particles and fluxes while keeping all the relevant symmetries intact. Also note the close analogy with the order-disorder operator construction for two-dimensional classical statistical mechanical systems. This feature is also present in the two dimensional Ising model and it reflects the fact that the Onsager fermions are two-component spinors [Kadanoff 71].

We now use the lattice Green function $G(\vec{r}, \vec{r}\,')$, which is the solution of the partial difference equation

$$\Delta^2_{\vec{r}}\, G(\vec{r}, \vec{r}\,') = \delta_{\vec{r},\vec{r}\,'} - \frac{1}{L_1 L_2}. \tag{7.6.16}$$

The last term of this equation, while unimportant in the thermodynamic limit, is necessary in order to define the Green function in a finite system without boundaries, no matter how large it is. The solution for the scalar field has the form

$$\Phi(\vec{r}) = \frac{1}{\theta}\sum_{\vec{r}'} G(\vec{r}, \vec{r}\,')\rho(\vec{x}\,'). \tag{7.6.17}$$

Thus, by inserting Eq. (7.6.17) into Eq. (7.6.14), we can write the vector potentials $\mathcal{A}_j$ in the form

$$\mathcal{A}_j(\vec{x}) = \frac{1}{\theta}\epsilon_{jk}\Delta_k \sum_{\vec{r}'} G(\vec{r},\vec{r}\,')\rho(\vec{x}\,'). \tag{7.6.18}$$

Let us *define* the multivalued function $\Theta(\vec{x},\vec{r}\,')$ as the solution for the lattice version of the Cauchy-Riemann equation

$$-\Delta_j G(\vec{r},\vec{r}\,') = \epsilon_{jk}\Delta_k \Theta(\vec{x},\vec{r}\,'). \tag{7.6.19}$$

The function $\Theta(\vec{x},\vec{r}\,')$ is found by integrating the Cauchy-Riemann equation along a path $\Gamma(\vec{x},\vec{x}\,')$, on the direct lattice, going from $\vec{x}$ to $\vec{x}\,'$ which leaves the point $\vec{r}$ to its left. For a finite system, the function $\Theta(\vec{x},\vec{r}\,')$ obtained by this procedure is *path dependent*. Moreover, along a closed path Γ on the direct lattice, which has the point $\vec{r}$ of the dual lattice in its interior region, the function Θ has a discontinuity $(\Delta\Theta)_\Gamma$. We can compute this discontinuity by using the Cauchy-Riemann equation

$$(\Delta\Theta)_\Gamma = \sum_\Gamma s_j(\Gamma)\Delta_j\Theta = \sum_\Gamma s_j(\Gamma)\epsilon_{jk}\Delta_k G \tag{7.6.20}$$

where $s_j(\Gamma)$ is a vector field which is equal to one on the path Γ and zero everywhere else. The last "line integral" in this equation can be computed by first using a discrete version of Gauss' theorem and then inserting Eq. (7.6.16) to yield

$$(\Delta\Theta)_\Gamma = \sum_{\bar{\Gamma}} \Delta^2 G = 1 - \frac{\mathcal{A}(\bar{\Gamma})}{L_1 L_2} \tag{7.6.21}$$

where $\bar{\Gamma}$ is the region of the dual lattice inside the closed path Γ and $\mathcal{A}(\bar{\Gamma})$ is its area. Thus, *in the thermodynamic limit*, the function Θ has a jump equal to one as a closed path Γ is traversed. Equivalently, we can say that Θ is a *multivalued function* which has a branch cut representing a jump by one unit. Using the same line of reasoning, one can show that the following important identity holds

$$\Theta(\vec{x},\vec{r}\,') - \Theta(\vec{x}\,',\vec{r}) = \frac{1}{2}. \tag{7.6.22}$$

This equation can be derived by using the following geometric construction. Draw a rectangle centered at $\vec{x}$ which has corners at $\vec{x}+\vec{R}$ and $\vec{x}-\vec{R}$ along a diagonal. We now consider the paths Γ_1, $\vec{x}+\vec{R} \to \vec{x}-\vec{R}$ without crossing the cut, and Γ_2, $\vec{x}+\vec{R} \to \vec{x}-\vec{R}$ crossing the cut. By symmetry, we have

$$\Big(\Theta(\vec{x}-\vec{R},\vec{r}) - \Theta(\vec{x}+\vec{R},\vec{r})\Big)_{\Gamma_1} = \Big(\Theta(\vec{x}+\vec{R},\vec{r}) - \Theta(\vec{x}-\vec{R},\vec{r})\Big)_{\Gamma_2}. \tag{7.6.23}$$

Since the total discontinuity of Θ is one, $[\Delta\Theta]_{\Gamma_1+\Gamma_2} = 1$, we get just half that result for a "half-way trip".

We can now use the Cauchy-Riemann equation, Eq. (7.6.19), to write the vector potential $\mathcal{A}_j$ in Eq. (7.6.18) as the *gradient* of a scalar "function" $\phi(\vec{x})$:

$$\mathcal{A}_j(\vec{x}) = \Delta_j \phi(\vec{x}) \tag{7.6.24}$$

where ϕ is given by

$$\phi(\vec{x}) = \frac{1}{\theta} \sum_{\vec{x}'} \Theta(\vec{x}, \vec{r}\,')\rho(\vec{x}\,'). \tag{7.6.25}$$

Therefore, the vector potentials associated with the *local* gauge degrees of freedom are pure gradients, and they can be "eliminated" by means of the "gauge transformation"

$$a(\vec{x}) = e^{-i\phi(\vec{x})}\, c(\vec{x}). \tag{7.6.26}$$

However, since ϕ is a function of the local density $\rho(\vec{x})$, the phase factor $e^{-i\phi}$ is not a c-number but an operator. This operator creates a coherent state of vector potentials which represents the flux attached to the particles. The operators $a(\vec{x})$ *so defined* satisfy the anyon commutation relations and the hard-core condition. Indeed, after some straightforward algebra we get that the operators $a(\vec{x})$ satisfy the commutation relations

$$\hat{a}(\vec{x})\hat{a}^{\dagger}(\vec{y}) = \delta_{\vec{x},\vec{y}} - e^{i\delta}\hat{a}^{\dagger}(\vec{y})\hat{a}(\vec{x}) \tag{7.6.27}$$

where the "phase" δ is given by

$$\delta = \frac{1}{\theta}\Big(\Theta(\vec{x}, \vec{r}\,') - \Theta(\vec{x}\,', \vec{r})\Big) = \frac{1}{2\theta}. \tag{7.6.28}$$

The hard-core condition $a(\vec{x})^2 = 0$ is a consequence of the fact that the operator $c(\vec{x})$ is a fermion. Thus, the operators $a(\vec{x})$ and $a^{\dagger}(\vec{x})$ are anyon destruction and creation operators. The statistical angle δ and the Chern-Simons coupling constant θ are related by

$$\delta = \frac{1}{2\theta}. \tag{7.6.29}$$

Notice that this is the same result that we derived in section 7.2 by considering a first quantized path integral approach.

It is clear that much of what was done above for a lattice theory can also be done in the continuum case. Thus, the identification of anyons with either fermions or bosons coupled to Chern-Simons gauge fields is also valid for continuum systems [Semenoff 88] but with one caveat. The notion of attaching fluxes to particles in the continuum is a very tricky one. We remarked in section 7.2 that, in addition to fractional statistics, the particles may acquire a fractional induced spin depending on the definition of the problem at short distances. For example, if the particle and the charge literally "sit on top of each other", there is no relative winding and no extra phase can possibly appear. But, if the particle and the flux are separated by some distance, they can wind around each other. As a result an extra phase may appear in the propagation amplitudes. This extra phase can be interpreted as an induced fractional spin. The lattice theory that we have discussed above does separate particles from fluxes in a natural and gauge-invariant way. We then expect that this lattice anyons should have an induced fractional spin.

7.7 Periodicity and Families of Chern-Simons Theories

The results of the last section allow us to conclude that a theory of *anyons* with statistical phase δ can be *defined* in terms of a theory of *fermions* coupled to a *Chern-Simons gauge field* with a coupling constant $\theta = \frac{1}{2\delta}$. Likewise, the *same* theory of anyons can also be *defined* in terms of a theory of *bosons with hard cores* coupled to a Chern-Simons gauge field but with a coupling constant $\theta = 1/2(\delta \pm \pi)$. This equivalency is the starting point of the boson approach.

However, there is an apparent discrepancy between the fermion (or boson) and anyon theories. The problem is that the anyon commutation relations are *periodic* in the statistical phase δ. Nothing changes in the anyon problem if the statistical phase is shifted by $\delta \to \delta + 2\pi n$, where n is an *arbitrary integer*, not necessarily positive. On the other hand, the only information in Chern-Simons theory about the statistics of the particles is in the coupling constant θ. It is not obvious that the Chern-Simons theory is *invariant* under the change in its coupling constant $\frac{1}{\theta} \to \frac{1}{\theta} + 4\pi n$ as it is required by the anyon commutation relations. This issue is of particular importance since all approximations which are commonly made, such as the average field approximation of Laughlin, work only in one particular period *i.e.*, a choice of n. Fortunately, it is possible that the Chern-Simons theory is indeed invariant under shifts. Notice that a shift of δ by $2\pi n$ is equivalent to attaching an additional *even* number $2n$ of flux quanta to each one of the particles. The argument is the following.

Let us first prove that "an even number of flux quanta is the same as nothing". Consider a system of fermions coupled to a Chern-Simons gauge field with coupling constant $\theta = \frac{1}{4\pi n}$. In first quantization, the functional integral reduces to a sum over all the histories of the particles and gauge fields. In section 7.2 we showed that the trajectories of the fermions form braids. If we compare two histories which differ just by the relative braiding of two particles, the propagation amplitude changes just by a phase factor $\exp(i\frac{\Delta\nu}{2\theta})$, where $\Delta\nu$ is the change in the linking number. Thus, for $\theta = \frac{1}{4\pi n}$, all scattering amplitudes remain unchanged since the phase change is just an integer multiple of 2π.

This suggests that if we want to attach an additional even number of fluxes to each particle, then we have to couple the system of fermions to a *new* Chern-Simons gauge field, let us call it $\mathcal{A}'_\mu$, with a coupling constant $\theta' = \frac{1}{4\pi n}$. Thus, the fermions end up being coupled to *two* Chern-Simons gauge fields, one which is responsible for the fractional statistics and the other for the periodicity.

However, the resulting theory seems to be unnecesarily complicated. This problem can be remedied quite easily. Since the Chern-Simons action is bilinear in the fields, we can integrate out one of the two gauge fields. More precisely, let us consider a problem in which two Chern-Simons gauge fields, $\mathcal{A}_\mu$ and $\mathcal{A}'_\mu$, are both coupled to the same fermi field ψ through the Lagrangian

density $\mathcal{L}$ (I drop the subindex of the gauge fields)

$$\mathcal{L} = \mathcal{L}_{\rm F}[\psi, \mathcal{A} + \mathcal{A}'] + \theta_1 \mathcal{L}_{\rm cs}[\mathcal{A}] + \theta_2 \mathcal{L}_{\rm cs}[\mathcal{A}'] \tag{7.7.1}$$

where $\mathcal{L}_{\rm F}[\psi, \mathcal{A} + \mathcal{A}']$ is the fermion part of the Lagrangian. Note that the fermions are assumed to couple in the same awy to both gauge fields. This is needed for the fluxes to be additive. We can use the invariance of the integration measure to define a new gauge field $A = \mathcal{A} + \mathcal{A}'$. The fermion only couples to the field A.

Let us now compute the functional integral over the fields $\mathcal{A}'$. After the shift the Lagrangian reads

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\rm F}[\psi, A] + \theta_1 \mathcal{L}_{\rm cs}[A - \mathcal{A}'] + \theta_2 \mathcal{L}_{\rm cs}[\mathcal{A}'] \\ &= \mathcal{L}_{\rm F}[\psi, A] + (\theta_1 + \theta_2)\mathcal{L}_{\rm cs}[\mathcal{A}'] + \theta_1 \mathcal{L}_{\rm cs}[A] - \frac{\theta_1}{2}\epsilon_{\mu\nu\lambda}\mathcal{A}'_\mu F_{\nu\lambda}. \end{aligned} \tag{7.7.2}$$

The functional integral over the $\mathcal{A}'_\mu$ fields can be carried out exactly. As usual one first shifts the field $\mathcal{A}'_\mu \to \mathcal{A}'_\mu + \tilde{\mathcal{A}}_\mu$ and $\tilde{\mathcal{A}}_\mu$ is then determined from the condition that the terms linear in $\mathcal{A}'_\mu$ are exactly cancelled. This condition yields the result

$$\tilde{\mathcal{A}}_\mu = \left(\frac{\theta_1}{\theta_1 + \theta_2}\right) A_\mu. \tag{7.7.3}$$

The fermions are coupled to a *single* Chern-Simons gauge field A_μ with the effective Lagrangian

$$\mathcal{L}_{\rm eff} = \mathcal{L}_{\rm F}[\psi, A_\mu] + \theta_{\rm eff}\mathcal{L}_{\rm cs}[A_\mu]. \tag{7.7.4}$$

The effective Chern-Simons coupling $\theta_{\rm eff}$ given by

$$\frac{1}{\theta_{\rm eff}} = \frac{1}{\theta_1} + \frac{1}{\theta_2}. \tag{7.7.5}$$

If we make the choice $\theta_2 = \frac{1}{4\pi n}$ we get the desired result.

Thus, Chern-Simons theories with coupling constants θ of the form $\frac{1}{\theta} = 2\delta + 4\pi n$ have the same physical properties. However, approximations done on each member of this sequence yield quite different results. This property will be of great importance for our discussion of the Fractional Quantum Hall Effect in chapter 10.

7.8 The Jordan-Wigner Transformation in Two Dimensions

The identity

$$\hat{a}(\vec{x})\hat{a}^\dagger(\vec{y}) = \delta_{\vec{x},\vec{y}} - e^{i\delta}\hat{a}^\dagger(\vec{y})\hat{a}(\vec{x}) \tag{7.8.1}$$

is the two-dimensional analog of the Jordan-Wigner Transformation discussed in chapter 4. As a matter of fact, for $\theta = \frac{1}{2\pi m}$ we get $\delta = \pi m$. Hence, for m

odd the operators $a(\vec{x})$ obey equal-time *boson* commutation relations and a hard-core condition. If we recall the mapping in section 7.3 between bosons with hard cores and spin-$\frac{1}{2}$ Pauli operators:

$$\sigma^+(\vec{x}) = a^\dagger(\vec{x}), \quad \sigma^-(\vec{x}) = a(\vec{x}), \quad \sigma_3(\vec{x}) = 2a^\dagger(\vec{x})a(\vec{x}) - 1, \tag{7.8.2}$$

we get from Eq. (7.6.26)

$$\sigma^+(\vec{x}) = e^{+i\phi(\vec{x})}\, c^\dagger(\vec{x}), \quad \sigma^-(\vec{x}) = e^{-i\phi(\vec{x})}\, c(\vec{x}), \quad \sigma_3(\vec{x}) = 2c^\dagger(\vec{x})c(\vec{x}) - 1. \tag{7.8.3}$$

These equations tell us that the two-dimensional quantum Heisenberg antiferromagnet on a square lattice is *exactly equivalent* to a theory of spinless fermions *on the same lattice* coupled to a Chern-Simons gauge field. In addition, there is a direct density-density repulsive force among nearest neighbors. Thus, unlike the familiar results from one dimension, in which the fermions are *free* (see chapter 4) in the XY limit, there is a long-range gauge interaction in two dimensions even in the XY limit. This property is due to the fact that, in one dimension, the only possible flux that the fermions can feel is a global effect determined by the boundary conditions. In two-dimensions, the fermions feel both a local and a global flux. As we will see next, even the global flux is non-trivial. Although the resulting fermion theory is not free, approximations and perturbation theory in one scheme still turns into a non-perturbative feature in the other.

7.9 Quantizing the Global Degrees of Freedom

In this section we consider the global degrees of freedom. Here I follow the results of Wen, Dagotto and myself [Wen 90]. The *global* gauge degrees of freedom $\bar{A}_j$ are completely unaffected by the Jordan-Wigner transformation which only involves local transformations. They satisfy the *homogeneous* constraint equation

$$\epsilon_{jk}\Delta_j \bar{A}_k = 0. \tag{7.9.1}$$

As was discusse above, these degrees of freedom cannot be eliminated by local ("small") gauge transformations since they have non-vanishing circulation Γ_j along any large circles $\mathcal{C}_j$ of the torus. The "best" we can do is, for instance, to pick the gauge in which the fields $\bar{A}_j$ are constant in space (but not in time!)

$$\bar{A}_j = \frac{\Gamma_j(t)}{L_j} \tag{7.9.2}$$

(no sum over j is implied).

These relations allow us to derive an effective Lagrangian for the global degrees of freedom $\Gamma_j(t)$ and to extract from it the quantum dynamics of the global degrees of freedom. By carrying out the canonical formalism to completion, it is easy to check that the non-integrable phases obey the commutation

relations

$$[\Gamma_1, \Gamma_2] = \frac{i}{\theta}. \tag{7.9.3}$$

Hence, Γ_1 and $\theta\Gamma_2$ form a *canonical pair* and cannot be diagonalized simultaneously. This feature is not present in one-dimensional systems for which there is only one non-integrable phase which is just a c-number. The global degrees of freedom in one dimension are just boundary conditions. In 2+1 dimensions, we discover that the global degrees of freedom acquire a life of their own. We will see now that, as a result of this feature, the states of anyon systems on a torus are not determined by the location of the particle alone.

It is now easy to check that the operators $\exp(i\Gamma_j)$ satisfy the algebra

$$e^{i\Gamma_1}\, e^{i\Gamma_2} = e^{-\frac{1}{\theta}} e^{i\Gamma_2}\, e^{i\Gamma_1}. \tag{7.9.4}$$

Let us denote the exponential operators $\exp(i\Gamma_j)$ by T_j . These operators will give an extra phase to any state as the anyons move around each other. Furthermore, since Γ_1 and Γ_2 do not commute, the *eigenstates* of the Hamiltonian are only functions of either variable but not of both at the same time. Also, both Γ_1 and Γ_2 enter only through the exponential operators T_j. Thus we can always choose say Γ_1 to be an angle with a range $[0, 2\pi]$. Hence $\theta\Gamma_2$ is an angular momentum-like operator whose spectrum is the set of integers. In all cases of physical interest, the statistical angle θ can only take the restricted set of values $\theta = \frac{m}{2\pi n}$, where m and n are integers. After all *local* gauge degrees of freedom are eliminated, we find that the effective Hamiltonian for the anyon system has the form

$$H = \sum_{\vec{x}, j=1,2} a^\dagger(\vec{x}) e^{i\left(\mathcal{A}_j(\vec{x}) + \frac{\Gamma_j}{L_j}\right)}\, a(\vec{x} + e_j) + \text{h.c.} \tag{7.9.5}$$

where $\mathcal{A}_j$ is given by the solution of the local constraint. This Hamiltonian is *almost* identical to the "free anyon" Hamiltonian. The only difference here is the presence of the *global* degrees of freedom Γ_j which were not included in our original naive expression. We will adopt this generalized version as the *definition* of the anyon Hamiltonian. In other words, the global degrees of freedom are an intrinsic feature of the anyon system on a torus. Clearly, if the manifold on which the anyons move is not a torus, but some other manifold, the properties of the global degrees of freedom will be different. For instance, if the system is quantized on a manifold with a boundary, such as a disc, there are no global degrees of freedom. Instead, gauge invariance requires the existence of edge states which have very interesting properties.

The form of the Hamiltonian suggests that its eigenstates are not functions only of the coordinates of the anyons since H involves the global degrees of freedom as well. Let us denote by Ψ_0 an eigenstate of H. We can also choose Ψ_0 to be an eigenstate of Γ_1 with zero eigenvalue or, what is equivalent, to be an eigenstate of T_1 with unit eigenvalue

$$T_1 \Psi_0 = \Psi_0. \tag{7.9.6}$$

Let us consider now the state Ψ_p defined by

$$\Psi_p \equiv T_2^p \Psi_0. \tag{7.9.7}$$

The state Ψ_p is an eigenstate of T_1

$$T_1 \Psi_p \equiv T_1 T_2^p \Psi_0 = e^{-\frac{ip}{\theta}} T_2^p T_1 \Psi_0 = e^{-\frac{ip}{\theta}} \Psi_p \tag{7.9.8}$$

with the eigenvalue $e^{-\frac{ip}{\theta}}$. Thus, for all cases of physical interest ($\theta = \frac{m}{2\pi n}$), there are m distinct eigenstates, each labelled by the integer p. The states of the Hilbert space are thus labelled by the anyon coordinates *and* by the quantum number p describing the state of the global degrees of freedom. In particular the condensed states of the anyon system do exhibit this degeneracy. The idea that such topological degeneracies occur quite generally in spin liquid states and other topologically ordered states is originally due to Wen [Wen 90].

References

[Chen 89] Y-H. Chen, F. Wilczek, E. Witten and B.I. Halperin. Int. J. Mod. Phys. B3, 1001, (1989).

[Dotsenko 84] V. Dotsenko. Nucl. Phys. 235B (FS11), 54, (1984).

[Fradkin 80] E. Fradkin and L.P. Kadanoff. Nucl.Phys. 170B (FS1), 1, (1980).

[Fradkin 89] E. Fradkin. Phys. Rev. Lett. 63, 322, (1989).

[Fradkin 91] E. Fradkin and F. Schaposnik. Phys. Rev. Lett. 66, 276, (1991).

[Frohlich 88] J. Frohlich and P. Marchetti. Lett. Math. Phys. 16, 347, (1988).

[Itzykson 80] C. Itzykson and J.B. Zuber. *Quantum Field Theory*, McGraw-Hill (1980).

[Jackiw 82] R. Jackiw, S. Deser and S. Templeton. Phys. Rev. Lett. 48, 372 (1982).

[Jordan 28] P. Jordan and E.P. Wigner. Z. Phys. 47, 631, (1928).

[Kadanoff 71] L.P. Kadanoff and H. Ceva. Phys. Rev. B3, 3918, (1971).

[Kalmeyer 88] V. Kalmeyer and R.B. Laughlin. Phys. Rev. Lett. 59, 2095 (1988).

[Klauder 79] J. Klauder. Phys. Rev. D , 2349 (1979).

[Laughlin 83] R.B. Laughlin. Phys. Rev. Lett. 50, 1395, (1983).

[Laughlin 88] R.B. Laughlin. Science 242, 525 (1988).

[Leinaas 77] J.M. Leinaas and J. Myrheim. Il Nuovo Cimento, 37B, 1, (1977).

[Polyakov 87] A.M. Polyakov. *Gauge Fields and Strings*, Harwood (1987).
[Polyakov 88] A.M. Polyakov. Mod. Phys. Lett. A3,325 (1988).
[Schonfeld 81] J. Schonfeld. Nucl. Phys. B185, 157 (1981).
[Semenoff 88] G. Semenoff. Phys. Rev. Lett. 61, 517, (1988).
[Wen 90] X.G. Wen. Int. Jour. Mod. Phys. 4, 239, (1990).
[Wen 90] X.G. Wen. Int. Jour. Mod. Phys. 4, 239, (1990).
[Wilczek 82] F. Wilczek. Phys. Rev. Lett. 48, 1144 (1982).
[Witten 89] E. Witten. Comm. Math. Phys. 121, 351 (1989).

CHAPTER 8

Anyon Superconductivity

8.1 Anyon Superconductivity

In this chapter we will consider the problem of predicting the behavior of an assembly of particles obeying fractional statistics. In the past sections we considered the problem of the quantum mechanics of systems of anyons. However, we did not consider what new phenomena may arise if the system had a macroscopic number of anyons present. At the moment of writing these lines, the physical reality of this problem is still unclear. However, this is such a fascinating problem that we will discuss it despite the lack of firm experimental support for the model.

There are two different physical situations in which the problem of anyons at finite density is important. Halperin observed [Halperin 84] that the quasiparticles of the Laughlin state for the Fractional Hall effect obeyed fractional statistics (*i.e.*, they are *anyons*). In chapter 10 we will discuss Halperin's theory. Furthermore, Halperin and Haldane suggested that, for filling fractions different from the $\frac{1}{m}$ Laughlin sequence, the ground state of a two dimensional electron gas in a strong magnetic field could be understood as a Laughlin state of anyons. Shortly afterwards, Arovas, Schrieffer, Wilczek and Zee [Arovas 85] studied the high temperature behavior of a gas of anyons and calculated the second virial coefficient.

Much of the recent interest on this problem is connected to its possible relevance to high temperature superconductors. Since anyons "interpolate" between fermions and bosons it is natural to ask if an assembly of anyons at finite density is more "fermion-like" or "boson-like". Fermions have non-condensed ground states with Fermi surfaces while bosons undergo Bose condensation and are superfluids. In two remarkable papers, Laughlin [Laughlin 88] argued that anyons generally form "condensates" in the sense that their ground states exhibit superfluid properties. Fetter, Hanna and Laughlin [Fetter 89] (FHL) developed a mean-field-theoryanyon gas for the free anyon gas in the continuum which has generally confirmed these conjectures. They argued that, if one represents anyons in terms of fermions coupled to fractional fluxoids, then the fermions feel an effective average flux determined by the

fermions coupled to fractional fluxoids, then the fermions feel an effective average flux determined by the particle density. A quantum Hall effect-like picture could then be used, at least within mean-field-theory. In a sense this is a very surprising result since a Hall Effect system is *incompressible* and, thus, it does not have any low energy modes. However, the flux is uniform only on average since the constraints force it to fluctuate together with the particle density. Fetter, Hanna and Laughlin showed that this was indeed the case. They did a calculation with the flavor of a Random Phase Approximation (RPA) and found a Goldstone pole in the (fermion) current-current correlation function. Hence, the fluctuations restore the *compressibility* which is necessary for the system to behave like a condensate. They argued that this pole implies the presence of a Meissner effect for an external electromagnetic field. This picture relies on two crucial assumptions: 1) the fermions can effectively be stripped off their fluxes and 2) the Goldstone pole is robust against fluctuations.

The predictions of Fetter, Hanna and Laughlin have, to some extent, been confirmed by extensive numerical calculations [Canright 89]. Chen, Wilczek, Witten and Halperin [Chen 89] have offered two (complementary) arguments to explain why the Goldstone pole is exact. One argument is based on the f-sum rule which is a consequence of gauge invariance (see chapter 9). For a nice derivation see, for instance, the book of Martin [Martin 75]. Their other argument (spontaneous violation of a *fact*) is based on translation invariance and on the fact that the translation generators of this theory become anomalous in the mean field ground state, reflecting the fact that they are actually the generators of magnetic translations. This latter argument is strongly reminiscent of a field theoretic anomaly. I will show below that this is indeed the case and that the "anyon mechanism" for superconductivity is the two-dimensional analog of the Schwinger [Schwinger 62] or Anderson [Anderson 63] mechanism more familiar from one space dimension.

Wen and Zee [Wen 90], Fisher and Lee [Fisher 89] and Kitazawa and Murayama [Kitazawa 90] have considered this problem from a bosonic point of view. In this language, one focuses more directly on the role of vortices, anyons binding into "Cooper bound states", *etc.* The emerging picture is complementary to the fermion description. Local operators in one language are non-local "disorder" operators in the other. It is worth noting that a similar picture has been developed for the Fractional Hall Effect [Girvin 87], [Zhang 89], [Read 89].

In our discussion here, I will follow my own work which is based on the path-integral approach for fermions coupled to the Chern-Simons theory [Fradkin 90], see also [Salam 90], [Hosotani 90]. In this approach, the exactness of the Goldstone modes follows from the topological invariance of an effective Hall conductance. In chapters 9 and 10, we discuss these issues of topological invariance and quantization at great length within the framework of the theory of the Quantum Hall Effect.

8.2 Functional Integral Formulation of the Chern-Simons Theory

In this section I consider the functional integral formalism for a system of $\mathcal{N}_a$ anyons at zero temperature. I will use the fermion formalism discussed above. I will work with a chemical potential μ which will be determined later from the requirement that the density ρ be equal to $\frac{\mathcal{N}_a}{L^2}$ for a system with L^2 sites (I assume a square lattice of L^2 sites with lattice constant $a_0 = 1$).

The functional integral representation for the partition function of this system at zero temperature (in real time) with chemical potential μ and background electromagnetic fields A_μ ($\mu = 0, 1, 2$) is given by

$$Z = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi\mathcal{D}\mathcal{A}\; e^{i\int dt L} \tag{8.2.1}$$

where the ψ's are Grassmann fields and some gauge fixing procedure is implicitly assumed. This functional integral has to be understood as a Coherent State path-integral. Let us consider the gauge field sector for the moment. The fermion sector is already known to be a Coherent State path-integral. In section 7.6, I showed that $\mathcal{A}_1$ and $\theta\mathcal{A}_2$ form a canonical pair. Notice that $\mathcal{A}_1$ resides on the link $(\vec{x}, \vec{x} + \hat{e}_1)$ while $\theta\mathcal{A}_2$ resides on the orthogonal link $(\vec{x}, \vec{x} + \hat{e}_2)$. Let $q(\vec{x})$ denote $\mathcal{A}_1(\vec{x})$ and $p(\vec{x})$ denote $\theta\mathcal{A}_2(\vec{x})$. The operators $p(\vec{x})$ and $q(\vec{x})$ obey canonical commutation relations. In the derivation of the path-integral one has to introduce complete sets of states at every intermediate time of the evolution. However, since $\mathcal{A}_1$ and $\mathcal{A}_2$ do not commute, we cannot define a complete set of states in which both are diagonal. Let us say we choose a basis in which $\mathcal{A}_1$ is diagonal and that we now insert a complete set of such states at every intermediate time. In addition, the states have to be restricted so as to satisfy the *local* constraint. This is implemented by means of a Lagrange multiplier $\mathcal{A}_0(\vec{x}, t)$ at every lattice site and at all times. The matrix elements of the time evolution operator for an infinitesimal time δt is not easy to compute in such a basis. Thus, it is convenient to also introduce a complete set of states in which $\mathcal{A}_2$ is diagonal. It is easy to show that, in addition to a term of the form $\theta\mathcal{A}_0\mathcal{B}$ which arises from the constraint, we get an extra term in the Lagrangian of the form $\theta\sum_{\vec{x}}\mathcal{A}_2(\vec{x})\partial_0\mathcal{A}_1(\vec{x})$. This term is generated by the overlaps of $\mathcal{A}_1$ and $\mathcal{A}_2$ states on neighboring time slices. Both sets of terms can be condensed into a single expression: the Chern-Simons Lagrangian. Hence, the functional integral is just the phase-space integral for the canonical pair $\mathcal{A}_1$ and $\theta\mathcal{A}_2$.

In the last section we derived the Lagrangian of the system. The anyons are coupled to the electromagnetic field via the minimal coupling prescription. Thus, all we need to do in order to include the chemical potential μ and the electromagnetic fields A_μ is to modify the derivatives and amplitudes in the

usual manner

$$D_0 = \partial_0 - i\mathcal{A}_0 \rightarrow \partial_0 - i\,(\mathcal{A}_0 + A_0 + \mu)\,,$$
$$e^{i\mathcal{A}_j(x)} \rightarrow e^{i(\mathcal{A}_j(x)+A_j(x))}. \tag{8.2.2}$$

Notice that, as usual, the chemical potential μ can be regarded as a constant shift of A_0. The integration measures are invariant measures.

8.3 Correlation Functions

The response of the system to slowly varying electromagnetic fields, can be studied in terms of the current correlation functions. In addition, we will be also interested in correlations which probe other features of the spectrum of the system. In particular it is of interest to study the *gauge-invariant* fermion propagator

$$G_\Gamma(x,x') = \left\langle \bar{\psi}(x)\; e^{i\int_\Gamma \mathcal{A}}\; \psi(x') \right\rangle \tag{8.3.1}$$

where $\int_\Gamma \mathcal{A}$ is a short hand for the line integral of the statistical vector potentials along some path Γ. Likewise the pair correlation function can be calculated in terms of the gauge-invariant four-point function, and so on. Other probes of interest are Wilson loops for the statistical vector potential $\mathcal{A}$ along a closed path Γ:

$$W_{\mathcal{A}}[\Gamma] = \left\langle e^{i\oint_\Gamma \mathcal{A}} \right\rangle. \tag{8.3.2}$$

In particular a space-like Wilson loop for a closed path Γ on the square lattice must represent, as a result of the constraint, the fluctuation in the number of fermions $\mathcal{N}_a(\Sigma)$ (and hence anyons) inside the region Σ bounded by the loop Γ:

$$W_{\text{space}}\,[\Gamma = \partial\Sigma] = \left\langle e^{i\oint_\Gamma \mathcal{A}} \right\rangle = \left\langle e^{\frac{i}{\theta}\int_\Sigma j_0(x)} \right\rangle \equiv \left\langle e^{\frac{i}{\theta}\mathcal{N}_a(\Sigma)} \right\rangle. \tag{8.3.3}$$

In the case of a time-like loop Γ, the constraint implies that a static particle has been added at one point and one subtracted at another point. Thus $\langle e^{i\oint_\Gamma \mathcal{A}}\rangle$ for time-like loops roughly represents the energy cost for adding a particle, say at $\vec{r}$, and removing it at $\vec{r}\,'$. This is the standard interpretation of the Wilson loop. Notice however, that now a particle is added without adding a flux. Hence we are creating a mismatch between charge and flux.

Analogously we can create a coherent state which represents a (static) flux piercing a given plaquette at dual site $\vec{r}$. The operator which creates this state is

$$K(\vec{x}) \equiv e^{i\theta\sum_{x'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\mathcal{A}_j^c(\vec{x}\,')} \tag{8.3.4}$$

where $\mathcal{A}_j^c(\vec{x}\,')$ is a background static vector potential with a curl equal to the flux. For $\mathcal{A}_j^c(\vec{x}\,')$ to represent a flux we demand that $\mathcal{A}_j^c(\vec{x}\,') = \frac{1}{\theta}\Delta_j\Theta(\vec{x}\,',\vec{r}\,)$.

It is easy to show that this operator $K(\vec{x})$ is precisely identical to the operator $e^{i\phi}$ defined in Eq. (7.6.25). Indeed, using Gauss' law, Eq. (7.6.6),

$$\theta\epsilon_{jk}\Delta_j\mathcal{A}_k - j_0 = 0 \tag{8.3.5}$$

one finds (up to boundary terms)

$$\begin{aligned}\theta\sum_{\vec{x}'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\mathcal{A}_j^c(\vec{x}\,') &= \sum_{\vec{x}'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\Delta_j\Theta(\vec{x}\,',\vec{r}\,)\\ &= -\sum_{\vec{x}'}\epsilon_{jk}\Big(\Delta_j\mathcal{A}_k(\vec{x}\,')\Big)\Theta(\vec{x}\,',\vec{r}\,)\\ &= -\frac{1}{\theta}\sum_{\vec{x}'}j_0(\vec{x}\,')\Theta(\vec{x}\,',\vec{r}\,).\end{aligned} \tag{8.3.6}$$

By means of the identity (see Eq. (7.6.22))

$$\Theta(\vec{x},\vec{r}\,') - \Theta(\vec{x}\,',\vec{r}\,) = \frac{1}{2} \tag{8.3.7}$$

one finds

$$\begin{aligned}\theta\sum_{\vec{x}'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\mathcal{A}_j^c(\vec{x}\,') &= \frac{1}{2\theta}\sum_{\vec{x}'}j_0(\vec{x}\,') - \frac{1}{\theta}\sum_{\vec{x}}j_0(\vec{x}\,')\Theta(\vec{x},\vec{r}\,')\\ &\equiv +\frac{1}{2\theta}\mathcal{N}_a - \phi(\vec{x}).\end{aligned} \tag{8.3.8}$$

Hence

$$K(\vec{x}) = e^{\frac{i}{2\theta}\mathcal{N}_a}e^{-i\phi(\vec{x})}. \tag{8.3.9}$$

Clearly $K(\vec{x})$ *is not invariant* under local gauge transformation of the statistical gauge field. Indeed, for a gauge transformation $\mathcal{A}_j(\vec{x}) \to \mathcal{A}_j(\vec{x}) + \Delta_j\varphi(\vec{x})$, we get

$$\theta\sum_{\vec{x}'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\mathcal{A}_j^c(\vec{x}\,') \to \theta\sum_{\vec{x}'}\epsilon_{jk}\mathcal{A}_k(\vec{x}\,')\mathcal{A}_j^c(\vec{x}\,') + \theta\sum_{\vec{x}'}\epsilon_{jk}\Delta_k\varphi(\vec{x})\mathcal{A}_j^c(\vec{x}). \tag{8.3.10}$$

By integrating by parts and with the help of Eqs. (7.6.19) and (7.6.16) one finds

$$K(\vec{x}) \to K(\vec{x})e^{+i\varphi(\vec{x})} \tag{8.3.11}$$

and thus the product

$$a(\vec{x}) = K(\vec{x})c(\vec{x}) \tag{8.3.12}$$

is gauge invariant. Obviously the operator $a(\vec{x})$ is nothing but the anyon operator.

8.4 The Semi-Classical Approximation

We are interested in studying the physical properties of the partition function of a gas of anyons. In particular we want to understand the following issues: (1) the spectrum of low lying excitations, (2) statistics of the quasiparticles, (3) does it exhibit superfluidity, (4) if there is a Meissner effect, and (5) behavior of correlation functions.

I will study this problem by treating the functional integral within the semi-classical (saddle-point) expansion. Formally this requires the presence of a large coefficient in front of the action $S = \int dt L$. This system does not have such a coefficient (apart from $\frac{1}{\hbar}$ itself). It is plausible that at large-densities the saddle-point approximation may become accurate. Such is the case for the (weakly interacting) electron gas where the Random Phase Approximation (RPA) works very well. Since the statistical angle δ happens to be equal to $\frac{1}{2\theta}$ one expects that this approximation may also work *for large values of* θ (*i.e.*, almost a fermion). This is the limit studied by Chen, Wilczek, Witten and Halperin. In the Bose limit ($\theta = \frac{1}{2\pi}$) the results depend crucially on the density. In fact it is well known that the hard-core Bose gas, at moderate densities, can be treated within RPA due to the effective softening of the hard-cores at such densities. At high-densities on a lattice, this approximation breaks down and the hard-cores cause the existence of crystalline states or off-diagonal long range order (ODLRO) for the spin one-half XY model. However, it is conceivable that regimes of Bose systems may exist for which the results of a *fermion* mean field theory may still be qualitatively correct. The results of Fisher and Lee [Fisher 89] suggest that this may be the case.

The saddle-point approximation (SPA) may also be formally justified by considering a system of M *species* of anyons (each with $\mathcal{N}_a$ particles) which are "free" in the sense that there is no explicit interaction term in the Hamiltonian. The requirement of fractional statistics, of course, amounts to an interaction since it is equivalent to the statement that all M species of *fermions* interact through the *same* statistical vector potential $\mathcal{A}_\mu$. At large M, with $\theta = \theta_0 M$, the saddle-point approximation is formally correct. For the sake of simplicity I will consider only $M = 1$ and assume that the approximation is, *at least*, qualitatively correct.

The SPA is now carried out in the usual fashion. One first observes that the action is a bilinear form in fermion variables. Thus the fermions can be integrated out explicitly. The result is naturally a determinant

$$\int \mathcal{D}\bar{\psi}\mathcal{D}\psi \; e^{iS_F} = \mathrm{Det}\Big[iD_0 - h[\mathcal{A}_j + A_j]\Big] \tag{8.4.1}$$

where S_F is the *fermion* part of the action

$$S_F = \int dt \sum_{\vec{x},\vec{x}'} \bar{\psi}(x)\Big(iD_0\delta_{\vec{x},\vec{x}'} - h[\mathcal{A}_j + A_j]\Big)\psi(x') \tag{8.4.2}$$

and the *one-particle Hamiltonian* $h[\mathcal{A}]$ is

$$h[\mathcal{A}_j] = \tau \sum_{j=1,2} e^{i\mathcal{A}_j(x)} \delta_{\vec{x}',\vec{x}+\hat{e}_j}. \tag{8.4.3}$$

The condition $\mathcal{A}_j(x) = -\mathcal{A}_{-j}(x+e_j)$ guarantees the hermiticity of the Hamiltonian.

Therefore, the statistical vector potentials $\mathcal{A}_\mu$ have the *effective action* S_{eff}

$$S_{\text{eff}}[\mathcal{A}_\mu, A_\mu] = -i \text{ Tr } \ln\left(iD_0 - h[\mathcal{A}+A]\right) - \frac{i\theta}{4} S_{\text{cs}}[\mathcal{A}]. \tag{8.4.4}$$

We can use the invariance of the measure to shift the statistical vector potentials $\mathcal{A}_\mu + A_\mu \to \mathcal{A}_\mu$. The result is that the effective action now reads

$$S_{\text{eff}}[\mathcal{A}_\mu, A_\mu] = -i \text{ Tr } \ln(iD_0 + \mu - h[\mathcal{A}]) - \frac{i\theta}{4} S_{\text{cs}}[\mathcal{A}-A] \tag{8.4.5}$$

where we have pulled the chemical potential μ out of the definition of D_0. In this form, the electromagnetic fields only appear in the Chern-Simons term which is *quadratic* in the fields. We can thus write

$$S_{\text{cs}}[\mathcal{A}-A] = S_{\text{cs}}[\mathcal{A}] + S_{\text{cs}}[A] - \epsilon_{\mu\nu\lambda}\left(\mathcal{A}^\mu F^{\nu\lambda} + A^\mu \mathcal{F}^{\nu\lambda}\right). \tag{8.4.6}$$

We will assume that the electromagnetic field A_μ is small and has zero *average* strength. In this case we may treat A_μ as a perturbation (*i.e.*, linear response theory). Note that a non-zero uniform external magnetic field cannot be treated in perturbation theory. This is crucial for the correct study of the Meissner effect. Let us consider, for the moment, the SPA in the abscence of external electromagnetic fields.

We demand that S_{eff} be stationary around some configuration $\mathcal{A}_\mu$ which is assumed to be time independent (*i.e.*, zero "electrical" statistical field $\bar{\mathcal{E}}$) and with uniform statistical "magnetic" field $\bar{\mathcal{B}}$. Thus

$$\left.\frac{\delta S_{\text{eff}}}{\delta \mathcal{A}_\mu}\right|_{\bar{\mathcal{A}}_\mu} = 0 \tag{8.4.7}$$

yields the Saddle-Point Equation (SPE)

$$\langle j_\mu^F \rangle_{\bar{\mathcal{A}}} = \frac{\theta}{2} \epsilon_{\mu\nu\lambda}\left(\bar{\mathcal{F}}^{\nu\lambda} - F^{\nu\lambda}\right) \tag{8.4.8}$$

where $\langle j_\mu^F \rangle$ is the gauge-invariant fermion current.

Since the electromagnetic field will be assumed to be small and with zero average, we will set $F^{\nu\lambda} = 0$ in the SPE for the rest of this section. In this case, and for solutions with ($\bar{\mathcal{B}} = \text{const}$, $\bar{\mathcal{E}} = 0$), we find

$$\rho = \theta\, \bar{\mathcal{B}} \tag{8.4.9}$$

where ρ is the fermion density.

The requirement that there should be $\mathcal{N}_a$ particles is met by requiring

$$-\frac{i}{Z}\frac{\partial Z}{\partial \mu} = \mathcal{N}_a. \tag{8.4.10}$$

Since μ is nothing but a constant shift of $\mathcal{A}_0$, one finds

$$\mathcal{N}_a = \theta\Phi \tag{8.4.11}$$

where Φ is the *total flux*

$$\Phi = \bar{\mathcal{B}}L^2. \tag{8.4.12}$$

Thus, we find

$$\theta\bar{\mathcal{B}} = \frac{\mathcal{N}_a}{L^2} \tag{8.4.13}$$

which is Laughlin's result. Thus, at the Saddle-Point level, the fermions feel an effective flux $\bar{\mathcal{B}}$ per plaquette.

The spectrum of this problem was studied by D. Hofstadter and its properties are summarized in chapter 9. He found that if the *number of particles is fixed*, then as $\bar{\mathcal{B}}$ *varies* the spectrum of the effective one-particle Hamiltonian is very rich and complex and, as a function of $\bar{\mathcal{B}}$, it has a fractal structure. However, in the problem at hand, $\bar{\mathcal{B}}$ *is determined by the number of particles.* In fact, for a system of $\mathcal{N}_a$ anyons on a lattice with L^2 sites, the density ρ is $\frac{\mathcal{N}_a}{L^2}$ and therefore can be written as a ratio of two relatively prime integers r and q, *i.e.*,

$$\rho = \frac{r}{q}. \tag{8.4.14}$$

Similarly, we can also write the statistical phase δ also in the form of an irreducible fraction in terms of two relatively prime integers n and m

$$\delta = \pi\frac{n}{m}. \tag{8.4.15}$$

Equivalently, the Chern-Simons coupling constant θ is given by

$$\theta = \frac{m}{2\pi n}. \tag{8.4.16}$$

The effective field $\bar{\mathcal{B}}$ is a fraction of the flux quantum, 2π,

$$\bar{\mathcal{B}} = 2\pi\frac{P}{Q} \tag{8.4.17}$$

where the two relatively prime integers P and Q are given from Eq. (8.3.13) by

$$2\pi\frac{P}{Q} = \frac{\rho}{\theta}. \tag{8.4.18}$$

Hence, we can write

$$\frac{P}{Q} = \frac{nr}{mq}. \tag{8.4.19}$$

The spectrum of one-particle states, the Hofstadter problem , for *rational fields* $\bar{\mathcal{B}} = 2\pi\frac{P}{Q}$, consists of q Landau bands each with $\frac{L^2}{Q}$ degenerate states (see section 9.2). In the continuum limit, these bands become the usual Landau levels. If we denote by f the fraction of occupied Landau bands, then f must be $\mathcal{N}_a \times \frac{Q}{L^2}$ since there is a total of $\mathcal{N}_a$ particles. The density is then $\frac{f}{Q}$. Using Eq. (8.3.14), we get $f = \frac{r}{q}Q$. Thus, f is an integer if and only if q is a factor of Q.

Let (a,b) denote the largest common factor of the pair of integers a and b. Let k and l be two integers defined satisfying $k = (n,q)$ and $l = (m,r)$. Hence, there exist four integers $\bar{n}$, $\bar{m}$, $\bar{r}$ and $\bar{q}$ such that

$$\begin{array}{ll} n = k\bar{n}, & q = k\bar{q}, \\ m = l\bar{m}, & r = l\bar{r}, \\ (\bar{n},\bar{q}) = 1, & (\bar{m},\bar{r}) = 1, \\ (\bar{n},\bar{m}) = 1, & (\bar{r},\bar{q}) = 1. \end{array} \tag{8.4.20}$$

Thus,

$$\frac{P}{Q} = \frac{\bar{n}\bar{r}}{\bar{m}\bar{q}} \tag{8.4.21}$$

and

$$\begin{aligned} P &= \bar{n}\bar{r}, \\ Q &= \bar{m}\bar{q}. \end{aligned} \tag{8.4.22}$$

Therefore, the fraction f of occupied Landau bands is

$$f = \frac{r}{q}Q = \frac{l}{k}\bar{r}\bar{m}. \tag{8.4.23}$$

It is easy to show that k does not have any common factors with either l, $\bar{r}$ or $\bar{m}$. In general, f is an irreducible fraction, unless one of the following conditions is satisfied

$$\begin{aligned} (n,q) &= 1, \\ (n,q) &= \frac{m}{(m,r)}, \\ (n,q) &= \frac{r}{(m,r)}. \end{aligned} \tag{8.4.24}$$

If f is not an integer, then there is no gap. Fluctuation effects should overwhelm the saddle-point results and this theory will generally be unstable. Hence, whenever possible, one must have f integer since, except for one special case, there is always a gap. In summary, for arbitrary density ρ and Chern-Simons coupling constant θ, it is not always possible to require f to be an integer. On the other hand, for the "happy fractions" listed above f is an integer and we have an integer number f of filled Landau bands. The physical behavior of the system will depend on which of the conditions listed above is realized. Thus, the physics of this problem is not just determined by the density and the statistics, but is also determined by number-theoretic conditions, *i.e.*, on the commensurability.

Of particular importance will be the sequence $\theta = \frac{m}{2\pi}$, *i.e.*, $n = 1$. In this case we have $k = (n,q) = 1$ and f is indeed an integer, $f = \frac{mr}{(m,r)}$. For this sequence, we have an integer number f of Landau bands filled for a system with *arbitrary* density $\rho = \frac{r}{q}$ and statistical parameter $\delta = \frac{\pi}{m}$. This is the case considered by Chen, Wilczek, Witten and Halperin. The saddle-point approximation is expected to work for θ large (*i.e.*, large m) which is the limit in which the anyons are *almost* fermions. The case of semions has $m = 2$.

An exceptional case occurs if $\rho = \frac{1}{2}$ (*i.e.*, half-filling) and $\theta = \frac{m}{2\pi n}$, with n an *odd* integer. In this case we get $P = n$, $Q = 2m$ and $f = m$. This means that all states with energy less than zero are filled and that the Fermi level is at $E = 0$. It has been known since Hofstadter's work that, in this case, there is a band crossing in the spectrum (see chapter 9). These bands cross at $E = 0$ at Q points of the Brillouin Zone. In fact the case $m = 1$ and $\rho = \frac{1}{2}$ corresponds to a "flux phase" with $\bar{\mathcal{B}} = \pi$ (*i.e.*, half-flux quantum per plaquette). In this case the fermion spectrum is effectively relativistic. In fact it has long been recognized that hopping in a frustrated lattice is an efficient way to set up the Dirac equation on a lattice. In general one finds Q species of Dirac fermions. Fluctuations in the statistical gauge fields may open up a gap in the spectrum. It is possible that this may be done in a manner in which time reversal invariance is violated explicitly or it may be spontaneously broken by fluctuations. In the field theory language, one is asking whether a Parity Anomaly is present. This problem is exactly the same one we have already encountered in our study of the Chiral Spin Liquid (chapter 7). For lattice systems one has to deal with the (infamous) "doubling" problem (here it is Q fold!). In most cases one expects no anomalies unless a perturbation which breaks time reversal is explicitly introduced. We are going to see in the next section that these issues are quite relevant for our problem.

Thus, the saddle-point approximation to the partition function yields Laughlin's result that the mean-field-theory for the anyon system should be equivalent to a system of particles (say fermions) moving in an effective magnetic field determined by their density. It is clear that this approximation assumes that the flux subsystem is rigid in the sense that the *average field*, determined at the saddle-point level, will not be modified by the fluctuations. In this high-density-like approximation, the fluctuations around the average field $\bar{\mathcal{B}}$ should be small for this approximation to be stable. The local value of the field is however still being determined by local fluctuations of the density. In this sense, the system is *compressible*. If the local fluctuations are massive, the spectrum should have a gap and the system will truly be rigid. But, if the fluctuations have a gapless state, the system will not be rigid. Indeed, this "fluctuation-induced compressibility" is the very origin of the superfluidity.

8.5 Fluctuations, Topological Invariance and Effective Action

8.5.1 Effective Action

In the past section we discussed the saddle-point approximation to the path-integral for the anyon gas. Fetter, Hanna and Laughlin claim that the fluctuations around the state with average flux $\bar{\mathcal{B}} = \frac{2\pi}{Q}$ induce a pole in the

current-current correlation function which, in turn, is responsible for the superfluidity. This is the FHL Goldstone boson . At first sight, this result seems to be quite surprising. In fact, fermions in a *background* magnetic field always lead to a spectrum with a gap, as in Quantum Hall Effect systems. What is different here is that the magnetic fields *do not constitute a fixed background*, since they are generated by the particles themselves. The fluctuations of the system retain this character. The SPA only fixes the average field, not its fluctuations and one is led to study the effects of fluctuations of the statistical gauge fields about the mean field. It is natural to compute the effective action of the statistical gauge fields including the effects of fermion loops. In this sense this calculation is close to the standard RPA.

Purely on the grounds of gauge and translation invariance, we can assert that the effective action for the statistical gauge fields at low energies and long-distances (*i.e.*, to leading order in a gradient expansion) should only be a function of the fluctuating part of $\mathcal{F}_{\mu\nu}$ (with the smallest number of gradients) plus a term with the same form as the bare Chern-Simons term. Banks and Lykken [Banks 90] argued that if the effective action has an induced Chern-Simons term which happens to cancel the bare one, then the FHL Goldstone boson is found and it is nothing but the *massless transverse* component of the fluctuating statistical vector potential. However, it is necessary to explain why does this crucial cancellation, present to leading order, survive renormalization. This is in fact far from obvious since the coefficients of the other terms do get renormalized.

Let us now investigate how does the FHL Goldstone pole appear within this path-integral framework. We will see that the exactness of the FHL Goldstone boson is a consequence of the *topological invariance* of the quantized Hall conductance for this system of fermions. Thus our problem is naturally related to the IQHE *integral* Hall Effect on lattices. In fact, we are going to be using many results of the theory of the Quantum Hall Effect. Most of these results are discussed in chapter 9.

Let us first consider the quadratic (*i.e.*, gaussian) fluctuations around the mean field. The effective action for the fluctuating part of the statistical gauge fields, hereafter denoted by $\mathcal{A}_\mu$, to quadratic order, $S^{(2)}$, is given by

$$S^{(2)}[\mathcal{A}_\mu] = \sum_{x,x'} \frac{\delta^2 S_{\text{eff}}}{\delta \mathcal{A}_\mu(x) \delta \mathcal{A}_\nu(x')} \mathcal{A}_\mu(x) \mathcal{A}_\nu(x') \tag{8.5.1}$$

where $x \equiv (\vec{x}, t)$ and $x' \equiv (\vec{x}\,', t')$, $\vec{x}$ and $\vec{x}\,'$ take values on the square lattice and t, t' are continuous (time) variables. Since S_{eff} is a sum of a fermionic part and a Chern-Simons term, $S^{(2)}$ also is a sum of two terms. The first term, which comes from the fermion loops, is nothing but the *polarization operator* $\Pi_{\mu\nu}(x, x')$. The second term is just the Chern-Simons term itself

$$S^{(2)}[\mathcal{A}_\mu] = \sum_{x,x'} \Pi_{\mu\nu}(x, x') \mathcal{A}_\mu(x) \mathcal{A}_\nu(x') - \frac{\theta}{4} S_{\text{CS}}[\mathcal{A}_\mu]. \tag{8.5.2}$$

$\Pi_{\mu\nu}(x, x')$ is the polarization operator for a system of fermions on a lattice in the presence of the background magnetic field $\bar{\mathcal{B}}$. Thus it is just the usual

linear response theory kernel, the current-current correlation function. The long-distance, low-energy behavior of $S^{(2)}$ can be obtained simply by noting that it has to satisfy the requirements of translation and gauge invariance. If there is a gap in the spectrum, $\Pi_{\mu\nu}$ is also local and it has a gradient expansion. Thus the effective action for fluctuations at distances larger than the interparticle separation and energies less than the gap has the form

$$S^{(2)}[\mathcal{A}] \approx \int d^2x dt \left[\frac{\epsilon}{2}\vec{\mathcal{E}}^{\,2}(\vec{x},t) - \frac{\chi}{2}\mathcal{B}^2(\vec{x},t) + \frac{1}{4}(\sigma_{xy}-\theta)\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu\mathcal{F}^{\nu\lambda}\right] + \text{h.o.t.} \tag{8.5.3}$$

where ϵ, χ and σ_{xy} are the (long-wavelength, low frequency) dielectric constant, diamagnetic susceptibility and Hall conductance of the Fermi system respectively. Note that the term which contains the Hall conductance has the *same form* but *opposite sign* as the bare Chern-Simons term which determines the statistics of the anyons.

The parameters (ϵ, χ and σ_{xy}) are in principle determined by integrating out all fluctuations from the highest energies allowed in this problem down to the only physical scale this system has: the gap. One expects that these coefficients will be heavily renormalized away from their saddle-point values. For the "almost-fermion" limit of large θ, the renormalizations are expected to be small, of order $\frac{1}{\theta}$. Thus, although explicit expressions for these coefficients can be found (they are given by various pieces of the polarization operator $\Pi_{\mu\nu}$), their precise form is not in principle very important due to the above mentioned renormalization effects.

While these considerations apply to ϵ and χ, as well as to the higher order terms in the effective action which we have neglected, the value of σ_{xy} *is completely determined already at the saddle-point level.* This is so because σ_{xy} is the Hall conductance for a system of fermions on a lattice, with an integer number f of Landau bands exactly filled, which has been shown to be *quantized.*

8.5.2 Quantized Hall Conductance and Compressibility

The quantization of σ_{xy} has been studied extensively in the contex of the Quantum Hall Effect. Thouless, Kohmoto, den Nijs and Nightingale (TKNN) [Thouless 82] have shown that the σ_{xy} associated with the Hofstadter problem is quantized in terms of an integer $t, \left(-\frac{Q}{2} < t < \frac{Q}{2}\right)$, which, in turn, is determined by a Diophantine equation. The theory of TKNN is discussed in chapter 9. The following results are relevant to our problem. If j denotes the j-th gap of a Hofstadter problem with $\bar{\mathcal{B}} = \frac{2\pi P}{Q}$, there exist two integers t_j and s_j (with t_j in the same range as t and s_j unconstrained) such that

$$j = Qs_j + Pt_j. \tag{8.5.4}$$

If the Fermi energy lies in the f-th gap, the Hall conductance is given by

$$\sigma_{xy} = \frac{e^2}{\hbar} \sum_{j=1}^{f} (t_j - t_{j-1}) \tag{8.5.5}$$

with $t_0 = 0$. Thus, in units in which $e^2 = \hbar = 1$, we find

$$\sigma_{xy} = \frac{t_f}{2\pi} \tag{8.5.6}$$

where t_f is the solution of the Diophantine equation for the f-th gap. We may now combine these results to get

$$\sigma_{xy} = \theta(1 - \frac{s_f}{\rho}). \tag{8.5.7}$$

The Diophantine equation has solutions in the form of a pair of integers (s_f, t_f). The solution is, in most cases, unique and, in general, both s_f *and* t_f will be different from zero. Under special circumstances, we will find families of solutions with $s_f = 0$. Also, in some special cases, the solution is not unique. The solutions with $s_f = 0$ play a special role for, as we will see, they represent the *compressible* states .

Let us first consider the sequence $\theta = \frac{m}{2\pi}$ and ρ arbitrary. The mean field theory yields the values $P = \frac{r}{(m,r)}$, $Q = \frac{mq}{(m,r)}$ and it requires that exactly $f = \frac{mr}{(m,r)}$ Landau bands are filled. The Diophantine equation has, for $j = f$, the *unique* solution

$$\begin{cases} s_f = 0, \qquad t_f = m, & \text{if } |m| < \frac{mq}{2(m,r)}, \\ s_f = \frac{r}{(m,r)}, t_f = m - \frac{mq}{(m,r)}, & \text{if } |m| > \frac{mq}{2(m,r)}. \end{cases} \tag{8.5.8}$$

There are degenerate solutions whenever $|m| = \frac{mq}{2(m,r)}$. In this case both solutions are possible and the value of t_f is ambiguous. It is easy to see that a degeneracy occurs whenever $m = \frac{mq}{2(m,r)}$, *i.e.*, for $q = 2(m,r)$ and q is even. This includes the half-filled even-denominator case $\rho = \frac{1}{2}$. Which solution is realized depends on how this degeneracy is lifted by additional terms in the Hamiltonian. It is natural to assume that it is always possible to find terms which will remove this degeneracy. The physical properties of the system will depend on the way we choose to render the system non-degenerate.

Thus, in the absence of degeneracies, the solution is unique and one finds $t_f = m$ and $s_f = 0$ if $q > 2(m,r)$. Hence, we get

$$\sigma_{xy} = \frac{m}{2\pi} \tag{8.5.9}$$

which is exactly identical to θ! We then conclude that, at least at the level of the saddle-point approximation and in the absence of degeneracies, $\sigma_{xy} = \theta$ and the Chern-Simons term in the effective action for the fluctuating statistical gauge fields is cancelled provided that $q > 2(m,r)$. As Banks and Lykken observed, this is a *sufficient* condition for the existence of the FHL Goldstone boson. Conversely, for $q < 2(m,r)$, the solution has $s_f = \frac{r}{(m,r)} \neq 0$, and there is no cancellation and no FHL Goldstone boson .

For other sequences, such as $n \neq 1$, it is not possible to find a solution of the Diophantine equation with $s_f = 0$. It easy to check that this solution exists only if n is a factor of m, which is impossible since $(n, m) = 1$ except for the case $n = 1$. Thus, the other sequences do not exhibit superfluidity. These non-superfluid states cannot be found in the continuum theory. They are the result of *diffraction* effects generated by the underlying lattice. It is clear that, in the low-density limit, these effects do not impose an overwhelming constraint, provided that the Fermi energy lies in one of the main energy gaps. In this case, there is a smooth continuum limit at low densities. However, if the Fermi energy is in one of the lower gaps, we will not get a cancellation, even in the low density limit. Thus, the continuum limit is tricky to get. We should then expect that the properties of the ground state should depend on some details of the behavior of the system at short distances. This problem will come back when we consider the role of higher order fluctuations.

8.5.3 Stability of the Mean Field State

One might wonder about the stability of this crucial result once fluctuations about mean field theory are considered. Two problems naturally arise. Firstly one must worry about infinite renormalizations. In continuum relativistic field theories it is known that the Chern-Simons term does not acquire infinite renormalizations [Semenoff 89]. Non-relativistic theories are not expected to be any more singular. Thus divergent renormalization of σ_{xy} are not to be expected. However, *finite* renormalizations are not excluded by such arguments. The stability of the FHL Goldstone boson requires no renormalization at all, neither infinite nor finite.

No-renormalization theorems usually follow from symmetry considerations or as a result of topology (or both). For the case of the lattice system, Kohmoto and Avron, Seiler and Simon showed that σ_{xy} is a topological invariant (see chapter 9). The topological invariance of σ_{xy} follows from the fact that the Brillouin zone of a two dimensional system with periodic boundary conditions is a two torus. The integer t_m is the first Chern number of the fiber bundle associated with the Berry connection induced by the wave functions on the two-torus. Small changes in the microscopic Hamiltonian will not change this number provided that no band crossings occur as a result of such changes. Qualitatively speaking, fluctuations about a solution with a finite gap are expected to have the same effect. After all, the fluctuations, configuration by configuration, will modulate the gap. Since each configuration yields the *same* value for σ_{xy}, the final result should be the same, provided that the sum over configurations makes sense. Once again, this argument requires the existence of a non-zero energy gap. Niu, Thouless and Wu have also given an argument for the stability of the quantization of σ_{xy} including many-body effects (*i.e.*, fluctuations). They showed that if the *many-body wave-function* for the ground state winds by the phases α and β along the x_1 and x_2 directions, then the value of σ_{xy}^{av}, *averaged over* α *and* β, is a topological invariant

and hence it is quantized in the full theory. For a system with specific boundary conditions, say periodic, they showed that $\sigma_{xy}^{\rm pbc}$ differs from $\sigma_{xy}^{\rm av}$ by terms which vanish exponentially fast in the thermodynamic limit provided that the system has a non-zero energy gap. For more details, see the discussion in chapter 9.

I thus conclude that the topological invariance of σ_{xy} guarantees that the FHL Goldstone boson is stable to all orders in perturbation theory.

8.5.4 Low Energy Spectrum

We must then conclude that the anyon gas can exist in only one of two possible states, each defined by a low-energy effective action of the form of a QED-type theory with possibly a Chern-Simons term with some effective coupling. For the case of the "happy fractions", $\theta = \frac{m}{2\pi}$ and $\rho = \frac{r}{q}$, the effective action does not have a net Chern-Simons term. The effective action has the form

$$S^{(2)}[\mathcal{A}] = \int d^2x dt \left[\frac{\epsilon}{2}\vec{E}^2(\vec{x},t) - \frac{\chi}{2}B^2(\vec{x},t)\right] + \text{h.o.t.} \tag{8.5.10}$$

which is the action of free "Maxwell" electrodynamics in 2+1 dimensions. Here I have neglected terms which vanish in the infrared limit. Some of these terms violate time reversal and parity explicitly. However their effects are very small and do not affect our main conclusions.

This theory has only one transverse degree of freedom, the "photon". Note that this has nothing to do with the *real* electromagnetic field. It originates from the fluxes associated with the anyons. This "photon" is the only massless excitation of this theory. It is precisely the FHL Goldstone boson. It is responsible for both the phase mode necessary for superconductivity and for a direct Coulomb-like static interaction among sources (or excitations) which couple to the *statistical* gauge field. At long distances, the 2+1 dimensional Coulomb interaction goes like $\ln R$ where R is the separation between two *sources* of the field $\mathcal{A}$. Thus, the energy necessary to create a *fermion* diverges logarithmically with the size of the system. The same happens with the energy to add a flux to the system.

An anyon, however, is a gauge-invariant state. As such it only couples weakly to the fluctuations of the statistical gauge field since it is neutral but not quite point-like. Thus, we expect that the energy of an anyon-like excitation be finite and its value be determined primarily by short-distance effects. Let us consider an operator which creates an anyon at point $\vec{x}$. It is easy to compute correlation functions of these gauge invariant operators in the Coulomb (or *anyon*) gauge. In this gauge, we can write

$$\mathcal{A}_j(\vec{x},t) = \epsilon_{jk}\Delta_k\phi(\vec{x},t) \tag{8.5.11}$$

where ϕ also obeys periodic boundary conditions. If we now substitute eq. (8.5.11) back into eq. (8.3.4) then, after an integration by parts the argument of the exponential, we find

$$K(\vec{x},t) = e^{i\theta\Sigma_{\vec{y}}\phi(\vec{y},t)\Delta_j\mathcal{A}_j^c(\vec{y})}. \tag{8.5.12}$$

If we also choose the Coulomb gauge to describe the classical fluxes, *i.e.*, $\Delta_j \mathcal{A}_j^c = 0$, we see that, *in this gauge*, $K(\vec{x},t)$ is equivalent to the identity operator. Thus, the correlation function for anyon operators is, in the Coulomb gauge, the same as the (gauge dependent) fermion propagator evaluated in the same gauge. In the Coulomb gauge, the fermion propagator has the following properties: (a) it is multivalued and (b) it is short ranged. It is multivalued, since the one-particle wave-functions are multivalued in this gauge. It is short-ranged, since the ground state has *filled* Landau bands and the only possible one-particle states available are in the next unfilled Landau band. These states are separated from the ground state by the energy gap between Landau bands which is *finite*.

In contrast, the elementary fermion excitations have a logarithmically divergent energy. This is so because the operators that create these states are not gauge invariant, reflecting the fact that these are not neutral states. A gauge-invariant fermion operator can be defined. This is achieved by inserting the usual exponential of the line integral, along some path Γ, of the statistical vector potential between a pair of fermion creation and anihilation operators some distance apart from each other:

$$c^\dagger(\vec{x},t)e^{i\int_\Gamma \mathcal{A}_\mu dx^\mu} c(\vec{x}\,',t'). \tag{8.5.13}$$

The massless "photon" gives rise to a logarithmically divergent fermion self-energy. A similar treatment can be given to flux states. The operator K which creates fluxes is not gauge invariant. A way to make it invariant is to multiply this operator by a fermion operator which represents anyons, not fermions or fluxes. However, it is still possible to multiply $K(\vec{x},t)$ by a line integral, just as in the fermion case. The resulting operator is a boson and it is manifestly gauge invariant. The one-particle states created by these operators also have logarithmically divergent energy and exactly for the same reason: the exchange of massless "photons".

Let us end this section by briefly considering the state in which the effective action has a non-zero effective Chern-Simons term. I will call this phase the Quantum Hall state. The effective Chern-Simons coupling constant $\bar{\theta}$ equals

$$\bar{\theta} = -s_f \frac{\theta}{\rho}. \tag{8.5.14}$$

Thus, a non-zero s_f means $\bar{\theta} \neq 0$. A theory with a non-zero Chern-Simons coupling constant is known to contain a *massive* photon. The mass of the photon is proportional to $\bar{\theta}$ and hence it is determined by s_f. Thus, the Quantum Hall State has short range *gauge* interactions mediated by the statistical gauge field. These fluctuations are effectively supressed and the state is effectively incompressible. We will see in the next section that *both* Superfluid and Quantum Hall states exhibit a *quantized* Hall effect. While this is to be expected in a Quantum Hall state it is quite a surprising result for a superfluid state.

8.6 The Anyon Superfluid State

We want to understand the superconducting properties of the anyon system. In the last section, an effective low-energy action for the statistical gauge fields was derived. We saw that the spectrum of this theory contains a massless excitation. I now give an argument which shows that this system is in fact a superconductor.

8.6.1 Duality Transformation and Landau-Ginzburg Action

Let us consider the effective low-energy action in the presence of weak, slowly varying electromagnetic fields A_μ. The effective action now has the form

$$S_{\text{eff}} = \int dtd^2x \left(\frac{\epsilon}{2} \vec{\mathcal{E}}^{\,2}(\vec{x},t) - \frac{\chi}{2} \mathcal{B}^2(\vec{x},t) - \frac{\theta}{2} \epsilon_{\mu\nu\lambda} \mathcal{A}_\mu(\vec{x},t) F_{\nu\lambda}(\vec{x},t) \right) - \frac{\theta}{4} S_{\text{cs}}[A] \tag{8.6.1}$$

where A is the electromagnetic field.

Let us now consider the *dual* theory. Here we understand duality in the statistical mechanical sense in which a *gauge* theory in 2+1 dimensions is *dual* to a theory with a *global* symmetry. Since the gauge field of this problem, the statistical gauge field, has a $U(1)$ symmetry, its dual is a *phase* field. Let $\Lambda_{\mu\nu}(\vec{x},t)$ be a real antisymmetric tensor field. Since we are dealing with an anisotropic theory, it is convenient to define $\Lambda_{0i} = \vec{e}_i$ and $\Lambda_{ij} = \epsilon_{ij} b$, where $\vec{e}_i$ and b are real functions of space and time.

Consider now the modified action S'

$$S' = \int dtd^2x \left(-\frac{1}{2\epsilon} \vec{e}^{\,2} + \frac{1}{2\chi} b^2 + \frac{1}{2} \Lambda_{\mu\nu} \mathcal{F}^{\mu\nu} - \frac{\theta}{4} \epsilon_{\mu\nu\lambda} A^\mu \mathcal{F}^{\nu\lambda} \right) - \frac{\theta}{4} S_{\text{cs}}[A]. \tag{8.6.2}$$

We can identify the path integrals with actions S and S' after a shift of the gaussian variables $\Lambda_{\mu\nu}$, except for an irrelevant constant. The fluctuating statistical gauge fields $\mathcal{A}_\mu$ can now be integrated out, yielding the constraint on the $\Lambda_{\mu\nu}$ fields

$$\partial^\mu \left(\Lambda_{\mu\nu} - \theta \epsilon_{\mu\nu\lambda} A^\lambda \right) = 0. \tag{8.6.3}$$

This constraint can be solved by means of the phase field ω defined by

$$\frac{1}{2\pi} \epsilon_{\mu\nu\lambda} \partial^\lambda \omega = \Lambda_{\mu\nu} - \theta \epsilon_{\mu\nu\lambda} A^\lambda. \tag{8.6.4}$$

By substituting back into the effective action, we get the effective Lagrangian density in terms of the ω field

$$\mathcal{L} = \frac{1}{8\pi^2\chi} \left(\partial_0 \omega + m A_0 \right)^2 - \frac{1}{8\pi^2\epsilon} \left(\partial_i \omega + m A_i \right)^2 + \text{h.o.t.} - \frac{\theta}{4} \epsilon_{\mu\nu\lambda} A^\mu F^{\nu\lambda}. \tag{8.6.5}$$

This effective Lagrangian was first obtained by Banks and Lykken. In this derivation I used the fact that $\theta = \frac{m}{2\pi}$. For systems in the Quantum Hall phase,

i.e., $\sigma^{xy} \neq \theta$, the dual theory contains a non-local term which effectively suppresses the fluctuations of the phase field ω.

8.6.2 Physical Picture and the Schwinger Mechanism

The first two terms of the effective action for the ω field are identical to the Lagrangian of a Landau-Ginzburg Theory, in the *London* limit, of a system with an order parameter $e^{i\omega}$, which is a single component scalar field with fixed amplitude. The last term is an induced Chern-Simons term for the electromagnetic field. All the effects of broken time reversal invariance are due to this term. Let us remark that the coefficient of this effective Chern-Simons term for the electromagnetic field is not universal. As a matter of fact, if we add a *bare* Maxwell term for the *statistical vector potential*, such as the one we derived in section 7.1 in the framework of the CSL, it is straightforward to see that its main effect is to change the value of the Chern-Simons coupling for the *electromagnetic field*. Under some circumstances, this coupling may effectively be equal to zero [Halperin 90].

This is a remarkable result. We have obtained an effective Lagrangian density for the low energy fluctuations (*i.e.*, a Landau Theory) of a superconductor *without* having to go through an intermediate *pairing* Hamiltonian! In other words, the anyon gas is a superconductor without the need of Cooper pairs. There has been some speculation in the literature that this theory may have "hidden pairs". Also, it has been claimed that the anyon superconductor does not have flux quantization. Neither of these statements are correct. The anyon gas is a superconductor which does not require Cooper pairing as a microscopic mechanism and which does exhibit flux quantization.

There is a simple physical way to understand these issues. The abscence of *explicit* Copper pairing is essentially a kinematic effect. The condensed state can be viewed as a set of *fermions* moving in a uniform magnetic field with flux $\bar{\mathcal{B}} = \frac{\rho}{\theta}$. There are an integer number f of *filled Landau bands*. In other words, there is an integer number of *Landau orbits* which are completely filled with *fermions* (not anyons!) very much in the same way as there are filled shells in atomic physics. The fermions, which carry the statistical gauge charge, and hence are not gauge-invariant, move on circles of radii equal to the cyclotron radius of the average magnetic field $\bar{\mathcal{B}}$. This state is *coherent*, which means that *all* the fermions rotate in phase at the same angular frequency, the cyclotron frequency ω_c, which is equal to the gap. The average distance between *any* pair of fermions is thus constant in time and we expect pairing correlations to be present in the system even though we did not have to solve a "Cooper problem" to make the superconductivity possible.

Thus, the superconductivity of the anyon gas is not due a microscopic BCS mechanism by which the particles would form bosonic bound states which later form a Bose condensed ground state. The BCS picture assumes that the fermions are good quasiparticles which would form a Fermi surface if superconductivity would not have occured. The instability in BCS theory is essentially caused by short range attractive forces. In contrast, the *fermions*

(or *bosons*) which become anyons when coupled to the Chern-Simons gauge field, are always experiencing a long range gauge force. The "normal state excitations" of the mean-field-theory of FHL do not have a Fermi surface. Rather, they fall into effective Landau levels. The best the system can do to lower its energy is to adjust the field to the particle density at every point.

Furthermore, the FHL Goldstone boson is a charge-carrying *transverse* excitation of this state. Imagine first a set of filled Landau orbits, with radii fixed by the field (and hence by the density), filling-up the plane. Imagine next a density fluctuation, by which a particle is removed from an orbit, centered around some point $\vec{x}_0$, and added to another orbit located some distance away say on the x axis. At the point in which we *removed* a particle, the density has been *decreased*, the effective magnetic field also *decreased* (because of the constraint) and, hence, the radius of the orbit has *increased*. Conversely, at the point at which the particle was *added*, the density and the field have *increased* but the radius of the orbit has *decreased*. We can now imagine a wave of this sort in which the *fermions* move on closed orbits whose radii excecute an oscillatory motion as the wave goes bye. The motion of the fermions is *transverse* to the direction of propagation of the wave. The calculation of the effective action shows that this wave is a *massless* mode, the FHL Goldstone boson.

That there is flux quantization is self-evident by inspection of the Lagrangian. The phase of the order parameter field, ω, is *minimally coupled* to the electromagnetic field A_μ with an effective charge me, where e is the electric charge. Thus, we get a flux quantum $\phi_0 = \frac{hc}{me}$. Barring a microscopic derivation of the anyon Hamiltonian, which should yield a value of θ, we can only argue that phenomenology requires $m = 2$ and, hence, semions.

This picture is strongly reminiscent of the *Schwinger mechanism* for the generation of mass to a gauge field in $1 + 1$ dimensions coupled with massless fermions, the Schwinger Model. Schwinger argued that this system has a spectrum equivalent to that of a *massive* boson. In his analysis, in which the chiral anomaly plays a crucial role, the density fluctuations of the fermions couple to the longitudinal (and only!) fluctuations of the vector potential which becomes massive. An alternative picture is found by *bosonization* of the Schwinger Model which makes the equivalence with a theory of a single massive scalar field apparent. The origin of this phenomenon can be traced back to the kinematical restrictions implied by the fact that the space dimension is equal to one. Any pair of massless particles in one dimension which move in the same direction have the same speed. Thus, their distance is constant in time, up to quantum fluctuations. It is indeed this special property of massless fermion in one dimensional space which makes bosonization work. In higher dimensions the situation is more complicated. The vector potential has both longitudinal and transverse componenents, *both* of which become massive in a superconductor: the longitudinal component is massive due to *screening* whereas the transverse component is massive (Meissner effect) by "eating" the phase degree of freedom of the order parameter field. This mechanism is usually called the Higgs mechanism. This phenomenon was discussed

in precisely these terms by P.W. Anderson in 1963 who drew an analogy with the Schwinger mechanism. However, in dimensions of space larger than one, there aren't enough density degrees of freedom to make *all* the components of the vector potential massive, unless a Cooper instability is present. This is the BCS mechanism. In this sense, anyon superconductivity is a higher dimensional analog of the Schwinger mechanism and, in fact, it is the only one known to this date!

8.6.3 Electrodynamics of Anyon Superconductivity

The effective Lagrangian density exhibits several features which are characteristic of a superconductor: a) Meissner effect, b) screening and c) supercurrent. However the last term, which is the result of the breakdown of Time-Reversal Invariance, modifies this picture in some significant ways.

Let us now include the dynamics of the electromagnetic field A_μ in the problem. For simplicity I will just add an extra term to the Lagrangian density to represent the dynamics of the electromagnetic field *in the plane*. The new Lagrangian density, keeping just leading order in gradients is

$$\mathcal{L} = \frac{1}{8\pi^2\chi}\left(\partial_0\omega + mA_0\right)^2 - \frac{\epsilon}{8\pi^2}\left(\partial_i\omega + mA_i\right)^2 - \frac{\theta}{4}\epsilon_{\mu\nu\lambda}A^\mu F^{\nu\lambda} - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu}. \tag{8.6.6}$$

The main drawback of this Lagrangian density is that the dynamics of the electrodynamic gauge fields is purely two-dimensional. There are only two space directions and two space components of the electromagnetic field. This is not physically correct, unless one wants to study a purely (2+1)-dimensional model. A more realistic model requires the addition of the third direction of space and the third spacial component of the electromagnetic field. It is not possible, for the moment, to write such a model since it would require to know how anyons in different planes of a three-dimensional system couple to each other. Clearly, there is no way for an anyon to meaningfully exist in three dimensions. Anyons only exist on a given plane. Thus, a simple anyon-hopping is not allowed. In the particular case of semions, bound states of two anyons are bosons (in general an m-anyon bound state is a boson), thus they can hop (or tunnel) together. Also, a direct electromagnetic coupling between planes is certainly allowed. Which one of these possibilities (or even others) are realized, requires a more detailed microscopic theory which, presently, is not available. Thus I will consider the above simpler choice for the Lagrangian density while keeping in mind that we are only allowing for electromagnetic fluctuations, polarized in the plane (both longitudinal and transverse) and propagating only in the plane.

Let us now derive the classical equations of motion in the Coulomb gauge. In this gauge, the electromagnetic vector potential can be written in the form

$$\begin{aligned} A_i &= \epsilon_{ij}\partial_j\Phi, \\ \epsilon_{ij}\partial_i A_j &= \nabla^2\Phi. \end{aligned} \tag{8.6.7}$$

The equations of motion are

$$\begin{aligned}
&\left(\partial_0^2 - \frac{\chi}{\epsilon}\nabla^2\right)\omega + m\partial_0 A_0 = 0, \\
&\frac{\theta}{2\pi\chi}\left(\partial_0\omega + mA_0\right) + \theta\nabla^2\Phi - \frac{1}{e^2}\nabla^2 A_0 = -J_0^{\rm ext}, \\
&\nabla^2\left(\theta A_0 - \frac{1}{e^2}\nabla^2\Phi\right) + \frac{\theta^2}{\epsilon}\nabla^2\Phi + \frac{1}{e^2}\partial_0^2\nabla^2\Phi = \epsilon_{ij}\partial_i J_j^{\rm ext},
\end{aligned} \tag{8.6.8}$$

where $J_0^{\rm ext}$ and $J_i^{\rm ext}$ represent an external distribution of charges and currents which we will use to probe the system.

8.6.4 Screening of a Static Charge:

Consider first the case of a static probe of electric charge Q_0, located at the origin, and zero external currents

$$\begin{aligned}
J_0^{\rm ext} &= Q_0\delta^2(\vec{x}), \\
J_i^{\rm ext} &= 0.
\end{aligned} \tag{8.6.9}$$

We look for static (*i.e.*, time-independent) solutions of the equations of motion with sources . In this case, the equations of motion take the form

$$\begin{aligned}
&\frac{\chi}{\epsilon}\nabla^2\omega = 0, \\
&\frac{\theta^2}{\chi}A_0 + \theta\nabla^2\Phi - \frac{1}{e^2}\nabla^2 A_0 = -Q_0\delta^2(\vec{x}), \\
&\theta A_0 - \frac{1}{e^2}\nabla^2\Phi + \frac{\theta^2}{\epsilon}\Phi = 0.
\end{aligned} \tag{8.6.10}$$

The solution gives A_0 and Φ in terms of the potential $U(\vec{x})$

$$\begin{aligned}
A_0 &= -Q_0 e^4\left(\frac{\theta^2}{\epsilon} - \frac{1}{e^2}\nabla^2\right)U(\vec{x}), \\
\Phi &= Q_0\theta e^4 U(\vec{x}).
\end{aligned} \tag{8.6.11}$$

The potential $U(\vec{x})$ is a linear combination of the two-dimensional static Green function $G_0(\vec{x}; M)$

$$U(\vec{x}) = \frac{1}{M_+^2 - M_-^2}\Big(G_0(\vec{x}; M_-) - G_0(\vec{x}; M_+)\Big) \tag{8.6.12}$$

where

$$\left(-\nabla^2 + M_\pm^2\right)G_0(\vec{x}, M_\pm^2) = \delta^2(\vec{x}) \tag{8.6.13}$$

and $M_\pm$, which we will see below are related to the masses of longitudinal and transverse modes, are given by

$$M_\pm^2 = \frac{\theta^2 e^2}{2\chi}\left((1+\frac{\chi}{\epsilon}+e^2\chi) \pm \left((1-\frac{\chi}{\epsilon})^2 + e^4\chi^2 + 2e^2\chi(1+\frac{\chi}{\epsilon})\right)^{\frac{1}{2}}\right). \tag{8.6.14}$$

The Green function G_0, for $|x| = R$, is

$$G_0(R) = \frac{1}{2\pi}K_0(MR) \tag{8.6.15}$$

where K_0 is a Bessel function. At long distances, $MR \gg 1$, G_0 decays exponentially fast

$$G_0(R,M) \approx \frac{e^{-MR}}{(8\pi MR)^{\frac{1}{2}}}. \tag{8.6.16}$$

Thus, the electric field $\vec{E}$ and the magnetic field B induced by the static charge Q_0, are

$$\begin{aligned} \vec{E} &= -\vec{\nabla}A_0 = Q_0 e^4(e\theta^2 - \frac{1}{e^2}\nabla^2)\nabla U, \\ B &= -Q_0\theta e^4\nabla^2 U. \end{aligned} \tag{8.6.17}$$

Therefore a static electric charge Q_0 is screened over a distance ξ, the screening length,

$$\xi \approx \frac{1}{M_-}. \tag{8.6.18}$$

Notice that *both* electric and magnetic fields are present in the vicinity of the charge, although the magnetic field certainly decays faster than the electric field does. These results indicate that the integrated induced magnetic flux on the entire sample is zero. The induced field is largest at a distance of the order of the screening length itself. This result is a consequence of the induced Chern-Simons term and was first derived by Paul and Khare [Paul 86] and by Lozano, Manias and Schaposnik [Lozano 88]. It is important to stress that this calculation is strictly two-dimensional and it does not apply to the problem of a single plane embedded in three dimensions since in that case the field would escape from the plane. However, it is the calculation of the screening field surrounding a line of charge in a three-dimensional stack of planes. The more physical situation of a single charge in three dimensions cannot be addressed by the two-dimensional anyon model since it requires to make assumptions about the behavior of a stack of planes with the anyon confined to the planes. Presently, there is no satisfactory microscopic model which could describe the behavior of a three dimensional system with anyons confined to its planes.

8.6.5 Longitudinal and Transverse Masses: Meissner Effect

Let us now solve the equations of motion in the absence of external sources. We are looking for the normal modes, namely for plane wave solutions which represent electromagnetic waves (both longitudinal and transverse) polarized

and propagating *in the plane*, with momentum $\vec{p}$, frequency p_0 and amplitudes $\bar{\omega}$, $\bar{\Phi}$ and $\bar{A}_0$. These amplitudes are related to each other by

$$\begin{aligned} \bar{\omega} &= \frac{-imp_0}{p_0^2 - \frac{\chi}{\epsilon}\vec{p}^2}\bar{A}_0, \\ \bar{\Phi} &= \frac{\theta e^2}{p_0^2 - \vec{p}^2 - e^2\frac{\theta^2}{\epsilon}}A_0. \end{aligned} \tag{8.6.19}$$

The solution of the equations of motion has two branches. In terms of E_1^2 , E_2^2 and Δ defined by

$$\begin{aligned} E_1^2 &= \vec{p}^{\,2} + e^2\frac{\theta^2}{\epsilon}, \\ E_2^2 &= \frac{\chi}{\epsilon}\vec{p}^2, \\ \Delta &= \frac{1}{2}\left(E_1^2 + E_2^2 + \theta^2 e^2(e^2 + \frac{1}{\epsilon})\right), \end{aligned} \tag{8.6.20}$$

the energy-momentum relations are

$$p_0^2 = \Delta \pm \sqrt{\Delta^2 - E_1^2E_2^2 - \frac{\theta^2}{\epsilon}e^2E_1^2 - \theta^2e^4E_2^2}. \tag{8.6.21}$$

There are two interesting limits to consider: i) $\vec{p}^{\,2} \to \infty$ and ii) $\vec{p}^{\,2} \to 0$. In the limit of *large* momentum, the disperion relation becomes linear, but the speed of light $c_\pm$ is different for each solution

$$|p_0| \approx c_\pm|\vec{p}|. \tag{8.6.22}$$

for $|\vec{p}| \to \infty$, where the speed of light $c_\pm$ is

$$\begin{aligned} c_+ &= 1, \\ c_- &= \sqrt{\frac{\chi}{\epsilon}}. \end{aligned} \tag{8.6.23}$$

We identify the solution with the positive root, which represents waves moving at the speed of light in the vacuum, with the transverse photon mode. The solution with the negative root, represents waves propagating at speed $\sqrt{\frac{\chi}{\epsilon}} < 1$, and we identify it with the plasmon, the longitudinal photon mode.

In the opposite limit, $|\vec{p}| \to 0$, both branches intercept the frequency axis at *different* values of p_0, *i.e.*, $M_\pm$. The value $p_0 = M_+$ is the *mass* of the transverse photon, whereas $p_0 = M_-$ is the *plasma frequency* associated with the plasmon. In a *Time-Reversal Invariant* superconductor, these two frequencies coincide($M_+ = M_-$). In a *Time-Reversal Noninvariant* superconductor, such as the anyon gas, these frequencies *do not* coincide. Indeed, the splitting of the square of the frequencies δM^2 is a measurement of the violation of Time-Reversal invariance,

$$\delta M^2 = \theta^2e^4\sqrt{1 + \frac{4}{e^2\epsilon}}. \tag{8.6.24}$$

This is a very important prediction. This splitting should be observable in light-scattering experiments in which the light beam propagates in the plane of the anyon gas, *i.e.*, the *ab* plane of the CuO_2 superconductor, and it is also polarized in the plane. What is important is that this spectroscopic probe is insensitive to the pattern in which Time-Reversal invariance is broken. For example, if different CuO_2 planes have different values of θ, such as a staggered pattern, the splitting will still be there. Other probes, such as the rotation of the angle of polarization of a beam perpendicular to the plane or the Condensate Hall effect, which we will discuss next, will cancel out for such patterns of symmetry breaking.

8.6.6 Supercurrent and Condensate Hall effect

Let us finally discuss the supercurrent and the Condensate Hall effect. The Lagrangian implies that the total current J_i of this system is

$$J_i \equiv -\frac{\delta \mathcal{L}}{\delta A_i} = \frac{\theta}{2\pi\epsilon}(\partial_i \omega + m A_i) - \frac{\theta}{2}\epsilon_{ij} E_j \tag{8.6.25}$$

where E_j is the electric field. The first term of J_i is clearly the *London current* of a standard superconductor. The second term, due to the induced Chern-Simons term, has the form of a *Hall current* with a Hall conductance $\sigma_{\text{eff}}^{xy} = -\theta = -\frac{m}{2\pi}$, if renormalization effects are ignored. This effect is *due to the condensate*, not to quasiparticles. Indeed, since we are studying the system at zero temperature, there are no quasiparticles left since this system has a gap. Also, this effect is taking place even though flux is expelled from the system due to the Meissner effect. This remarkable effect unfortunately is sensitive to the *sign* of θ and would not be observable if θ is staggered between different planes.

8.6.7 Vortices

The supercurrent derived above also implies that the *vortices* of this system are charged. Consider now the charge density J_0 associated with $\vec{J}$,

$$J_0 \equiv \frac{\delta \mathcal{L}}{\delta A_0} = \frac{\theta}{2\pi\chi}(\partial_0 \omega + m A_0) - \theta B. \tag{8.6.26}$$

A vortex is a configuration with non-trivial winding number for the phase ω of the order-parameter field over a very large circle $\mathcal{C}$ in space. A configuration with *total vorticity* k, with k integer, has $\Delta\omega = 2\pi$, $A_0 = 0$, it is time independent and has no current in the abscence of external fields. The total charge $\bar{Q}$ carried by the vortex is

$$\bar{Q} = \int d^2x J_0 = -\theta \int d^2x B = -\theta \int_{\mathcal{C}} dx_\mu A^\mu. \tag{8.6.27}$$

Using the zero-current condition, $\partial_i \omega + m A_i = 0$, we find

$$\bar{Q} = \theta \frac{\Delta\omega}{m} = \theta \frac{2\pi}{m} k = k. \tag{8.6.28}$$

In other words, the vortex is *charged* and carries an integer charge equal to the vorticity, $\bar{Q} = k$. By tracing back the duality transformation, it is easy to prove that to add a vortex with charge $\bar{Q} = +1$ is equivalent to add an extra slowly moving *charge* to the system (not an anyon!). This result is consistent with the fact that an extra charge has a logarithmically divergent self-energy, as any vortex should.

References

[Anderson 63] P.W. Anderson. Phys. Rev. 130, 439, (1963).

[Arovas 85] D. Arovas, J.R. Schrieffer, F. Wilczek and A. Zee. Nucl. Phys. 251B, 117, (1985).

[Banks 90] T. Banks and J.D. Lykken. Nucl. Phys. B336, 500 (1990).

[Canright 89] G. Canright, S. Girvin and A. Brass. Phys. Rev. Lett. 63, 2291, (1989).

[Chen 89] Y-H. Chen, F. Wilczek, E. Witten and B.I. Halperin. Int. J. Mod. Phys. B3, 1001, (1989).

[Fetter 89] A. Fetter, C. Hanna and R.B. Laughlin. Phys. Rev. B 39, 9679, (1989).

[Fisher 89] M.P.A. Fisher and D.H. Lee. Phys. Rev. Lett. 63, 903, (1989).

[Fradkin 90] E. Fradkin. Phys. Rev. B 42, 570, (1990).

[Girvin 87] S. Girvin and A. MacDonald. Phys. Rev. Lett. 58, 1252, (1987).

[Halperin 84] B.I. Halperin. Phys. Rev. Lett. 52, 1583, (1984).

[Halperin 90] B.I. Halperin, J. March-Russell and F. Wilczek. Phys. Rev. B 40, 8726, (1990).

[Hosotani 90] Y. Hosotani and S. Chakravarty. Phys. Rev. B 42, 342 , (1990).

[Kitazawa 90] K. Kitazawa and H. Murayama. Nucl. Phys. B338, 777, (1990).

[Laughlin 88] R.B. Laughlin. Phys. Rev. Lett. 60, 2677, (1988).

[Laughlin 88] R.B. Laughlin. Phys. Rev. Lett. 60, 2677, (1988).

[Lozano 88] G. Lozano, M. Manias and F. Schaposnik. Phys. Rev. D 38, 601, (1988).

[Martin 75] P.C. Martin. *Measurements and Correlation Functions*, Gordon & Breach 1975.

[Paul 86] S. Paul and A. Khare. Phys. Lett. B174, 420, (1986).

[Read 89] N. Read. Phys. Rev. Lett. 62, 86, (1989).

[Salam 90] S. Randjbar-Daemi, A. Salam and J. Strathdee. Nucl. Phys. B340 ,403, (1990).

[Schwinger 62] J. Schwinger. Phys. Rev. 125, 397, (1962).

[Semenoff 89] G.W. Semenoff, P. Sodano and Y.S. Wu. Phys. Rev. Lett. 62, 715, (1989).

[Thouless 82] D. Thouless, M. Kohmoto, M. denNijs and M. Nightingale. Phys. Rev. Lett. 49, 405, (1982).

[Wen 90] X.G. Wen and A. Zee. Phys. Rev. B 41, 240, (1990).

[Zhang 89] S.C. Zhang, T. Hannson and S. Kivelson. Phys. Rev. Lett. 62, 82, (1989).

CHAPTER 9

Topology and Quantum Hall Effect

In this chapter I discuss the problem of electrons moving on a plane in the presence of an external uniform magnetic field perpendicular to the system. This is a subject of great interest from the point of view of both theory and experiment. The explanation of the remarkable quantization of the Hall conductance observed in MOSFETS and in heterostructures has demanded a great deal of theoretical sophistication. Concepts drawn from branches of mathematics, such as topology and differential geometry, have become essential to the understanding of this phenomenon. In this chapter I will only consider the Quantum Hall Effect in non-interacting systems. This is the theory of the *integer* Hall Effect. The Fractional Quantum Hall Effect is discussed in chapter 10.

The chapter begins with a description of the one-electron states, both in the continuum and on a two-dimensional lattice, followed by a summary of the observed phenomenology of the Quantum Hall Effect. A brief discussion of Linear Response theory is also presented. The rest of the chapter is devoted to the problem of topological quantization of the Hall conductance.

9.1 Particle States in the Presence of a Magnetic Field

Let us review the Landau problem of the states of charged particles moving on a plane in the presence of a perpendicular uniform magnetic field B. We will consider both, the continuum and lattice versions of the problem.

We consider first the continuum problem. Let us think of a spinless particle of mass M and charge e. The one-particle Hamiltonian which describes the dynamics of this system is

$$\mathcal{H} = \frac{1}{2M}[(-i\hbar\frac{\partial}{\partial x_1} + \frac{e}{c}A_1)^2 + [(-i\hbar\frac{\partial}{\partial x_2} + \frac{e}{c}A_2)^2)]. \qquad (9.1.1)$$

The vector potential $\vec{A}$ is such that its curl is equal to B, the perpendicular

component of the field,

$$B = \epsilon_{ij}\partial_i A_j. \tag{9.1.2}$$

If the linear size of the plane is L, the total flux Φ is

$$\Phi = BL^2. \tag{9.1.3}$$

In what follows, I will assume that there is an exact *integer* number N_ϕ of flux quanta ϕ_0 piercing the plane.

$$\Phi = N_\phi \phi_0 \equiv N_\phi \frac{hc}{e}. \tag{9.1.4}$$

If we choose units such that $\hbar = e = c = 1$, the flux quantum ϕ_0 is just equal to 2π. In these units we can write $\Phi = 2\pi N_\phi$. Also, we are going to measure lengths in units of the magnetic length l_0 defined to be $l_0 = B^{-\frac{1}{2}}$.

We will work in the isotropic gauge

$$A_i = -\frac{1}{2} B \epsilon_{ij} x_j. \tag{9.1.5}$$

In this gauge, it is convenient to work in complex coordinates $z = x_1 + ix_2$. Let us factor an exponentially decaying function of $|z|^2$ out of the wave function. This procedure automatically introduces an apparently special point, the origin $z = 0$. Since the location of the origin must be arbitrary, there should exist an operator which will remove this arbitrariness. We will see that this is the case. As a by product, we will also find that the energy eigenvalues, the Landau levels, not only are degenerate but that this degeneracy is generated by a special group of transformations, the group of magnetic translations [Zak 64]. If we are dealing with rotationally invariant system, such as a disc, it is convenient to write the wave functions in the form

$$\Psi(z, \bar{z}) = f(z, \bar{z})\, e^{-\lambda |z|^2} \tag{9.1.6}$$

which decays exponentially fast at infinity. For this Hilbert space, the disk is topologically equivalent to a two-sphere.

If we now choose for λ the value

$$\lambda = \frac{e|B|}{4\hbar c} = (\frac{e}{\hbar c})\frac{|B|}{4} \equiv \frac{1}{4l_0^2} \tag{9.1.7}$$

(where we introduced the cyclotron radius l_0), the function $f(z, \bar{z})$ is found to satisfy an equation which, in complex coordinates, has the form

$$-\frac{2\hbar^2}{M}\partial_z \partial_{\bar{z}} f + \frac{e|B|\hbar}{Mc} z \partial_z f + \frac{e|B|\hbar}{2Mc} f = Ef \tag{9.1.8}$$

for $B > 0$. For $B < 0$ we must replace z by $\bar{z}$. In Eq.(9.8), we have introduced the operators ∂_z and $\partial_{\bar{z}}$ defined by

$$\partial_z = \frac{1}{2}(\partial_1 - i\partial_2) \tag{9.1.9}$$

$$\partial_{\bar{z}} = \frac{1}{2}(\partial_1 + i\partial_2). \tag{9.1.10}$$

It will be sufficient to discuss the case of $B > 0$.

Any *analytic* function $f(z)$ is a solution of Eq.(9.8). A complete basis $\{f_n(z)\}$ has the form

$$f_n = z^n \tag{9.1.11}$$

which are also eigenstates of angular momentum operator L_z

$$L_z = -i\hbar(x_1\partial_2 - x_2\partial_1) \equiv +\hbar(z\partial_z - \bar{z}\partial_{\bar{z}}) \tag{9.1.12}$$

with eigenvalue

$$L_z z^n = \hbar n z^n. \tag{9.1.13}$$

The energy E_n of the n^{th} Landau level is

$$\begin{aligned} E_n &= \hbar\omega_c(n + \frac{1}{2}), \\ \omega_c &= \frac{eB\hbar}{Mc} \end{aligned} \tag{9.1.14}$$

is the cyclotron frequancy. The Landau levels have a huge degeneracy which the same for all the Landau levels and it is equal to N_ϕ.

In order to make this degeneracy more apparent, let us introduce the operators of *magnetic translation* and the group of transformations induced by them. Let $\vec{a}$ and $\vec{b}$ be two vectors on the plane. For a system in a magnetic field B ($B > 0$), the canonical momentum operator $\vec{P}$ is given by the usual minimal coupling definition

$$\begin{aligned} \vec{P} &= -i\hbar\vec{\nabla} + \frac{e}{c}\vec{A}, \\ B &= \epsilon_{ij}\partial_i A_j \end{aligned} \tag{9.1.15}$$

The generator of infinitesimal magnetic translations $\vec{k}$ [Zak 64] is

$$k_i = P_i - \frac{eB}{c}\epsilon_{ij}x_j \equiv P_i(-B). \tag{9.1.16}$$

A *finite* magnetic translation by a vector $\vec{a}$ is represented by the operator $\hat{t}(\vec{a})$

$$\hat{t}(\vec{a}) = e^{i\vec{a}\cdot\frac{\vec{k}}{\hbar}}. \tag{9.1.17}$$

These *magnetic translation operators* obey the so called *magnetic algebra*

$$\hat{t}(\vec{a})\hat{t}(\vec{b}) = e^{-i\frac{(\vec{a}\times\vec{b})\cdot\hat{z}}{l_0^2}}\hat{t}(\vec{b})\hat{t}(\vec{a}) \tag{9.1.18}$$

where $\hat{z}$ is a unit vector normal to the plane.

The magnetic translations form a group in the sense that the operators $\hat{t}(\vec{a})$ obey the composition law

$$\hat{t}(\vec{a})\hat{t}(\vec{b}) = \exp(\frac{i}{2l_0^2}(\vec{a}\times\vec{b})\cdot\hat{z})\ \hat{t}(\vec{b}+\vec{a}). \tag{9.1.19}$$

Thus, the operators $\hat{t}(\vec{a})$ form a representation of the group of magnetic translations. Eq.(9.19) has an extra phase factor, not present in the usual group

composition law. The existance of this phase, which is known in mathematics as a *cocycle*, indicates that the operators $\hat{t}(\vec{a})$ form a *ray representation* of the group of magnetic translations.

The Hamiltonian for a charged particle moving in a magnetic field can now be written in the standard form, $H = \frac{\vec{P}^2}{2M}$. The canonical momentum operators $\vec{P}$ and the generators of magnetic translations $\vec{k}$ commute with each other

$$[k_i, P_j] = 0 \tag{9.1.20}$$

although the *different* components of $\vec{k}$ (and $\vec{P}$) do not commute among themselves

$$[k_i, k_j] = -[P_i, P_j] = \frac{ie\hbar B}{c}\epsilon_{ij}. \tag{9.1.21}$$

Thus, the two components of $\vec{k}$ commute with the Hamiltonian

$$[k_i, H] = \frac{1}{2M}[k_i, \vec{P}^2] = 0 \tag{9.1.22}$$

and are constants of motion. However, since k_1 and k_2 do not commute with each other, they cannot be diagonalized simultaneously. We can then use k_1 or k_2, or some linear combination of both, to label the degenerate states. Which combination is convenient depends on the choice of boundary conditions.

Let us assume, for the moment, that the system has the shape of a rectangle with linear dimensions L_1 and L_2 along the (orthogonal) directions $\hat{e}_1$ and $\hat{e}_2$ respectively ($\hat{e}_i \cdot \hat{e}_j = \delta_{ij}, i = 1, 2$). The total flux Φ passing through the rectangle is $\Phi = BL_1L_2$. In units of the flux quantum $\Phi_0 = \frac{hc}{e}$, the total flux is an integer $N_\phi = \Phi/\Phi_0$. Alternatively, N_ϕ can be given in terms of the magnetic length l_0 and the area of the system L_1L_2 in the equivalent form

$$\frac{L_1L_2}{l_0^2} = 2\pi N_\phi. \tag{9.1.23}$$

Let us now consider the operator $\hat{T}_1$ and $\hat{T}_2$ which represent magnetic translations by $\frac{L_1}{N_\phi}$ and $\frac{L_2}{N_\phi}$ along the directions $\hat{e}_1$ and $\hat{e}_2$ respectively

$$\begin{aligned} \hat{T}_1 &\equiv \hat{t}\,(\frac{L_1}{N_\phi}\hat{e}_1), \\ \hat{T}_2 &\equiv \hat{t}\,(\frac{L_2}{N_\phi}\hat{e}_2) \end{aligned} \tag{9.1.24}$$

These operators obey the algebra

$$\hat{T}_1\hat{T}_2 = e^{-i\frac{2\pi}{N_\phi}}\,\hat{T}_2\hat{T}_1 \tag{9.1.25}$$

which is often also referred to as the *magnetic algebra*. In chapter 7 we discussed this algebra in the context of the commutation relations for anyon operators.

Let us now assume that we have a state $\Psi_{n,0}$ which is an eigenstate of the Hamiltonian in the n-th Landau level and that it is also an eigenstate of,

say, $\hat{T}_1$, i.e.

$$\begin{aligned} \hat{H}\Psi_{n,0} &= E_n\Psi_{n,0}, \\ \hat{T}_1\Psi_{n,0} &= e^{i\lambda_0}\Psi_{n,0}, \end{aligned} \tag{9.1.26}$$

where E_n and λ_0 are the eigenvalues. Consider now the state $\Psi_{n,m}$

$$\Psi_{n,m} = \hat{T}_2^m\ \Psi_{n,0}. \tag{9.1.27}$$

Since both $\hat{T}_1$ and $\hat{T}_2$ commute with $\hat{H}$, it follows that all the states $\Psi_{n,m}$ have energy E_n.

$$\hat{H}\Psi_{n,m} = \hat{H}\hat{T}_2^m\Psi_{n,0} = \hat{T}_2^m\hat{H}\Psi_{n,0} = E_n\Psi_{n,m}. \tag{9.1.28}$$

However, the states $\Psi_{n,m}$ have different eigenvalues of $\hat{T}_1$

$$\hat{T}_1\Psi_{n,m} = e^{-i2\pi\frac{m}{N_\phi}+i\lambda_0}\ \Psi_{n,m}. \tag{9.1.29}$$

Thus, there are exactly N_ϕ linearly independent degenerate eigenstates of the Hamiltonian in a given Landau level. For a system with wave functions satisfying vanishing at spacial infinity (i.e. a "disk") the operators k_1 and k_2 are replaced by their counterparts in complex coordinates, k and $\bar{k}$.

$$\begin{aligned} k &= \frac{i}{2\hbar}(k_1 - ik_2) = \partial_z - \frac{\bar{z}}{4l_0^2} \\ \bar{k} &= \frac{i}{2\hbar}(k_1 + ik_2) = \partial_{\bar{z}} + \frac{z}{4l_0^2} \end{aligned} \tag{9.1.30}$$

which also commute with the momenta (in complex coordinates) P and $\bar{P}$

$$\begin{aligned} P &= \frac{i}{2\hbar}(P_1 + iP_2) = \partial_{\bar{z}} - \frac{z}{4l_0^2} \\ \bar{P} &= \frac{i}{2\hbar}(P_1 - iP_2) = \partial_z + \frac{\bar{z}}{4l_0^2} \end{aligned}. \tag{9.1.31}$$

The complex coordinate analogs of $\hat{T}_1$ and $\hat{T}_2$, T and $\bar{T}$ are defined by

$$\begin{aligned} T &= e^{\frac{2L}{N_\phi}k}, \\ \bar{T} &= e^{i\frac{L}{N_\phi}\bar{k}}, \end{aligned} \tag{9.1.32}$$

for a system with $L_1 = L_2 = L$. The operators $T, \bar{T}$ also satisfy the magnetic algebra Eq.(9.1.25). The operator $\bar{k}$ anhilates the wave function $\Psi_n(z, \bar{z})$

$$\bar{k}\Psi_n = 0 \qquad \text{with} \qquad \Psi_n = c_n z^n e^{-\frac{|z|^2}{4l_0^2}}. \tag{9.1.33}$$

Thus, Ψ_n is an eigenstate of $\bar{T}$ with unit eigenvalue

$$\bar{T}\Psi_n = e^{i\frac{L}{N_\phi}\bar{k}}\ \Psi_n = \Psi_n. \tag{9.1.34}$$

A complete set of eigenstates of the n-th Landau level $\{\Psi_{n,m}\}$ can now be constructed quite easily ($m = 1, ..., N_\phi$)

$$\Psi_{n,m}(z, \bar{z}) = T^m\Psi_n(z, \bar{z}) \equiv C_{n,m}\ e^{+2L\frac{m}{N_\phi}k}\Psi_n(z, \bar{z}). \tag{9.1.35}$$

The states in the set $\{\Psi_{n,m}(z,\bar{z})\}$ have eigenvalues

$$\begin{aligned} H\Psi_{n,m}(z,\bar{z}) &= E_n \Psi_{n,m}(z,\bar{z}) \\ \bar{T}\Psi_{n,m}(z,\bar{z}) &= e^{-i2\pi\frac{m}{N_\phi}} \Psi_{n,m}(z,\bar{z}) \end{aligned} \tag{9.1.36}$$

with

$$H = \frac{2\hbar^2}{M}[-P\bar{P} + \frac{eB}{4\hbar c}]. \tag{9.1.37}$$

If instead of open (or vanishing) boundary conditions we want to consider a system on a *torus*, i.e. periodic boundary conditions along the directions $\hat{e}_1$ and $\hat{e}_2$ of a rectangle, the wave functions will have to satisfy a periodicity condition. It is customary to demand that

$$\Psi(x_1,x_2) = \Psi(x_1+L_1,x_2) = \Psi(x_1,x_2+L_2). \tag{9.1.38}$$

However, it is not possible to satisfy this condition if a non-zero magnetic field is present. The vector potential violates translation invariance. Thus, the wave functions cannot obey periodic boundary conditions (PBC's) since no flux could possibly go through the system if PBC's are to be obeyed. In such a case, the circulation of $\vec{A}$ around the boundary equals zero. In order to accomodate a non-zero external flux, the vector potentials *and* the wave functions, have to change by a large gauge transformation as we traversethe system [Haldane 85a]

$$\begin{aligned} A_i(x_1+L_1,x_2) &= A_i(x_1,x_2) + \partial_i\beta_1(x_1,x_2) \\ A_i(x_1,x_2+L_2) &= A_i(x_1,x_2) + \partial_i\beta_2(x_1,x_2) \end{aligned} \tag{9.1.39}$$

such that the circulation around the boundary Γ equals the flux Φ. This requirement implies that β_1 and β_2 must satisfy the condition

$$[\beta_2(x_1+L_1,x_2) - \beta_2(x_1,x_2)] - [\beta_1(x_1,x_2+L_2) - \beta_2(x_1,x_2)] = \Phi. \tag{9.1.40}$$

It is sufficient to give just one solution to this equation, which we choose to be

$$\beta_i = -\frac{1}{2}\Phi\ \epsilon_{ij}\frac{x_j}{L_j}. \tag{9.1.41}$$

The requirement of gauge invariance forces the wave functions $\Psi(x_1,x_2)$ to transform like

$$\Psi(x_1,x_2) \rightarrow \exp(-i\frac{e}{\hbar c}\Lambda(x_1,x_2))\ \Psi(x_1,x_2) \tag{9.1.42}$$

under a gauge transformation in which A_i changes by $\partial_i\Lambda(x_1,x_2)$. Thus, under the large gauge transformation of Eq.(9.1.39), the wave functions must change like

$$\begin{aligned} \Psi(x_1+L_1,x_2) &= e^{i\frac{e}{\hbar c}\beta_1(x_1,x_2)}\ \Psi(x_1,x_2) \\ \Psi(x_1,x_2+L_2) &= e^{i\frac{e}{\hbar c}\beta_2(x_1,x_2)}\ \Psi(x_1,x_2) \end{aligned} \tag{9.1.43}$$

The boundary conditions of Eq.(9.1.39) and Eq.(9.1.43) are consistent provided that the translations $(x_1,x_2) \rightarrow (x_1+L_1,x_2) \rightarrow (x_1+L_1,x_2+L_2)$ and $(x_1,x_2) \rightarrow (x_1,x_2+L_2) \rightarrow (x_1+L_1,x_2+L_2)$ lead to the same value of the

wave function. It is easy to check that this consistency condition leads to flux quantization $\Phi = N_{\Phi_0}$. This result should come to no surprise since we are in the situation of the Aharonov-Bohm effect. In other words, the system has single valued wave functions on the torus only if the flux is quantized. The (single valued) wave functions constructed with this prescription are (doubly) periodic and form N_ϕ-fold degenerate multiplets. If the flux is not quantized (e.g. a rational multiple of ϕ_0) the wave functions are multivalued and have branch cuts.

9.2 The Hofstadter Wave Functions

In the past section we considered the quantum mechanical motion of charged particles which move in a plane in the presence of an external magnetic field perpendicular to the plane. There are many physical situations in which the presence of a *lattice* cannot be ignored. In most cases these effects are quite small. Magnetic fields are relativistic effects and if we want to pass a sizable fraction of the flux quantum ϕ through a plaquette of a physical lattice (with spacing $a_0 \approx 10\mathring{A}$), astronomically large magnetic fields are required. Thus, for problems such as electrons in a heterostructure, lattice effects are, in practice, negligible. However, when we are dealing with a Chiral Spin State, we discover the existance of dynamically generated gauge fields with *large* fluxes. Here, of course, lattice effects become dominant.

The problem of the quantum motion on two dimensional lattices in external magnetic fields was first studied by D. Hofstadter [Hofstadter 76]. He considered the problem of particle of charge e hopping on a square lattice, with hopping amplitude t, in the presence of an external uniform magnetic field B. Let $|\vec{x}\rangle$ denote the (Wannier) state localized at site $\vec{x}$ of the square lattice. The hopping (tight-binding) Hamiltonian H is

$$H = -t \sum_{\vec{x},j=1,2} |\vec{x}\rangle \, e^{i\frac{e}{\hbar c}A_j(\vec{x})} \, \langle \vec{x} + e_j| + h.c. \tag{9.2.1}$$

The vector potentials $A_j(\vec{x})$ reside on the links and represent the external flux. The total flux Φ going through any individual plaquette (with lattice spacing $a_0 = 1$) is B

$$\sum A_j = \Delta_1 A_2 - \Delta_2 A_1 = B. \tag{9.2.2}$$

If we demand that the system be a torus, it is customary to work in the *Landau gauge*

$$A_1 = -Bx_2 \qquad A_2 = 0 \tag{9.2.3}$$

where x_1 and x_2 are integers ($0 \leq x_i \leq L_i, i = 1, 2$). From now on, I will assume that $Ba^2 = \frac{p}{q}\phi_0$ with p and q a pair of relatively prime integers. In other words, the flux going through an elementary plaquette is a fininte fraction $(\frac{p}{q})$ of the flux quantum ϕ_0.

The eigenstates $|\Psi\rangle$ of the system can be expanded in terms of a set of site (or Wannier) states

$$|\Psi\rangle = \sum_{\vec{x}} \Psi(\vec{x})|\vec{x}\rangle \tag{9.2.4}$$

and obey the discrete Schrödinger equation

$$\begin{aligned} &-t\left\{e^{-i2\pi\frac{p}{q}x_2}\,\Psi(x_1+1,x_2)+e^{+i2\pi\frac{p}{q}x_2}\,\Psi(x_1-1,x_2)\right\}+ \\ &-t\left\{\Psi(x_1,x_2+1)+\Psi(x_1,x_2-1)\right\}=E\,\Psi(x_1,x_2) \end{aligned} \tag{9.2.5}$$

This Hamiltonian is not invariant under translations by one lattice spacing. However, in the Landau gauge, it is invariant under the translations

$$\begin{aligned} (x_1,x_2) &\rightarrow (x_1+q,x_2) \\ (x_1,x_2) &\rightarrow (x_1,x_2+1)\cdot \end{aligned} \tag{9.2.6}$$

Hence, the unit cell has q elementary plaquettes. With the present choice of gauge, the unit cell is $1 \times q$. The total flux passing through the unit cell is

$$\Phi_{\text{cell}} = q\,\Phi_{\text{plaquette}} = p \tag{9.2.7}$$

which is an integer. Naturally, this is not an accident.

The gauge-invariant operator for translations $e^{i\hat{P}_j}$ is (in units such that $e=\hbar=c=a_0=1$)

$$e^{i\hat{P}_j} = \sum_{\vec{x}} |\vec{x}\rangle\; e^{iA_j(\vec{x})}\;\langle\vec{x}+\hat{e}_j|\cdot \tag{9.2.8}$$

These operators satisfy the algebra

$$e^{i\hat{P}_1}e^{i\hat{P}_2} = e^{i2\pi\frac{p}{q}}e^{i\hat{P}_2}e^{i\hat{P}_1} \tag{9.2.9}$$

and, hence, do not generally commute with each other. But $e^{in_1\hat{P}_1}$ and $e^{in_2\hat{P}_2}$ commute if

$$\frac{p}{q}n_1n_2 = \text{integer}\cdot \tag{9.2.10}$$

Thus, the translations $e^{in_1\hat{P}_1}$ and $e^{in_2\hat{P}_2}$ commute if and only if the flux passing through the rectangle with edges n_1 and n_2 is an integer multiple of the flux quantum. The smallest rectangle satisfying Eq.(9.54) is known as the *magnetic unit cell.*

The hopping Hamiltonian can now be written in terms of the operators $e^{i\hat{P}_j}$ in the form

$$H = -t\sum_{j=1,2}(e^{i\hat{P}_j}+e^{-i\hat{P}_j})\cdot \tag{9.2.11}$$

The eigenstates of H are not eigenstates of e^{iP_j}, but of the operators $e^{i\hat{k}_j}$ which generate finite (i.e. lattice) *magnetic translations.* The operators $e^{i\hat{k}_j}$ are defined by

$$e^{i\hat{k}_j} \equiv \sum_{\vec{x}} |\vec{x}\rangle e^{iA'_j(\vec{x})}\langle\vec{x}+\hat{e}_j| \tag{9.2.12}$$

where the vector potentials $A'_j(\vec{x})$ have to be chosen so that the *magnetic translation* operators $e^{i\hat{k}_j}$ *commute* with the elementary *lattice translations* $e^{i\hat{P}_j}$ and, hence, with the Hamiltonian H. This conditions are met if we choose $(j \neq k)$

$$\Delta_j A'_k(\vec{x}) = \Delta_k A_j(\vec{x}). \tag{9.2.13}$$

So, once again, we find

$$\hat{k}_j = \hat{P}_j(-B) \tag{9.2.14}$$

but in the specific choice of gauge:

$$A'_1(\vec{x}) = 0 \qquad A'_2(\vec{x}) = -2\pi\frac{p}{q}x_1. \tag{9.2.15}$$

The operators $e^{i\hat{k}_j}$ do not commute with each other. Rather, they obey

$$e^{i\hat{k}_1}e^{i\hat{k}_2} = e^{i2\pi\frac{p}{q}}\, e^{i\hat{k}_2}\, e^{i\hat{k}_1}. \tag{9.2.16}$$

Consider now the magnetic translations by n_1 steps along x_1 and n_2 steps along x_2 (no sum over j)

$$\hat{T}_j^{n_j} = e^{in_j\hat{k}_j}. \tag{9.2.17}$$

These operators commute with each other if n_1 and n_2 satisfy Eq.(9.2.10). Thus, the eigenstates of H are also eigenstates of $\hat{T}_1^{n_1}$ and $\hat{T}_2^{n_2}$. With the choice of Eq.(9.2.15), we see that $\hat{T}_1$ and $\hat{T}_2^q$ satisfy

$$[\hat{T}_1, \hat{T}_2^q] = [\hat{T}_1, \hat{H}] = [\hat{T}_2^q, H] = 0 \tag{9.2.18}$$

and their eigenstates can be used to label the eigenstates of H. The eigenstates of $\hat{T}_1$ and $\hat{T}_2^q$ are of the form $|k_1, k_2\rangle$

$$\begin{aligned} \hat{T}_1 \mid k_1, k_2\rangle &= e^{ik_1}|k_1, k_2\rangle \\ \hat{T}_2^q|k_1, k_2\rangle &= e^{iqk_2}|k_1, k_2\rangle \end{aligned} \tag{9.2.19}$$

and must satisfy periodic boundary conditions

$$\begin{aligned} \hat{T}_1^{L_1} \mid k_1, k_2\rangle &= |k_1, k_2\rangle \\ (\hat{T}_2^q)^{\frac{L_2}{q}}|k_1, k_2\rangle &= |k_1, k_2\rangle. \end{aligned} \tag{9.2.20}$$

These conditions can only be met if (k_1, k_2) belongs to the *magnetic Brillouin zone* $(-\pi \leq k_1 < \pi)$ and $(-\frac{\pi}{q} \leq k_2 < \frac{\pi}{q})$. Clearly, this boundary conditions can only be imposed if L_2 is an integer multiple of q. That is to say, the total flux Φ going through the entire system has to be an integer N_ϕ multiple of the flux quantum ϕ_0, with $N_\phi = \frac{p}{q}L_1L_2$.

The magnetic Brillouin zone labels a total of $\frac{L_1L_2}{q}$ states. We will see now that this system has q Landau (or Hofstadter) bands each with a $\frac{L_1L_2}{q}$-fold degeneracy. This is the discrete version of the degeneracy of the continuum problem.

Let us now expand the states $\Psi(\vec{x})$ in terms of magnetic translation eigenstates

$$\Psi(x_1,x_2)=\frac{1}{q}\sum_{r=1}^{q}\int_{-\pi}^{\pi}\frac{dk_1}{2\pi}\int_{\frac{-\pi}{q}}^{\frac{\pi}{q}}\frac{dk_2}{2\pi/q}\,e^{i(k_1x_1+k_2x_2)}\,\Psi(k_1,k_2+2\pi\frac{p}{q}r)\cdot \quad (9.2.21)$$

It is now convenient to define the q-component vector, $\Psi_r(k_1,k_2)$

$$\Psi_r(k_1,k_2)\equiv\Psi(k_1,k_2+2\pi\frac{p}{q}r) \qquad r=1,...,q. \quad (9.2.22)$$

We recognize in this vector a generalization of the spinons used to study the flux phase where $\frac{p}{q}=\frac{1}{2}$. The (discrete) Schroedinger equation now reads

$$-t[e^{ik_1}\Psi_{r+1}(k_1,k_2)+e^{-ik_1}\Psi_{r-1}(k_1,k_2)]-2t\cos(k_2+2\pi\frac{p}{q}r)\Psi_r(k_1,k_2)=$$
$$=E(k_1,k_2)\Psi_r(k_1,k_2)\cdot \quad (9.2.23)$$

This equation is also known under the name of Harper's equation and plays an important role in the theory of incommensurate systems. The amplitudes $\Psi_r(r_1,r_2)$ are periodic functions on the magnetic Brillouin zone and thus satisfy

$$\begin{aligned}\Psi_r(k_1+2\pi n_1,k_2)&=\;=\Psi_r(k_1,k_2)\\ \Psi_r(k_1,k_2+\frac{2\pi}{q}n_2)&=\Psi_{r+n_2}(k_1,k_2)\end{aligned} \quad (9.2.24)$$

and

$$\Psi_{r+q}(k_1,k_2)=\Psi_r(k_1,k_2) \quad (9.2.25)$$

where n_1 and n_2 are integers. Eq.(9.2.24) implies that the magnetic Brillouin zone has the topology of a two-torus. The amplitudes $\Psi_r(k_1,k_2)$, which are solutions of Eq.(9.2.23), form an r-component complex vector field which is continuous on the torus.

For arbitrary values of the integers p and q (p and q relatively prime), the spectrum determined from Eq.(9.2.23) has a very complex structure. For instance, if p and q are chosen to belong to some infinite sequence such that, in the limit, $\frac{p}{q}$ becomes arbitrarily close to an irrational number, the spectrum becomes a Cantor set [Hofstadter 76] and the wave functions exhibit self-similar behavior [Kohmoto 83]. Even if the problem is restricted to commensurate flux only ($\Phi=2\pi\frac{p}{q}\phi_0$), the spectrum has energy gaps which , as q is increased, exhibits a hierarchical structure. We will not consider these issues here. Rather, we will consider only the broad qualitative properties of the spectrum and wave functions. In general, Eq.(9.2.23) has to be solved numerically.

For generic values of p and q, the spectrum has q bands. For any arbitrary pair of relatively prime integers p and q, the Hamiltonian $\mathcal{H}(k_1,k_2)$ associated with the Schroedinger equation Eq.(9.2.23) has a number of symmetries [Wen

89]. Let $\hat{A}$ and $\hat{B}$ be a pair of $q \times q$ matrices defined by

$$\begin{aligned} \hat{A}_{jk} &= \omega^k \delta_{jk} \\ \hat{B}_{jk} &= \delta_{j,k-1} \end{aligned} \tag{9.2.26}$$

where $j, k = 1, \ldots, q$ and $\omega = e^{-i2\pi p/q}$, and satisfy the algebra $AB = e^{i2\pi\frac{p}{q}}BA$. The Hamiltonian $\mathcal{H}(k_1, k_2)$ is given by

$$\mathcal{H}(k_1, k_2) = e^{-ik_2}\hat{A} + e^{+ik_1}\hat{B} + h.c. \tag{9.2.27}$$

Given p and q, we can always find a pair of (relativley prime) integers n and m such that

$$1 = np + mq. \tag{9.2.28}$$

It is easy to check that the matrices $\tilde{A} \equiv \hat{A}^n$ and $\tilde{B} \equiv \hat{B}^m$ satisfy the following identities

$$\tilde{A}\mathcal{H}(k_1, k_2)\tilde{A}^{-1} = \mathcal{H}(k_1 + \frac{2\pi}{q}n, k_2) \tag{9.2.29}$$

$$\tilde{B}\mathcal{H}(k_1, k_2)\tilde{B}^{-1} = \mathcal{H}(k_1, k_2 + \frac{2\pi}{q}) \tag{9.2.30}$$

$$\mathcal{H}(k_1 + \frac{2\pi}{q} - 2\pi m, k_2) = \mathcal{H}(k_1 + \frac{2\pi}{q}, k_2) \tag{9.2.31}$$

$$\tilde{A}\tilde{B} = e^{-i2\pi p\frac{n^2}{q}}\tilde{B}\tilde{A}. \tag{9.2.32}$$

Thus, if $\Psi(k_1, k_2)$ is an eigenstate of $\mathcal{H}(k_1, k_2)$ with energy $E(k_1, k_2)$, the state $\Psi'(k_1, k_2)$

$$\Psi'(k_1, k_2) = \tilde{A}\Psi(k_1, k_2) \tag{9.2.33}$$

is an eigenstate of $\mathcal{H}'(k_1, k_2) = \mathcal{H}(k_1 + \frac{2\pi}{q}, k_2)$ with the *same* eigenvalue $E(k_1, k_2)$. In other words, there is a one-to-one correspondence between the spectrum at (k_1, k_2) and at $(k_1 + \frac{2\pi}{q}, k_2)$. An analogous argument shows that the spectra at $(k_1, k_2 + \frac{2\pi}{q})$ and (k_1, k_2) are also indentical to each other. In addition, under the translation $(k_1, k_2) \rightarrow (k_1 + \pi, k_2 + \pi)$, the Hamiltonian changes sign, i.e. $\mathcal{H}(k_1 + \pi, k_2 + \pi) = -\mathcal{H}(k_1, k_2)$, and $E(k_1 + \pi, k_2 + \pi) = -E(k_1, k_2)$. For q *even* this operation is a particular case of Eq.(9.2.29). Thus, if q is even, for each eigenstate of $\mathcal{H}$ with energy E, there exists an eigenstate with energy $-E$. The operator which connects states with opposite signs of the enery, let us call it Γ, must anticommute with $\mathcal{H}$ and be hermitian. It is easy to check that the matrix Γ_{jk} with elements

$$\Gamma_{jk} = (-1)^j i^{\frac{q}{2}} \delta_{k,j+\frac{q}{2}} \tag{9.2.34}$$

has the desired properties

$$\{\mathcal{H}, \Gamma\} = 0 \qquad \Gamma^2 = I \tag{9.2.35}$$

Γ also anticommutes with $\hat{A}, \hat{B}, \tilde{A}$ and $\tilde{B}$.

Furthermore, it is possible to show that for q even there are at least q eigenstates of $\mathcal{H}$ with zero energy [Wen 89]. The argument uses the topology of the torus in an essential way. It can be regarded as a generalization of the

Nielsen-Ninomiya theorem for the absence of Weyl fermions in lattice systems [Nielsen 81]. The magnetic Brillouin zone is locally isomorphic to the complex plane ($w = k_1 + ik_2$) and globally equivalent to a torus. Let us consider a point w on the magnetic Brillouin zone and assume that the eigenstates of $\mathcal{H}$ at $w = k_1 + ik_2$ are all different from zero. It is possible to choose a basis of states in which Γ is diagonal. In this basis we can write

$$\Gamma = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad \mathcal{H} = \begin{pmatrix} 0 & h^+ \\ h & 0 \end{pmatrix} \tag{9.2.36}$$

h is a $\frac{q}{2} \times \frac{q}{2}$ non hermitean matrix and h^+ is its adjoint. If in the neighborhood of w there are no zero energy eigenvalues, the determinant of $\mathcal{H}$

$$\det\mathcal{H} = -|\det h| \tag{9.2.37}$$

is non zero. Let us denote by D the determinant of h, $D = \det h$. D is locally an analytic function of w. Thus, it is possible to define the vector field $\mathcal{A}_i$ $(i = 1, 2)$

$$\mathcal{A}_i = D^{-1}\frac{\partial}{\partial k_i}D \tag{9.2.38}$$

which, in fiber-bundle parlance, is a one-form. In any neighborhood of w free of zero energy eigenvalues, the one-form $\mathcal{A}_i$ is closed, i.e.

$$\epsilon_{ij}\partial_i\mathcal{A}_j = \epsilon_{ij}\partial_i\partial_j \ln D = 0 \tag{9.2.39}$$

but, in general, it is not exact. The circulation ν of $\vec{\mathcal{A}}$ on an arbitrary contour $\mathcal{C}$ of the magnetic Brillouin zone

$$\nu = \frac{1}{2\pi}\oint_{\mathcal{C}} d\vec{k}\cdot\mathcal{A} \tag{9.2.40}$$

is in general different from zero. If ν is not zero, the determinant $D(\vec{k}) = \det h(\vec{k})$ must have a zero at some point $\vec{k}_0$ somewhere inside $\mathcal{C}$. We now follow Wen and Zee and consider a path $\mathcal{C}$ which is a rectangle with corners at $(k_1, k_2), (k_1 + \frac{2\pi}{q}, k_2), (k_1 + \frac{2\pi}{q}, k_2 + \frac{2\pi}{q})$ and $(k_1, k_2 + \frac{2\pi}{q})$. From the symmetries of $\mathcal{H}$, it is possible to show that $D(k_1, k_2)$ satisfies

$$D(k_1, k_2) = -D^*(k_1 + \frac{2\pi}{q}, k_2) = -D(k_1 + \frac{2\pi}{q}, k_2 + \frac{2\pi}{q}) = D^*(k_1, k_2 + \frac{2\pi}{q}). \tag{9.2.41}$$

Eq.(9.2.41) implies that the *phase* of D must wind as the path $\mathcal{C}$ is traversed. In general, $D(\vec{k})$ being a complex number, will trace a closed path $\mathcal{D}$ on the complex plane as $\vec{k}$ traces the path $\mathcal{C}$. If the D does not have a zero inside $\mathcal{C}$, the winding number ν will vanish and $\mathcal{C}$, and hence $\mathcal{D}$, can be smoothly shrunk to zero. If there is a zero, D will have a singularity and $\mathcal{C}$ cannot be deformed to zero. The path $\mathcal{D}$ will now wind around the origin $\mathcal{D} = 0$ a number of times before closing. The winding number ν of Eq.(9.81) is precisely this winding number. Since $D(\vec{k})$ is not a constant, we conclude that it must have zeroes at certain isolated locations. However, the translation symmetries of Eq.(9.2.29) require that if $\vec{k}_0$ is a zero of $\mathcal{H}$, then $\vec{k}_0 + \frac{2\pi}{q}(n_1\hat{e}_1 + n_2\hat{e}_2)$ must also be

zeroes of $\mathcal{H}$. This lattice of zeroes of $\mathcal{H}$ must be periodic. The only values of $\vec{k}_0$ consisent with these demands are $\vec{k}_0 = (\frac{\pi}{2}, \frac{\pi}{2})$ and its translations.

There are exactly q distinct points in this lattice. Thus, for q even, the Hamiltonian has exactly q zeros. Note that the flux phase is a particular case of this problem. The doubling of spinon species that we found there, is a particular case of the q-fold multiplicity discussed in this section.

9.3 The Quantum Hall Effect

In this section we will discuss the most qualitative features of a very fascinating problem: the Quantum Hall Effect. It is not within the scope of this book to give an exhaustive review on this subject. Reviews are already available, in particular the excellent volume by R. Prange and S. Girvin [Prange 90].

However, there are very close analogies and connections between the theories of the Fractional and Integer Quantum Hall Effects and the theories of Chiral Spin Liquids. We will devote considerable attention to these analogies.

In 1980, K. von Klitzing, G. Dorda and M. Pepper [Klitzing 80] announced the discovery of very unusual transport properties of a two-dimensional electron gas in a high magnetic field. They were studying the Hall conductance of two-dimensional inversion layers or MOSFETS. In these systems, the electrons of a semiconductor move on quantum states which are localized whithin atomic scales of the layer. They are almost free to move inside the layer. Von Klitzing and his collaborators noticed that when they measured the Hall conductance σ_{xy} of the layer at very low temperatures, the conductance had a stepwise dependence on the external magnetic field. At the same time, the longitudinal conductivity σ_{xx}, appeared to be essentially zero when σ_{xy} was nearly constant, the so called plateaus. For values of the field in which σ_{xy} varied, σ_{xx} was non zero.

What was very unusual was the values that σ_{xy} attained at the plateaus. It appeared to be *quantized* at integer multiples of $\frac{e^2}{h}$. Furthermore, the quantization appeared to be sharper at lower temperatures and, oddly enough, for the more disordered samples. This phenomenon is known today as the Integer Quantum Hall Effect.

In 1982, D. Tsui, H. Stormer and A. Gossard [Tsui 82], performed a similar series of experiments but on highly pure $Ga\,As - Al\,As$ heterojunctions. Here too, the electrons are bound to a surface and are essentially free to move inside the surface. They found a Fractional Quantum Hall Effect. In fact, their results were very similar to what von Klitzing et al have seen, except that σ_{xy} was not an integer multiple of $\frac{e^2}{h}$ but a *fraction*. In particular, they were able to observe the fractions $\frac{1}{3}, \frac{2}{5}$ and others. It is a simple matter to argue that, if a Landau level is completelly filled, the Hall conductance has to be quantized. In the case of a *translationally invariant* system a simple argument can be made. Let us imagine that we have an external magnetic field B perpendicular to the sample and that there is an external electric field

$\vec{E}$ paralell to the sample. By coupling the system to a source and a sink of electrons, a current is established. In such a situation, there is a Lorentz force which pushes all the electrons sideways. Also, if a number of Landau levels are completely filled, leaving all others empty, there cannot be any component of the current parallel to $\vec{E}$ since it would require processes which are suppressed by an energy gap equal to $\hbar\omega_c$. If the electric field is small, and the system is translationally invariant, there is a reference frame moving at a velocity $\vec{v}$, relative to the laboratory such that $\frac{\vec{v}}{c} \times \vec{B} = -\vec{E}$. In this frame the electric field is absent. A completely filled Landau level has $N = N_\phi$ electrons. If there are n Landau levels which are filled, the total charge Q is $Q = nN_\phi$. The current $\vec{J}$ is then equal to $\vec{J} = +Qe\vec{v}$. Putting it all together we conclude that the current density, $\vec{j} = \frac{\vec{J}}{L^2}$ has components

$$j_i = \frac{Qe}{L^2} v_i = (\frac{Qec}{BL^2})\epsilon_{ij} E_j \tag{9.3.1}$$

From Eq.(9.83) we conclude that the Hall conductance σ_{xy} , i.e. the coefficient of E_j, is equal to $\frac{Qec}{BL^2}$. By using the fact that there are n filled Landau levels and that the flux BL^2 is equal to N_ϕ times the flux quantum $\frac{hc}{e}$, we get

$$\sigma_{xy} = \frac{Qec}{BL^2} = \frac{nN_\phi ec}{\frac{hc}{e} N_\phi} = n\frac{e^2}{h} \tag{9.3.2}$$

Notice that h, and hence quantum mechanics, comes only through the flux quantum $\frac{hc}{e}$.

This is an appealing argument but it is deceptive. First of all, it does not apply to systems which are not translationally invariant. However, a detailed calculation shows that Eq.(9.3.2) is valid even in that case. The second and most serious problem with this argument is that it cannot predict the existence of the fractional values of σ_{xy}. In fact, the absence of the parallel, or dissipative, component of the current was argued by recalling the fact that , if an integer number of levels is exactly filled, no scattering is possible. If some level is only partially filled, there are states availables for scattering and the argument seems to fall apart. Thirdly, this argument *alone* cannot explain the fact that the effect is actually observed. It cannot explain either the incredible accuracy to which the quantization is measured (one part per million for the integer steps). In the experimental setup, the charge density or the external magnetic field can be varied. In either case, the chemical potential must lie between Landau levels for one Landau level to be filled and the next one to be empty. As the density increases, the chemical potential (i.e. Fermi energy) jumps discontinuously from Landau level to Landau level. It remains fixed at the energy of a given level until the level gets filled completely. This argument suggests that σ_{xy} should be a monotonically increasing function of the electron density. So, why do we see steps?

The resolution of all of these paradoxes has required a significant amount of theoretical effort. The explanation of the *observability* of the steps in σ_{xy} (i.e. the plateaus) involves both the presence of *impurities* and of *states at the edge* of the sample. The *accuracy* of the effect turned out to be connected to

the *topological properties* of the quantum states. The *fractional* effect required the discovery of a new condensed state of matter, the *Laughlin state.*

The in-depth study of all these issues lies far from the main scope of this book. Besides, there are excellent reviews widely available. I will then not discuss the role of disorder beyond a very qualitative description. The role of topology and the Laughlin wave function will be discussed in the next sections.

9.4 The Quantum Hall Effect and Disorder

Let us briefly discuss the role of disorder. In part for the sake of simplicity, but also because the problem is not fully understood, we will only focus in the non-interacting problem. It is widely suspected that disorder is as essential to the observability of the fractional effect as it is to the integer effect. So, we wish to understand why is the integer quantum Hall effect observed in the more disordered samples. We saw above that a simple model of free electrons in Landau levels does not explain the plateaus which are characteristics of the integer Hall effect. The reason behind the monotonic increase was the fact that the Fermi level *jumps* from one Landau level to the next as the level gets filled-up. If there were extra states "in the gap"(i.e. "between Landau levels"), the Fermi energy will have to progress through those levels until they too get filled. However, these extra states should not contribute to the value of σ_{xy} for the plateaus to remain sharp.

Disorder offers a natural way to generate states "between Landau levels". First of all, any degree of randomness, usually represented by a *random potential* $V(\vec{x})$, will lift the degeneracy of each Landau level which become narrow bands. From studies of electron states in random potentials one expects that at least some states should become *localized* [Anderson 58]. In the *absence* of a magnetic field, it is widely believed that *all electronic* states of two dimensional disordered system are localized [Abrahams 79]. The arguments involve both scaling ideas and a mapping of the problem onto a special type of non-linear sigma model [Wegner 79], [Wegner 83]. If the electrons move in the presence of a *weak* magnetic field, the same arguments apply. It turns out that the presence of the field has only two effects: (1) the symmetry of the non-linear sigma model is unitary (which reflects the fact that, in the presence of the field there is no time-reversal invariance) and (2) the presence of a *topological term* in the effective action [Levine 83] and [Pruisken 84]. There is an excellent review by Pruisken on this subject in the book by Prange and Girvin [Prange 90].

The non-linear sigma model represents the physics of the diffusive modes in the presence of the external field. It is a correct description if the elastic mean-free-path λ is *short* compared with the magnetic length $l_0, \lambda \ll l_0$. This condition can only be achieved in the weak field limit. The diffusive modes are represented in the replica formalism, by $2N \times 2N$ hermitian matrices

$Q_{i\alpha,j\beta}(x)$, where $i,j = 1,\ldots,N$ and $\alpha,\beta = \pm$. The latin indices i and j represent the "replicas" and the greek indices, α and β, represent the particle and hole channels. The non-linear sigma model has the effective Lagrangian [Wegner 83] [Levine 83]

$$\mathcal{L} = -\frac{\sigma^0_{xx}}{8}\,\mathrm{tr}[\partial_\mu Q \partial_\mu Q] + \frac{\sigma^0_{xy}}{8}(B)\,\mathrm{tr}[\epsilon_{\mu\nu} Q \partial_\mu Q \partial_\nu Q] \qquad (9.4.1)$$

valid in the replica limit, $N \to 0$.

The coefficients in this Lagrangian σ^0_{xx} and $\sigma^0_{xy}(B)$ represent the values of the longitudinal (σ^0_{xx}) and Hall (σ^0_{xy}) conductance at the length scales of the elastic mean-free-path (i.e. their Boltzmann values). This non-linear sigma model is invariant under global unitary transformations in the coset $U(2N)/U(N)\times U(N)$. Notice that the topological term in Eq.(9.4.1) has the *same* structure as the topological terms that we discussed for antiferromagnets in Chapter 5. At scales l *long* compared with λ but *short* compared with l_0, the effective values of $\sigma_{xx}(l)$ and $\sigma_{xy}(l)$ are strongly renormalized. The non-linear sigma model of Eq.(9.4.1) is asymptotically free which means that $\sigma_{xx}(l) \to 0$ for $l \ll \lambda$. In this infrared limit $\sigma_{xy}(l)$ is quantized, $\sigma_{xy}(l) \to \frac{n}{2\pi}(\frac{e^2}{\hbar})$. This quantization has the same topological origin of the quantization of spin and of the coefficients of the topological terms that we discussed in chapter 5.

Thus, this calculation shows that $\sigma_{xy}$is indeed quantized and that σ_{xx} is zero whenever we are in a plateau. However, the replica limit obscures the physical mechanism by which all of this takes place. It almost seems like magic!. Moreover, the actual mechanism by which the system manages to even support a Hall current is very obscure in this picture. But it does point to the fact that it is the physics of localization that makes the effect *observable* in the first place and that the *topological* properties of the quantum states are responsible for the exact *quantization* of σ_{xy}.

If topology is to be the source of both the quantization and accuracy of the Quantum Hall Effect, it appears that the mechanism which supports the Hall current should not be linked to disorder in an essential way. Halperin proposed that the states which carry the Hall current resides at the *edge* of the system [Halperin 82]. Roughly speaking, the electrons are kept inside the sample by a potential which rises towards the physical edge of the system. On some set of points close to the edge, the potential is equal to the Fermi energy. This set of points constitutes a closed curve. The edge states are the waves of the electron liquid spilling over this curve. The presence of disorder complicates the picture. The landscape of the potential can be quite rough. Semiclassically the ground state can be viewed as a set of equipotential curves. In the high field limit, equipotential curves will generally be closed and enclose regions which are quite small and are occupied by electrons. As the field is lowered, these regions will begin to merge and at some critical value B_c of the field, a percolation phenomenon occurs [Trugman 83]. At B_c there is at least one curve which percolates thoughout the system. This curve is a "new edge" which is thus capable of carrying current. The electron states associated with these "edges" have a very special property: they are "chiral" [Wen 90c].

What this means is that the electrons have to *drift* in the field and hence the *direction* of their motion is determined by the sign of the magnetic field. Roughly, the electrons move at the drift velocity $\frac{cE}{B}$. Since the electrons on the edge states move only in one direction, the only possible effect of impurities on them is only a phase shift of the wave propagating forward. There are no backward scattering processes. Localization is due to multiple scattering process in which forward and backward scattering events interfere so much that the electron is unable to propagate. In the absence of backward scattering, there are no localized states. The edge states carry the full current.

9.5 Linear Response Theory and Correlation Functions

In this section, we derive a set of formulas which will enable us to calculate the Hall conductance, as well as other response functions, in terms of the Green functions of the system. In the next section it will be shown that these formulas, when used to compute σ_{xy} for a system with an energy gap, have a hidden topological structure.

Let us consider the system of fermions coupled to an external electromagnetic field. We will consider cases of the fermions moving in free space and on a lattice in the tight-binding limit. In both cases, the generating functional of the fermion Green functions is a functional integral $\mathcal{Z}[A_\mu]$ which is a functional of the external electromagnetic field A_μ. Let us further assume that A_μ is a small fluctuating component of the external field. The average field $< A_\mu >$ is absorbed in the definition of the system. Under such circumstances, it makes sense to determine $\mathcal{Z}[A]$ in perturbation theory, i.e. a series expansion in powers of A_μ. The leading term in this expansion is known as linear response theory [Fetter 71]. This series can be written in the exponentiated form

$$\mathcal{Z}[A_\mu] = \mathcal{Z}[0] \exp\{\frac{i}{2} \int d^D x \int d^D y A_\mu(x) \Pi_{\mu\nu}(x,y) A_\nu(y) + \ldots\} \qquad (9.5.1)$$

where $\Pi_{\mu\nu}(x,y)$ is the *polarization tensor* and $D = d+1$, d being the dimension of the space. For a tight binding model the integrals are replaced by sums.

The underlying fermiongauge-invariant fermion system system is *gauge-invariant*. Thus, upon an arbitrary local gauge transformation $\phi(x)$

$$\begin{aligned} A_\mu &= A'_\mu + \partial_\mu \phi(x) \\ \psi(x) &= e^{i\frac{e}{\hbar c}\phi(x)}\ \psi'(x) \end{aligned} \qquad (9.5.2)$$

the functional $\mathcal{Z}[A]$ is invariant. Thus, the linear response term must also be gauge-invariant. This is only possible if the polarization tensor $\Pi_{\mu\nu}(xy)$ is *transverse*, i.e.

$$\partial^x_\mu \Pi_{\mu\nu}(x,y) = 0. \qquad (9.5.3)$$

To be more precise, we consider either a system without boundaries or one in which only "small" gauge transformations are allowed , i.e. those transformations which vanish at the boundaries, $\lim_{|x|\to\infty}\phi(x)=0$. If the actual boundaries are to be taken into account, such as in cases in which the system is physically coupled to external leads of batteries or measuring instruments, then the values of the gauge transformations at the boundaries become physical degrees of freedom (i.e. the voltage of a battery.) Similarly, for a system without boundaries, the circulation of the vector potential A_μ around closed loops Γ which wrap around the system are gauge-invariant operators and also represent physical degrees of freedom. An example are the loops Γ which are topologically equivalent to the large circles of a torus. The line integrals $\oint_\Gamma dx_\mu A_\mu$ are the so-called holonomies of the gauge fields on the torus.

The transversality condition Eq.(9.5.3) then follows from a simple algebraic manipulation of the exponent in Eq.(9.5.2)

$$\begin{aligned} I =& \frac{i}{2}\int d^Dx \int d^Dy\, A_\mu(x)\,\Pi_{\mu\nu}(x,y)\,A_\nu(y) \\ =& \frac{i}{2}\int d^Dx \int d^Dy\,[A'_\mu(x)+\partial_\mu\phi(x)]\,\Pi_{\mu\nu}(x,y)\,[A'_\nu(y)+\partial_\nu\phi(y)] \\ =& \frac{i}{2}\int d^Dx \int d^Dy\, A'_\mu(x)\,\Pi_{\mu\nu}(x,y)\,A'_\nu(y)+\delta I \end{aligned} \tag{9.5.4}$$

where δI is given by

$$\begin{aligned} \delta I =& \frac{i}{2}\int d^Dx \int d^Dy\,\{\partial_\mu\phi(x)\,\Pi_{\mu\nu}(x,y)\,A'_\nu(y)+ \\ &+A'_\mu(x)\,\Pi_{\mu\nu}(x,y)\partial_\nu\phi(y)+\partial_\mu\phi(x)\,\Pi_{\mu\nu}(x,y)\,\partial_\nu\phi(y)\}. \end{aligned} \tag{9.5.5}$$

Then $\mathcal{Z}[A]$ is gauge invariant if and only if $\delta I\equiv 0$. Upon an integration by parts we get

$$\begin{aligned} \delta I = -\frac{i}{2}\int d^Dx \int d^Dy\,\{&\phi(x)\,\partial^x_\mu\Pi_{\mu\nu}(x,y)\,A'_\nu(y)+ \\ &+A'_\mu(x)\,\partial^y_\nu\Pi_{\mu\nu}(x,y)\,\phi(y)-\phi(x)\partial^x_\mu\partial^y_\nu\Pi_{\mu\nu}(x,y)\phi(y)\}+ \\ &+\text{surface terms}. \end{aligned} \tag{9.5.6}$$

Since $\phi(x)$ is arbitrary, δI vanishes identically if and only if $\Pi_{\mu\nu}(x,y)$ is transverse. The surface terms are zero sinee either ϕ vanishes at the surface or there are no boundaries.

It is possible to relate $\Pi_{\mu\nu}$ to a fermion current,Green function. The gauge invariant fermion current $J_\mu(x)$ is

$$J_\mu(x)=\frac{\delta S}{\delta A_\mu(x)} \tag{9.5.7}$$

where S is the total action of the system. The current J_μ is gauge invariant because the action S itself is invariant. For the problem of fermions in free space, J_μ is just the usual fermion current with the diamagnetic term included

(the spin is omitted.)

$$J_0 = e\psi^\dagger\psi$$
$$J_j = \frac{e\hbar}{2imc}[\psi^\dagger(\partial_j\psi) - (\partial_j\psi^\dagger)\psi] - \frac{e^2}{mc^2}A_j\psi^\dagger\psi. \tag{9.5.8}$$

The spacial components of the current can be written in the more manifestly gauge invariant form

$$J_j = \frac{e\hbar}{2mic}[\psi^\dagger D_j\psi - (D_j\psi)^\dagger\psi] \tag{9.5.9}$$

where D_j is the covariant derivative

$$D_j = \partial_j - \frac{ie}{\hbar c}A_j \tag{9.5.10}$$

and e is the (negative) electron charge.

For a lattice, J_j has the form

$$J_j(\vec{x}) = \frac{t}{2i}\left(\psi^\dagger(x)\,\exp[\frac{ie}{\hbar c}\int_{\vec{x}}^{\vec{x}+\hat{e}_j}\vec{A}(\vec{z}).d\vec{z}]\;\psi(\vec{x}+\hat{e}_j) - \text{h.c.}\right) \tag{9.5.11}$$

where t is a hopping amplitude and $\hat{e}_j$ is the vector difference of the positions of two lattice sites along the direction j on the lattice.

Since $J_\mu = \frac{\delta S}{\delta A_\mu}$, we can compute expectation values of products of currents by functional differentation of $\mathcal{Z}[A]$. The average current $< J_\mu(x) >$ is given by

$$< J_\mu(x) >= \frac{-i}{\hbar}\frac{1}{\mathcal{Z}[A]}\frac{\delta\mathcal{Z}[A]}{\delta A_\mu(x)}. \tag{9.5.12}$$

The polarization tensor $\Pi_{\mu\nu}(x,y)$ can be computed from its definition. We get

$$\Pi_{\mu\nu}(x,y) = -i\hbar\frac{\delta^2}{\delta A_\mu(x)\delta A_\nu(y)}\ln\mathcal{Z}[A]. \tag{9.5.13}$$

A straightforward algebraic manipulation yields the expression

$$\Pi_{\mu\nu}(x,y) = -i\hbar\frac{\delta}{\delta A_\mu(x)}[\frac{1}{\mathcal{Z}[A]}\frac{\delta\mathcal{Z}[A]}{\delta A_\nu(y)}] =$$
$$= i\hbar(\frac{1}{\mathcal{Z}[A]}\frac{\delta\mathcal{Z}[A]}{\delta A_\mu(x)})(\frac{1}{\mathcal{Z}[A]}\frac{\delta\mathcal{Z}[A]}{\delta A_\nu(y)}) - i\hbar\frac{1}{\mathcal{Z}}\frac{\delta^2\mathcal{Z}[A]}{\delta A_\mu(x)\delta A_\nu(y)}. \tag{9.5.14}$$

Hence, we get

$$\Pi_{\mu\nu}(x,y) = \frac{i}{\hbar} < J_\mu(x)J_\nu(y) >_c + < \frac{\delta J_\mu(x)}{\delta A_\nu(y)} > \tag{9.5.15}$$

where $< J_\mu(x)J_\nu(y) >_c$ is the connected time-ordered current-current correlation function $D_{\mu\nu}(x,y)$ defined by

$$\frac{\hbar}{i}D_{\mu\nu}(x,y) \equiv < J_\mu(x)J_\nu(y) >_c = < J_\mu(x)J_\nu(y) > - < J_\mu(x) >< J_\nu(y) > . \tag{9.5.16}$$

The last term in Eq.(9.5.15) is usually called the "tadpole" term and follows from the diamagnetic piece of the current.

Since $\Pi_{\mu\nu}$ has to be transverse for the system to be gauge invariant, $D_{\mu\nu}$ must obey a similar conservation law. However $D_{\mu\nu}$ is not quite transverse because of the presence of the tadpole term in Eq.(9.5.15). Indeed from the transversality of $\Pi_{\mu\nu}$ we get the equation

$$0 = \partial_\mu^x \Pi_{\mu\nu}(x,y) = \partial_\mu^x D_{\mu\nu}(x,y) + \partial_\mu^x < \frac{\delta J_\mu(x)}{\delta A_\nu(y)} > . \tag{9.5.17}$$

Thus the divergence of $D_{\mu\nu}$ is

$$\partial_\mu^x D_{\mu\nu}(x,y) = -\partial_\mu^x < \frac{\delta J_\mu(x)}{\delta A_\nu(y)} > . \tag{9.5.18}$$

Since $D_{\mu\nu}$ is time ordered and J_ν is conserved ($\partial_\mu J_\mu = 0$) we can write the l.h.s of Eq.(9.5.14), as

$$\begin{aligned}
&\partial_\mu^x D_{\mu\nu}(x,y) = \\
&= \frac{i}{\hbar}\partial_\mu^x < T J_\mu(x) J_\nu(y) > \\
&= \frac{i}{\hbar}\partial_\mu^x [\theta(x_0 - y_0) < J_\mu(x) J_\nu(y) > + \theta(y_0 - x_0) < J_\nu(y) J_\mu(x) >] \\
&= \frac{i}{\hbar}\delta(x_0 - y_0) < [J_0(x), J_\nu(y)] > + \frac{i}{\hbar} < T \partial_\mu^x J_\mu(x) J_\nu(y) > \\
&= \frac{i}{\hbar}\delta(x_0 - y_0) < 0 | [J_0(x), J_\nu(y)] | 0 > .
\end{aligned} \tag{9.5.19}$$

The r.h.s. of Eq.(9.5.14) is equal to

$$\partial_\mu^x < \frac{\delta J_\mu(x)}{\delta A_0(y)} > = 0 \tag{9.5.20}$$

$$\partial_\mu^x < \frac{\delta J_\mu(x)}{\delta A_l(y)} > = \partial_k^x < \frac{\delta J_k(x)}{\delta A_l(y)} > = \partial_k^x [\delta(x-y)(-\frac{e}{mc^2} J_0(x))]. \tag{9.5.21}$$

Collecting terms, we get the following identities for the ground state equal-time expectation value of the commutators

$$\begin{aligned}
\delta(x_0 - y_0) < 0 | [J_0(x), J_k(y)] | 0 > &= \frac{ie}{\hbar mc^2} \partial_k^x [\delta(x-y) J_0(x)] \\
\delta(x_0 - y_0) < 0 | [J_0(x), J_0(y)] | 0 > &= 0
\end{aligned} \tag{9.5.22}$$

which are the Ward-Identities for this system. These identities are the key to the derivation of the f-sum rule [Kadanoff 61]. These identities show that, even though $D_{\mu\nu}(x,y)$ is a correlation function of conserved currents, $D_{\mu\nu}$ itself it is not conserved

$$\begin{aligned}
\partial_\mu^x D_{\mu o}(x,y) &= 0 \\
\partial_\mu^x D_{\mu k}(x,y) &= \frac{ie}{\hbar mc^2} \partial_k^x [\delta(x-y) J_0(x)]
\end{aligned} \tag{9.5.23}$$

$\Pi_{\mu\nu}$ is stricly conserved, $\partial_\mu^x \Pi_{\mu\nu} = 0$. The non-vanishing r.h.s of Eq.(9.5.23) is what in Quantum Field Theory is commonly called *Schwinger term* (S.T.). We

have already encountered a S.T. in section 4.3.1. There, the S.T. resulted from the lack of chiral symmetry in a gauge-invariant theory of one-dimensional relativistic fermions. In a sense, it is due to an effect produced by the "bottom" of the Fermi sea. In the problem discussed in the present section it follows from the definition of the current.

The results of this section are valid in the most general condensed matter systems. They hold regardless of the statistics of the charge carriers. In the derivation that is usually presented in textbooks ([Pines 66],[Mahan 90]), the proof is done within the framework of Fermi Liquid Theory. The argument presented here is more general and follows in spirit the discussion by Kadanoff-Martin. These conservation laws and sum rules are, in fact, a direct consequence of local gauge-invariance. In other words, they follow from local charge conservation. It is important to stress that they also hold in phases with "spontaneously broken gauge invariance," such as superconducting states. The quotation marks are meant to stress that *local* gauge invariance *cannot* be spontaneously broken, as dictated by Elitzur's theorem (see chapter 6). In superconducting states the *global* phase invariance (a subgroup of local gauge transformations) is spontaneously broken in the *absence* of electromagnetic gauge field. The sum rules are a statement about the system as a whole and they hold provided that both the normal and the superfluid contributions are taken into account.

Let us now find an explicit expression for $\Pi_{\mu\nu}(x,y)$ for a simple system. For the sake of simplicity I will discuss only the non-interacting fermion case. Interactions can be introduced in standard fashion. Let us discuss the problem of non-interacting electrons moving in free space coupled to an external electromagnetic field A_μ. Once again, A_μ represents a small fluctuating component with vanishing average. All averages $< A_\mu >$ are absorbed in the definition of the otherwise non-interacting fermions. The action for this system is (ignoring spin)

$$S[A] = \int d^d x \; \psi^*(iD_0 + \mu - h[< A_\mu > + A_\mu])\psi \tag{9.5.24}$$

where $h[A_\mu]$ is a one-particle Hamiltonian which describes the dynamics of particles coupled to a gauge field, such as

$$h[A_\mu] = -\frac{\hbar^2}{2m}\vec{D}_j^2 \tag{9.5.25}$$

and D_0 and D_j are covariant derivatives. The generating functional of the current correlation functions $\mathcal{Z}[A]$ is given by

$$\mathcal{Z}[A] = \int \mathcal{D}\psi^*\mathcal{D}\psi \; \exp\{\frac{i}{\hbar}S[A]\}. \tag{9.5.26}$$

Since the ψ fields represent fermions, we get ($\hbar = 1$)

$$\mathcal{Z}[A] = \mathrm{Det}(iD_0 + \mu - h[A]). \tag{9.5.27}$$

Thus, the effective action for the vector potential A_μ due to the motion of the charged particles is

$$S_{\text{eff}}[A] = -i \text{ tr} \ln \text{Det}(iD_0 + \mu - h[A]). \tag{9.5.28}$$

We have encountered several times expressions of this sort in the previous sections of this book. We will deal with it exactly in the same way. If A_μ is small, $S_{\text{eff}}[A_\mu]$ can be expanded in powers of A_μ and, if A_μ has zero average, the first non-zero term is quadratic in A_μ. A straightforward calculation yields the following expressions for $\Pi_{\mu\nu}(x,y)$ in terms of the one-particle fermion Green function $G(x,y)$. $G(x,y)$ satisfies the equation of motion

$$(iD_0 + \mu - h[< A >])_x \; G(x,y) = \delta(x-y) \tag{9.5.29}$$

i.e.

$$G(x,y) = < x| \frac{1}{iD_0 + \mu - h[< A >]} |y > . \tag{9.5.30}$$

The components of the polarization tensor $\Pi_{\mu\nu}(x,y)$ are ($\hbar = 1$)

$$\Pi_{00}(x,y) = i \; G(x,y)G(y,x) \tag{9.5.31}$$

$$\Pi_{0j}(x,y) = \frac{1}{2m}\{G(x,y)D_j^y G(y,x) - G(y,x)D_j^{y\dagger} G(x,y)\} \tag{9.5.32}$$

$$\Pi_{j0}(x,y) = + \frac{1}{2m}\{-G(x,y)D_j^{x\dagger} G(y,x) + G(y,x)D_j^x G(x,y)\} \tag{9.5.33}$$

$$\begin{aligned}
\Pi_{jk}(x,y) =& \frac{i}{m}\delta(x-y)\delta_{jk}G(x,y) + \\
& - \frac{i}{4m^2}(D_j^x G(x,y))(D_k^y G(y,x)) + \\
& - \frac{i}{4m^2}(D_j^{x\dagger} G(y,x))(D_k^{y\dagger} G(x,y)) + \\
& + \frac{i}{4m^2} G(y,x))(D_j^x D_k^{y\dagger} G(x,y)) + \\
& + \frac{i}{4m^2}(D_j^{x\dagger} D_k^y G(y,x))G(x,y).
\end{aligned} \tag{9.5.34}$$

These formulas, in addition to satisfying the requirements of gauge invariance, are also translation invariant if the external fields are uniform.

In the next section we will make use of these formulas, particularly Π_{0j}, to compute the Hall conductance. Notice that all the expressions in this section hold for *time-ordered* correlation functions. In order to compute the conductivities it is necessary to go to retarded functions [Fetter 71]. Fortunately, the static limit of the Hall conductance can also be calculated directly from the time-ordered functions.

The tight-binding case (on a cubic lattice) can be treated using a similar line of argument. In fact, the polarization tensor $\Pi_{\mu\nu}$ for the lattice case can be obtained in the following manner. First the spacial integrals are replaced by sums over lattice sites $\{\vec{x}\}$. The covariant derivatives are replaced by covariant difference according to the rule

$$D_j^x G(x,y) \to \Delta_j^x G(x,y) \equiv G(x+e_j;y) e^{i\int_x^{x+e_j} \vec{A}.\vec{dl}} - G(x,y) \tag{9.5.35}$$

and the hopping amplitude t and the mass m are related by $\frac{1}{t} = ma_0^2$, where a_0 is the lattice constant. Once these identifications are made, the continuum result becomes valid for the lattice case.

We will be interested primarily in the low frequency, long wave length limit of the effective action. On the basis of gauge and translation invariance we can write the effective action $S_{\text{eff}}[A]$ in terms of an expansion in powers of the gradients of A_μ. The leading order terms are (in two space dimensions)

$$S_{\text{eff}}[A_\mu] = \int d^2x dt [\frac{\epsilon}{2}\vec{E}^2 - \frac{\chi}{2}B^2 + \gamma(\vec{\nabla}\cdot\vec{E})B + \frac{\sigma_{xy}}{4}\epsilon_{\mu\nu\lambda}A_\mu F_{\nu\lambda} + \ldots] \quad (9.5.36)$$

where $\vec{E}$ and B are the fluctuating pieces of the external electromagnetic field. The coefficients ϵ, χ, γ and σ_{xy} can be determined from $\Pi_{\mu\nu}$. In particular ϵ and χ are the static dielectric constant and diamagnetic susceptibility of the system and σ_{xy} is the static Hall conductance. Notice that the Hall term is precisely the Chern-Simons term which we encountered in section (7.1). Indeed, the last term gives a contribution to the average current $< J_k >_{xy}$

$$< J_k >_{xy} = \sigma_{xy}\epsilon_{kl}E_l \quad (9.5.37)$$

which has precisely the correct form for the Hall current. The static Hall conductance σ_{xy} can be obtained from the Fourier transform $\Pi_{\mu\nu}(Q)$ of the polarization tensor

$$\sigma_{xy} = \lim_{Q\to 0} i\frac{\epsilon_{\mu\nu\lambda}Q_\lambda}{Q^2}\Pi_{\mu\nu}(Q) \quad (9.5.38)$$

where $Q = (Q_0, \vec{Q})$.

9.6 The Hall Conductance and Topological Invariance

The most remarkable feature of the Quantum Hall Effect is the quantization of the Hall conductance, i.e. the very existance of the effect itself!. The arguments of the previous section show that σ_{xy} is determined from $\Pi_{\mu\nu}$. However, the coefficients of the gradient expansion of the effective action $S_{\text{eff}}[A]$ are usually renormalized away from the values predicted by a theory of weakly interacting fermions. In effect, the $\Pi_{\mu\nu}$ of the last section is just the leading order (RPA) approximation to the full $\Pi_{\mu\nu}$. Furthermore, the higher order terms of the gradient expansion are also expected to give contributions at lower orders. This is so since the higher order terms are important for wave vector $|\vec{Q}|$ larger than the inverse cyclotron length and frequencies Q_0 larger than the inverse Landau gap. The effective low energy (hydrodynamic) theory is determined by integrating out (or summing over) the high-momentum and high-frequency modes. All these processes will contribute with effective (usually finite) renormalization of the parameters $\epsilon, \chi, \sigma_{xy}$, and γ. On these

grounds, it is not obvious why should σ_{xy} be given exactly by some integer (or fraction) multiple of $\frac{e^2}{\hbar}$.

In a general case (i.e. arbitrary density and arbitrary external field) σ_{xy} does get renormalized. However, there is a special, but very important, case in which σ_{xy} does not get renormalized. This happens whenever the ground state and the lower energy excitations of the system are separated by a non-zero energy gap.

We will show now, following the arguments due to D.J. Thouless and collaborators [Thouless 82], that in this case σ_{xy} is not renormalized by fluctuations. The key to the argument is the observation that σ_{xy} is determined by a *topological invariant*. We will follow the arguments first presented by M. Kohmoto [Kohmoto 85] and by Q. Niu, D. Thouless and Y.S. Wu [Niu 85]. In this section I will discuss the topological invariance in terms of the more general problem of boundary conditions in a many-body system with an enery gap.

(A) The Kubo Formula

Let us consider a system which is in its ground state $|\Psi_0>$ and that there is a gap to all excitations. Let us assume that, in addition to a uniform magnetic field B, the system is allowed to interact with a *small slowly varying* external electromagnetic field. In this limit, perturbation theory reduces to the *adiabatic approximation*. To first order in the time derivative, the perturbed eigenstates are

$$|\Psi_\alpha(t)>= \exp\{-\frac{i}{\hbar}\int_0^t dt'\epsilon_\alpha(t')\}[|\alpha(t)> +i\hbar\sum_{\beta\neq\alpha}\frac{|\beta(t)><\beta(t)|\frac{\partial}{\partial t}|\alpha(t)>}{E_\beta(t)-E_\alpha(t)}] \tag{9.6.1}$$

where $|\alpha(t)>$ is an instantaneous eigenstate of the time dependent Schroedinger equation

$$H(t)|\alpha(t)>= E_\alpha(t)\,|\alpha(t)> \tag{9.6.2}$$

which is a parametric function of t. If we choose the gauge $A_0 = 0$, time enters in the Hamiltonian $\hat{H}$ through the space components of the vector potential which have now the extra term $\delta\vec{A}$

$$\delta\vec{A} = \vec{E}(t)t \tag{9.6.3}$$

where $\vec{E}$ is a very weak, slowly varying electric field. The expectation value of an arbitrary operator $\hat{M}$ in this state is

$$<\Psi_\alpha(t)|\hat{M}|\Psi_\alpha(t)>= i\hbar\sum_{\beta\neq\alpha}\frac{<\alpha|\hat{M}|\beta><\beta|\frac{\partial}{\partial t}|\alpha> + <\alpha|\frac{\partial}{\partial t}|>\beta<\beta|\hat{M}|\alpha>}{\epsilon_\beta(t)-\epsilon_\alpha(t)}. \tag{9.6.4}$$

Let us compute the expectation value of the current operator $\hat{J}_k(x)$. Let us recall that the states $\{|\alpha>\}$ obey the time-dependent Schroedinger equation.

Thus $< \alpha|\frac{\partial}{\partial t}|\beta >$ is given by

$$< \alpha|\frac{\partial}{\partial t}|\beta >= \frac{< \alpha|\frac{\partial \hat{H}}{\partial t}|\beta >}{\epsilon_\beta(t) - \epsilon_\alpha(t)}. \tag{9.6.5}$$

The Hamiltoninan $\hat{H}$ is a slowly varying function of time. But time only enters in $\hat{H}$ through its dependence on the vector potential $\hat{A}$. From this observation, and from the definition of the current as the functional derivative of the Hamiltonian $\hat{H}$, it follows that the Hall conductance σ_{xy} can be written in the form

$$(\sigma_{xy})_\alpha = -i\hbar L_1 L_2 \sum_{\beta \neq \alpha} \frac{< \alpha|\hat{J}_1|\beta >< \beta|\hat{J}_2|\alpha > - < \alpha|\hat{J}_2|\beta >< \beta|\hat{J}_1|\alpha >}{(\epsilon_\beta(t) - \epsilon_\alpha(t))^2}. \tag{9.6.6}$$

This expression is known as the Kubo formula for the Hall conductance σ_{xy}. Analogous formulae can be derived for other components of the conductivity tensor as well as for other transport properties. It is important to stress that the states $\{|\alpha >\}$ are the *exact* eigenstates of the full *many-body system* described by $\hat{H}$ and that $\{\epsilon_\alpha(t)\}$ are the exact energy levels. They should not be confused with the one-particle states and levels of the non-interacting system which are quite different.

(B) Generalized Toroidal Boundary Conditions

There is an alternative approach which yields a more suggestive and useful expression for σ_{xy}. Let us use the Schroedinger Equation to write an equivalent, but more useful, expression for the Hall conductance. Let us imagine that the system under consideration has N particles inside a rectangle of sides L_1 and L_2. Since the external (weak) electric field is taken to be uniform in space, we can write the associated electrostatic potential $U(\vec{x})$ in the form

$$U(\vec{x}) = \vec{E} \cdot \vec{x} \tag{9.6.7}$$

and $\vec{E} = \vec{\nabla} U$. Thus, the extra term in the vector potential $\delta \vec{A}$ is

$$\delta \vec{A} = \vec{E} t = \vec{\nabla}[U(\vec{x})t]. \tag{9.6.8}$$

Since $\delta \vec{A}$ is a pure gradient, it can be eliminated by a suitable gauge transformation of the fermion operator of the form

$$\Psi(\vec{x}) \rightarrow e^{i\frac{e}{\hbar c}U(\vec{x})t}\Psi(\vec{x}). \tag{9.6.9}$$

Notice, however, that such *local* gauge transformations cannot change the value of the circulation of the vector potential $\delta \vec{A}$ on closed non-contractible loops. More specifically the line integrals I_j,

$$I_j = \oint_{\Gamma_j} \delta \vec{A}.d\vec{l} = t \oint_{\Gamma_j} \vec{E}.d\vec{l} \equiv t E_j L_j \tag{9.6.10}$$

on paths Γ_j which wrap around the system on the x_1 and x_2 directions respectively, are gauge invariant if the fermions move on the torus. Thus, although

the vector potential $\delta\vec{A}$ disappears from the problem, the holonomies do not. In fact, they enter in the boundary conditions. Line integrals of a gauge field on non-contractible loops in space (or space-time) are called the *holonomies* of the gauge field.

The problem of assigning boundary conditions to quantum mechanical systems on a closed manifold is a very subtle one. For instance, if the fermions move on a torus and no magnetic field is present, it is perfectly consistent to use periodic or twisted boundary conditions which, for an N-particle system, are

$$\Psi(\vec{x}_1,\ldots,\vec{x}_N) = e^{-i\vec{\theta}\cdot\vec{L}}\Psi(\vec{x}_1+\vec{L},\ldots,\vec{x}_N+\vec{L}) \tag{9.6.11}$$

where $\vec{\theta}$ is an arbitrary two-component vector and $\vec{L}$ is a displacement along x_1 by a distance L_1 or x_2 by a distance L_2. These boundary conditions are perfectly consistent since, in the absence of magnetic field, the total momentum $\vec{P}$ is a constant of motion. The momentum of the only eigenstate compatible with the boundary conditions is $\frac{\vec{\theta}}{|\vec{L}|}$. But, if a magnetic field is present, the situation is somewhat different. In section (9.1) we introduced the *magnetic translation operators*. These operators commute with the one-particle Hamiltonian. In fact, they also commute with the Hamiltonian of the full interacting system. In section (9.1), we also found that the only consistent boundary conditions for the wave functions (generalized now to the N-particle case) for charged particles moving on a torus in the presence of a non-zero magnetic field B are

$$\begin{aligned}
&a_1(x_1, x_2+L_2) = a_1(x_1,x_2) + \partial_1\beta_2(x_1,x_2)\\
&a_2(x_1+L_1, x_2) = a_2(x_1,x_2) + \partial_2\beta_1(x_1,x_2)\\
&\Psi(\{x_1^{(j)}+L_1\};\{x_2^{(j)}\}) = \exp\{\frac{-ie}{\hbar c}\sum_{j=1}^{N}\beta_1(x_1^{(j)};x_2^{(j)}) + i\theta_1\}\Psi(\{x_1^{(j)}\};\{x_2^{(j)}\})\\
&\Psi(\{x_1^{(j)}\};\{x_2^{(j)}+L_2\}) = \exp\{\frac{-ie}{\hbar c}\sum_{j=1}^{N}\beta_2(x_1^{(j)};x_2^{(j)}) + i\theta_2\}\Psi(\{x_1^{(j)}\};\{x_2^{(j)}\})
\end{aligned} \tag{9.6.12}$$

where we have included the effect of the electric fields through the angles θ_1, θ_2. These boundary phases are related to the electric field by

$$\theta_j = \frac{et}{\hbar c}E_j L_j \equiv \frac{et}{\hbar c}I_j. \tag{9.6.13}$$

Thus, in addition to the phase twist $\vec{\theta}$, the requirement that the states be eigenstates of the magnetic translation operator leads naturally to the generalized boundary conditions. We will see below that the additional phase factors arise from the impossibility of defining the phase of the wave function globally and smoothly on the torus. The wave functions for particles on a torus in the presence of a magnetic field form a fiber bundle. The conditions, or rather the requirement that the states be eigenstates of the magnetic translations, define the fiber bundle. We will see below, that a similar difficulty arises when one tries to define the dependence of the phase of the wave

function on the twist angles $\vec{\theta}$.

(C) The Kubo Formula for σ_{xy} and the First Chern Character

From now on, we will assume that the vector $\vec{\theta}$ represent two constant angles. In any case all the time dependence of the states enters through $\vec{\theta}$. All time derivatives become derivatives relative to the phase θ_j. The Kubo formulas can now be written in the form ($\partial_j \equiv \partial_{\theta_j}$)

$$(\sigma_{xy})_\alpha = \frac{ie^2}{\hbar}[\partial_1 < \alpha|\partial_2|\alpha > -\partial_2 < \alpha|\partial_1|\alpha >]. \tag{9.6.14}$$

In this form, this formula was first derived by Q. Niu, D. Thouless and Y.S. Wu [Niu 85]. They also considered the *average* $< (\sigma_{xy}) >$ over the torus of boundary conditions

$$\begin{aligned} < (\sigma_{xy})_\alpha > &= \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \, (\sigma_{xy})_\alpha \\ &= \frac{e^2}{i\hbar} \int_0^{2\pi} \int_0^{2\pi} d\theta_1 d\theta_2 \, [\partial_2 < \alpha|\partial_1|\alpha > -\partial_1 < \alpha|\partial_2|\alpha >] \end{aligned} \tag{9.6.15}$$

What matters to our discussion is the fact that $< (\sigma_{xy})_\alpha >$ is proportional to a quantity, known as the first Chern number C_1, which is a topological invariant. A similar expression also appears in the tight-binding case, which will be discussed below.

Before we consider what the average conductance is, we must face the fact that in any physically relevant situation the boundary conditions are fixed. Thus it might appear that while $< (\sigma_{xy})_\alpha >$ may be an interesting quantity to compute, it is not directly relevant. This is true. However, we are considering a special situation in which there is a finite energy gap between the ground state and the first excited state. It is easy to argue that, if the gap is finite, then the difference of the *measured* value of $(\sigma_{xy})^\theta_\alpha$, with *fixed* boundary conditions, and $< (\sigma_{xy})_\alpha >$ vanishes in the thermodynamic limit, at least like $\frac{1}{L}$.

Let us consider the differential change $\frac{\partial(\sigma_{xy})^\theta_\alpha}{\partial\theta_1}$. The dependence of the conductance on the phase angles θ_1 and θ_2 enters through the Hamiltonian $\hat{H}$. But $\hat{H}$ is a function of $\frac{\theta_1}{L_1}$ and $\frac{\theta_2}{L_2}$ only. Thus, all changes must be of the form $\frac{1}{L_j}\frac{\partial(\sigma_{xy})_\alpha}{\partial(\frac{\theta_j}{L_j})}$ $(j = 1, 2)$. Since there is a non-vanishing gap, all small changes in the parameters of the Hamiltonian $\hat{H}$ must lead to changes of order unity in all local quantities. This includes, changes in the energies and wave functions of local excitations. Thus, the derivatives $\frac{\partial(\sigma_{xy})_\alpha}{\partial(\frac{\theta_j}{L_j})}$ must have finite limits for thermodynamically large systems. Hence, $\frac{\partial(\sigma_{xy})_\alpha}{\partial\theta_j}$ behaves like $\frac{\text{const}}{L_1}$ for asymptotically large systems. This justifies the use of the conductance averaged over all boundary conditions [Niu 85].

(D) Fiber Bundles and the Quantum Hall Conductance

Let us now turn to the issue of the topological invariance of $< (\sigma_{xy})_\alpha >$. The argument goes as follows. The boundary condition angles θ_1 and θ_2, being phases, are defined modulo 2π. Each choice of a boundary condition amounts to a choice of a point $\vec{\theta}$ on the torus $S_1 \times S_1$ of boundary conditions. For each point $\vec{\theta}$ we have a unique eigenstate $\Psi_\alpha(\{\vec{x}\};\vec{\theta})$ of the full many-body Hamiltonian $\hat{H}$. In mathematical jargon, we have a *fiber bundle*. The wave function has an amplitude and a phase which are smooth functions of $\vec{\theta}$. Now, the total phase of the wave function is not a physical observable. But *changes* of the phase are. In particular, let us imagine that, at some initial time t_0, we have defined an initial boundary condition $\vec{\theta}(t_0)$ with a phase for the state $\arg\Psi(\theta(t_0))$. The external electromagnetic field is now allowed to couple to the system in such a way that the boundary conditions change as a function $\vec{\theta}(t)$ and return to the initial value $\vec{\theta}(t_0)$ after some very long time T. During this process the vector $\vec{\theta}(t)$ traces a closed curve Γ on the torus $S_1 \times S_1$. At the same time, the phase of the wave function changes by an amount δ_Γ

$$\delta_\Gamma = \Delta\arg\Psi = \Delta\mathrm{Im}\ln\Psi = \arg\Psi(\theta(t_0+T)) - \arg\Psi(\theta(t_0)) \qquad (9.6.16)$$

If Ψ is an *analytic non-vanishing function* of $\vec{\theta}$, the phase change, δ_Γ must be zero. This is so because, in such case the contour can be deformed to zero. However, the only analytic function on a torus is a constant. Thus, a non-vanishing adiabatic phase change δ_Γ requires that the function $\ln\Psi$ be non-analytic on the torus of boundary conditions. In this case, closed contours which enclose singularities of $\ln\Psi$ are non-contractible and δ_Γ is non-zero for such contours. Non-zero adiabatic changes of the phase of wave functions of quantum mechanical systems are known as *Berry phases* [Berry 84] and [Simon 83]. Since the wave function $\Psi_\alpha(\vec{x}_1,\ldots,\vec{x}_N;\vec{\theta})$ is a smooth function of its arguments, a non-analyticity in $\ln\Psi$ amounts to zeroes of Ψ for some values of $\vec{\theta}$. Smoothness requires that the zeroes be isolated points on the torus $S_1 \times S_1$. The Berry phase δ_Γ counts the number of zeroes of Ψ enclosed by the contour Γ.

How is the phase of the wave function $\Psi^{(\alpha)}(\{\vec{x}\};\vec{\theta})$ related to the Hall conductance? In order to investigate this issue let us introduce the following suggestive notation originally introduced by Kohmoto. Let $\mathcal{A}_k^{(\alpha)}(\theta_1,\theta_2)$ be on a vector field on the torus $S_1 \times S_1$ defined by

$$\mathcal{A}_k^{(\alpha)} = i < \alpha|\frac{\partial}{\partial\theta_k}|\alpha > \equiv i < \Psi_{\vec{\theta}}^{(\alpha)}|\frac{\partial}{\partial\theta_k}|\Psi_{\vec{\theta}}^{(\alpha)} > . \qquad (9.6.17)$$

With this notation, the averaged Hall conductance is

$$< (\sigma_{xy})_\alpha >= \frac{e^2}{\hbar}\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2\,(\partial_1\mathcal{A}_2 - \partial_2\mathcal{A}_1). \qquad (9.6.18)$$

This is the Niu-Thouless-Wu formula [Niu 85].

In other words, $< (\sigma_{xy})_\alpha >$ is the flux *through the torus* $S_1 \times S_1$ *of the vector field* $\vec{\mathcal{A}}(\vec{\theta})$.

Furthermore, the states $|\Psi^{(\alpha)}(\vec{\theta})>$ are defined up to an overall phase factor. Thus, the states $|\Psi^{(\alpha)}(\vec{\theta})>$ and $e^{if(\vec{\theta})}|\Psi^{(\alpha)}(\vec{\theta})>$ are physically equivalent. Notice that the phase factor does not modify the boundary conditions. Under a phase change, the vector field $\vec{\mathcal{A}}(\vec{\theta})$ transforms like a gauge transformation.

$$\mathcal{A}_k(\vec{\theta}) = i < \alpha|\partial_k|\alpha > \rightarrow i < \alpha|\partial_k|\alpha > -\partial_k f(\vec{\theta}). \tag{9.6.19}$$

Thus phase factors in the wave functions translate into a gauge transformation for the vector field $\mathcal{A}_k$ defined on the torus of boundary conditions. We can now use Stoke's Theorem to write the averaged Hall conductance in the form

$$< (\sigma_{xy})_\alpha >= \frac{e^2}{\hbar} \oint_{\bar{\Gamma}} \mathcal{A}_k(\vec{\theta})\, d\theta_k \tag{9.6.20}$$

where $\bar{\Gamma}$ is the rectangular contour with corners at (θ_1, θ_2), $(\theta_1 + 2\pi, \theta_2)$, $(\theta_1, \theta_2 + 2\pi)$, $(\theta_1 + 2\pi, \theta_2 + 2\pi)$. A *non-zero Hall conductance* means that the vector field $\vec{\mathcal{A}}$ *cannot be a periodic function on the torus* $S_1 \times S_1$ of boundary conditions. This, in turn, implies that along non-contractible closed contours Γ_1 and Γ_2, which wrap around the torus along the θ_1 and θ_2 directions respectively, $\mathcal{A}_k$ and the wave functions must change like

$$\mathcal{A}_k(\theta_1 + 2\pi, \theta_2) = \mathcal{A}_k(\theta_1, \theta_2) + \partial_k f_1(\theta_1, \theta_2) \tag{9.6.21}$$

$$\mathcal{A}_k(\theta_1, \theta_2 + 2\pi) = \mathcal{A}_k(\theta_1, \theta_2) + \partial_k f_2(\theta_1, \theta_2) \tag{9.6.22}$$

$$\Psi^{(\alpha)}(\{\vec{x}\}; \theta_1 + 2\pi, \theta_2) = e^{if_1(\theta_1,\theta_2)}\Psi^{(\alpha)}(\{\vec{x}\}; \theta_1, \theta_2) \tag{9.6.23}$$

$$\Psi^{(\alpha)}(\{\vec{x}\}; \theta_1, \theta_2 + 2\pi) = e^{if_2(\theta_1,\theta_2)}\Psi^{(\alpha)}(\{\vec{x}\}; \theta_1, \theta_2). \tag{9.6.24}$$

This topological structure is strongly reminiscent of the Wu-Yang construction for the wave functions of charged particles moving in the presence of a Dirac magnetic monopole [Eguchi 80] and [Nash 83]. We can make the analogy even sharper. Let us suppose that we have a wave function $\Psi_\alpha(\{\vec{x}\}, \vec{\theta})$ which satisfies boundary conditions determined by the point $\vec{\theta}$ on $S_1 \times S_1$. Now, given $\Psi_\alpha(\{\vec{x}\}, \vec{\theta})$, can we unambiguosly and completely determine $\Psi_\alpha(\{\vec{x}\}, \vec{\theta}')$ for some other *arbitrary* point $\vec{\theta}'$ on $S_1 \times S_1$?. The answer to this question is *no*. The phase of Ψ_α *cannot* be determined uniquely and smoothly over the boundary condition torus *unless* the Hall conductance is equal to zero. This is so because at the zeroes of Ψ_α its phase is undefined. Let us consider the simpler case of a wave function which vanishes at just one point $\vec{\theta}_0$ on $S_1 \times S_1$. We now split the torus $T \equiv S_1 \times S_1$ into two disjoint subsets T_I and T_{II} such that $\vec{\theta}_0$ is in T_I. Since T_{II} does not contain point $\vec{\theta}$ in which Ψ_α is zero the phase of Ψ_α can be determined globally on T_{II}. For instance, we can choose to make Ψ_α real on T_{II}. However, on T_I there is a point $\vec{\theta}_0$ in which Ψ_α does vanish. We can always define the phase of Ψ_α at $\vec{\theta} = \vec{\theta}_0$ to be some arbitrarily chosen value. Once a value is chosen, the phase of Ψ_α can be defined by continuity on an arbitrary neighborhood of $\vec{\theta}_0$ which is not equal to the whole torus T. Thus we have two different definitions of the phase of Ψ_α on T_I and T_{II}. Obviously these definitions must amount to a gauge transformation, i.e.

$$\Psi^I_\alpha(\{\vec{x}\}, \vec{\theta}) = e^{if(\vec{\theta})}\Psi^{II}_\alpha(\{\vec{x}\}, \vec{\theta}) \tag{9.6.25}$$

where $f(\vec{\theta})$ is a smooth function on the closed curve γ which is the boundary between T_I and T_{II}. The vector field $\mathcal{A}_k$ also has two different definitions on T_I and T_{II}, which differ again by a gauge transformation

$$\mathcal{A}_k^I(\vec{\theta}) - \mathcal{A}_k^{II}(\vec{\theta}) = \partial_k f(\vec{\theta}). \tag{9.6.26}$$

The Hall conductance reduces to a sum of two contributions, one from T_I and the other from T_{II}. Since both surfaces have boundaries, we can readily use Stoke's Theorem to write

$$\begin{aligned} < (\sigma_{xy})_\alpha > &\equiv \frac{e^2}{\hbar}\{\int_{T_I} (\partial_1 \mathcal{A}_2 - \partial_2 \mathcal{A}_1) + \int_{T_{II}} (\partial_1 \mathcal{A}_2 - \partial_2 \mathcal{A}_1)\} \\ &= \frac{e^2}{\hbar}\{\int_\gamma \vec{\mathcal{A}}_I d\vec{\theta} - \int_\gamma \vec{\mathcal{A}}_{II} d\vec{\theta}\} \end{aligned} \tag{9.6.27}$$

where we have used the fact that the boundaries of T_I and T_{II} have opposite orientation. Thus, we find

$$< (\sigma_{xy})_\alpha >= \frac{e^2}{\hbar}\int_\gamma (\vec{\mathcal{A}}_I - \vec{\mathcal{A}}_{II}) \cdot d\vec{\theta} = \frac{e^2}{\hbar}\int_\gamma \vec{\partial} f \cdot d\vec{\theta}. \tag{9.6.28}$$

Thus, $< (\sigma_{xy})_\alpha >$ counts the number of times the gauge transformation $f(\vec{\theta})$ winds around 2π as $\vec{\theta}$ traces the closedwinding number loop γ. The *winding number* C_1

$$C_1 = \frac{1}{2\pi}\int_\gamma \vec{\partial} f \cdot d\vec{\theta} \tag{9.6.29}$$

is a topological invariant known as the first Chern character. It is a topological invariant since it cannot change by any smooth deformation of the contour γ. However, if under a deformation, one or more additional zeroes of Ψ_α cross the boundary into region T_I, the winding number will jump by integer amounts. These processes correspond to crossings of energy levels.

The fiber bundle associated with this problem can be defined in the following way [Kohmoto 85]. To every point $\vec{\theta}$ on T we associate a state $\Psi_\alpha(\vec{\theta})$. States $\Psi'_\alpha(\vec{\theta})$, which differ from $\Psi_\alpha(\vec{\theta})$ by a gauge transformation, $f(\vec{\theta})$, are physically equivalent. Thus, at every point $\vec{\theta} \epsilon T$ we have associated the *ray or bundle* of states related to $\Psi_\alpha(\vec{\theta})$ by a gauge transformation. The torus T is partitioned into a union of sets $T_I, T_{II}, \ldots$ each containing *at most* one zero of Ψ_α. The phase of Ψ_α is defined on each set which results on a set of state vectors $\Psi_\alpha^I, \Psi_\alpha^{II}, \ldots$, whose phases are smoothly defined on $T_I, T_{II}, \ldots$. These state vectors differ from each other just by gauge transformations which are smooth functions $f(\vec{\theta})$ on the overlap between two regions, say T_I and T_{II}. The *transition function* $f(\vec{\theta})$ is a smooth map from the closed curve $\gamma \subset T_I \bigcap T_{II}$ to the group $U(1)$ of phases $e^{if(\vec{\theta})}$. Since γ is isomorphic to $U(1)$ the transition function is a smooth map from $U(1)$ onto $U(1)$. These maps can be classified into *homotopy classes*, each class defined by the winding number C_1 of equation Eq.(7.199). This map is known as the principal $U(1)$ bundle over the torus T. The vector field $\mathcal{A}_k(\vec{\theta})$ defines a *connection*.

Let's define now the one-form $d\mathcal{A} = \mathcal{A}_k \, d\theta_k$. A connection one-form can be written as $\Omega = \mathcal{A} + d\mathcal{A}$. The transition functions act on fibers (i.e. state vectors) by multiplication. Once a connection $\mathcal{A}_k$ is given a curvature two-form $F = d\mathcal{A}$ can be defined and it is known as the first Chern form. The integral of this two-form is the first Chern number.

Let us now note the following interesting analogy. In section (9.1) we discussed the problem of the quantization of the motion of a charged particle in an uniform magnetic field with the particle constrained to move on the surface of a torus in *space*. There, we found how the wave functions transform under magnetic translations. In this section, we showed how to construct the wave function on different *patches* of the torus of boundary conditions. The relation between the wave functions on different patches is analogous to the way the wave functions transform under magnetic translations. However, here we are discussing phases of many-body wave functions on the torus of boundary conditions! At the root of this analogy is the fact that the many-body wave functions are also representations of the group of magnetic translations. Here too, if the wave functions $\Psi^{(\alpha)}(\{\vec{x}\},\vec{\theta})$ are required to be single valued functions on the torus $S_1 \times S_1$, the same consistency condition discussed in section (9.1) implies that the total flux through the torus should be an *integer* n multiple of 2π. Otherwise different paths from $\vec{\theta} = (0,0)$ to, say, $\vec{\theta} = (2\pi, 2\pi)$ would lead to inequivalent phases for the wave function $\Psi^{(\alpha)}$. We conclude that, in this case, the averaged Hall conductance is quantized to be an integer multiple of $\frac{e^2}{\hbar}$.

This argument is actually much too strong. In fact, it appears to require that $< (\sigma_{xy})_\alpha >$ should *always* be an integer multiple of $\frac{e^2}{\hbar}$. The actual observation of the Fractional Hall Effect, as well as the success of Laughlin's theory, indicates that this argument has to be relaxed. Indeed, the observation of FQHE which has $\sigma_{xy} = \frac{e^2}{\hbar}(\frac{n}{m})$ requires that for the case of toroidal boundary conditions, the wave functions $\Psi^{(\alpha)}(\{\vec{x}\};\vec{\theta})$ must be *multivalued* functions on the *torus of boundary conditions*. This means that the eigenstates of $\hat{H}$ must have several components and behave like vectors under periodic changes of boundary conditions. Hence, rather than requiring that $\Psi_\alpha(\{\vec{x}\};\vec{\theta})$ be single valued on the torus $S_1 \times S_1$, we should demand that Ψ_α should have m components (where m is some integer). The wave function returns to its initial value only after the torus has been covered m times. In this case the averaged Hall conductance is equal to $\frac{n}{m}(\frac{e^2}{\hbar})$. The integers n and m cannot be determined by topological arguments alone. They have to be calculated from some microscopic theory. In the next sections we will discuss a few examples: (a) free electrons filling up one Landau level, (b) the tight-binding Hofstadter problem and (c) Laughlin's theory of the Fractional Quantum Hall Effect. In each case n and m turn out to be different. However, the importance of the topological argument is that, for the specific task of computing σ_{xy}, it *suffices* to consider just some simple limit in which the calculation can be done easily. The topological invariance of $< \sigma_{xy} >$ insures that it cannot change under smooth deformations of the underlying Hamiltonian (unless, of course,

during this process there is a level crossing)

(E) How Many Components does the Wave Function Have?

Let us point out that subtle, but important, differences in the behavior of the system arise depending on the choice of boundary conditions. In this section we have considered mainly the case of generalized periodic boundary conditions (GPBC'S.) These GPBC'S require that the fermions move on a two-dimensional torus in space. While this choice is *convenient* from the point of view of mathematics, it is not very natural from an experimental standpoint. Experimentally, the natural choice is a rectangle coupled to a four-point probe, which is a set of sources and sinks of charge. In practice this means taking charge from one point on the edge of the sample (sink) and injecting it back into the system at another point (source). Typically this process involves the use of wires, batteries, etc. In a sense the *measuring devices* implement the generalized periodic boundary conditions. The voltage drop accross the device is proportional to the boundary condition angles θ_1 and θ_2. Yet, another physically relevant situation is a *disk* without wires. If the disk is isotropic and thermodynamically large, then the wave function vanishes exponentially fast as the difference of the particle coordinates becomes large. This can happens due to the presence of an isotropic potential which confines the particles inside some region of the disk. In this case the points on the edge of the disk are asymptotically equivalent to each other. The thermodynamic limit of this case is thus identical to that of a set of particles moving on the surface of a sphere with uniform radial magnetic field. i.e. a magnetic monopole [Haldane 83]. Niu, Thouless and Wu [Niu 85] observed that GPBC'S *require* multicomponent wave functions. Spherical (or disk-like) boundary conditions have only one-component wave functions [Laughlin 83] and [Haldane 83]]. This issue has caused a great deal of confusion, which was partly due to the fact that the components of the wave functions for GPBC'S were originally thought as resulting from the spontaneous breakdown of some unknown discrete symmetry. Indeed, in systems in which a global discrete symmetry is spontaneously broken, there are a finite number of degenerate ground states which are related by a symmetry operation. This phenomenon is quite common in magnetic systems with discrete symmetries. The most common example is the Ising Model. It is also present in commensurate charge-density-wave systems, such us polyacetylene chains. However, these analogies are quite misleading. In the case of Quantum Hall systems, the multicomponent structure is a feature of the *entire* Hilbert space, not just of the ground state. The Hilbert space is split into a number of disconnected pieces not related by a symmetry operation. In other words, this structure is not the result of the spontaneous breakdown of any symmetry. Rather, this feature of the Hilbert space merely reflects the global non-triviality of the manifold on which the particles move. As a matter of fact, the number of components of the wave functions is different on different manifolds [Wen 90a]. For example, instead of a torus, let us consider a sphere. All closed loops on the surface of a sphere are contractible. Thus,

all the holonomies are trivial. The wave functions for charged particles moving on the surface of the sphere in the presence of a uniform radial magnetic field (i.e. a magnetic monopole) still form non-trivial fiber bundle, known as the monopole bundle [Wu 75]. But the arguments given above indicate that the states are now non-degenerate.

What is the physical significance of this degeneracy?. There are two schools of thought on this issue. According to one school, the degeneracy should not be regarded as being physical since it changes with boundary conditions. According to this point of view, the degeneracy merely reflects the fact that the location of the center of mass is quantized if the system is placed on a torus. Indeed, Haldane [Haldane 85b] has given a detailed study of the symmetries of the states on the torus and showed that degeneracy arises from the magnetic translations of the center of mass independently of the physical properties of the system. He further showed that, in general, there are no additional degeneracies and that the states for the relative coordinates are generally non-degenerate. But, for the same token, it is clear that there are no states on a sphere that can carry a current. Thus, if we wish to construct a state with a non zero current we must put the system on the torus. Wen [Wen 90b] has given a very general argument which shows that if the surface on which the fermions move have g handles (genus g Riemann surface) the degeneracy is k^g, if k is the degeneracy on the torus. From this point of view, the topological degeneracy is a fundamental qualitative feature of the system.

9.7 Quantized Hall Conductance of a Non-Interacting System

In this section we will discuss the fairly simple but interesting problem of the computation of the Hall conductance for an assembly of non-interacting electrons moving freely on a torus . We will assume that the external magnetic field and the electron density are such that there are an *integer* number of completely filled Landau levels.

Let us begin by discussing the nature of the one-particle states. Let $\vec{x}$ denote the coordinate of a particle of charge e and mass m. The magnetic field is B and the torus has linear dimensions L_1 and L_2 along its main circles. In section (9.1) we constructed the single particle state for the case of an isotropic disk. For simplicity we will restrict our discussion to the case of particles on the lowest Landau level. In section (9.1) we found that the single particle states for the lowest Landau level $\Psi(z,\bar{z})$ have the form

$$\Psi(z,\bar{z}) = f(z,\bar{z})e^{-|z|^2/4l_0^2} \tag{9.7.1}$$

where $f(z,\bar{z})$ is an analytic function, i.e. $\partial_{\bar{z}} f = 0$.

A basis of (analytic) functions are the powers z^m. For a system with N_ϕ flux quanta there are N_ϕ linear independent states. Thus, an arbitrary

state in the lowest Landau level is a polynomial in z of degree N_ϕ times the exponential factor.

Let us consider now the case of a system with exactly $N = N_\phi$ electrons in a magnetic field B with N_ϕ flux quanta. The ground state wave function Ψ_N for the N-particle system is the Slater determinant

$$\Psi_N(z_1,\ldots,z_N) = \begin{vmatrix} 1 & \ldots & 1 \\ z_1 & \ldots & z_N \\ \vdots & \vdots & \vdots \\ z_1^N & \ldots & z_N^N \end{vmatrix} \exp\{-\frac{1}{4l_0^2}\sum_{j=1}^{N}|z_j|^2\}. \tag{9.7.2}$$

This determinant has the Vandermonde form. The wave function Ψ_N can be written in the form

$$\Psi_N(z_1,\ldots,z_N) = \prod_{1\le j\le k\le N}(z_j - z_k)\ \exp\{-\frac{1}{4l_0^2}\sum_{j=1}^{N}|z_j|^2\}. \tag{9.7.3}$$

We want to compute the Hall conductance for this system. We will use the Niu-Thouless-Wu formula. However, in order to use that formula we need to write down a wave function which is an explicit function of the boundary condition angles θ_1 and θ_2. What we need is to generalize the state for a system on a torus (instead of a disk) of linear dimensions L_1 and L_2 satisfying the generalized periodic boundary conditions of Haldane [Haldane 85a], whose work I follow here. Since toroidal boundary conditions break rotational invariance, it is more natural to work in the axial (or Landau) gauge $A_1 = -Bx_1, A_2 = 0$. It can be easily checked that the wave functions for the states in the lowest Landau level have the form

$$\Psi(x_1,x_2) = f(z)\ e^{-\frac{x_2^2}{2l_0^2}} \tag{9.7.4}$$

where $z = x_1 + ix_2$ and $f(z)$ is an analytic function.

The generalized periodic boundary conditions imply that $f(z)$ must satisfy the consitency conditions

$$\begin{aligned} f(z+L_1) &= e^{i\theta_1} f(z) \\ f(z+iL_2) &= e^{i\theta_2 - i\pi N_s[(\frac{2z}{L_1})+\tau]} f(z) \end{aligned} \tag{9.7.5}$$

where $\tau = \frac{iL_2}{L_1}$. The analytic function $f(z)$ must have zeros inside the rectangle with vertices at $\frac{L_1}{2}(\pm 1 \pm \tau)$. Thus, $f(z)$ must have N_s zeroes. Indeed, the integral of $\frac{f'(z)}{f(z)}$ around the edges of the rectangle is equal to N_s since the total change of the *phase* of $f(z)$ is $2\pi N_s$.

The functions $f(z)$ which are analytic inside the rectangle and satisfy the consistency conditions, must be all analytic functions with exactly N_s zeroes. The most general form that $f(z)$ can take is [Haldane 85a]

$$f(z) = e^{ikz}\prod_{j=1}^{N_s}\vartheta_1(\frac{z-z_j}{L_1}|\tau) \tag{9.7.6}$$

where $\vartheta_1(u|\tau)$ is the first odd elliptic theta function defined by Ederlyi [Ederlyi 53]

$$\vartheta_1(u|\tau) = i \sum_{n=-\infty}^{+\infty} (-1)^n \exp[i\pi\tau(n-\frac{1}{2})^2 + i\pi(2n-1)u]. \tag{9.7.7}$$

The parameter k is a real number in the range $0 \leq |k| \leq \pi N_s \frac{L_2}{L_1}$. The solutions are thus parametrized by the set of N_s complex numbers $\{z_j\}$ which determine the location of the zeros of the function $f(z)$, and by k. By direct substitution we find that k and $z_0 \equiv \sum_{j=1}^{N_s} z_j$ are the solutions to the set of equations

$$\begin{aligned} e^{i\theta_1} &= e^{ikL_1(-1)^{N_s}} \\ e^{i\theta_2} &= e^{-kL_2 + i\pi\frac{z_0}{L_1}} \end{aligned} \tag{9.7.8}$$

which have the unique solution $k = \frac{\theta_1 + \pi N_s}{L_1}$ and $z_0 = \frac{\theta_2 L_1}{\pi} - ik\frac{L_1 L_2}{\pi}$. The location of the zeroes is determined by requiring that the wave functions $f(z)$ form a complete set of orthogonal wave functions which are eigenstates of the magnetic translation operators. A simple way to construct such a set (i.e. a basis for the Hilbert space of the lowest Landau levels) is to choose a set of zeros which satisfies $z_{j+i} = z_j + \frac{L_1}{N_s}$. Thus the dimension of the Hilbert space equals N_s as it should be.

The N-particle states are constructed very much in the same fashion. Here we consider the case of $N = N_s$ particles and, once again, we have filled up the lowest Landau level. The only difference here is that we will separate the coordinates $z = \sum_{j=1}^{N} z_j$ for the center of mass (C.M.) of the system from the set of relative coordinates $z_j - z_k$. The antisymmetric N-particle wave function Ψ_N has the form

$$\Psi_N = \mathcal{N}\,\Psi_{CM}(z) \prod_{1\leq j<k\leq N} f(z_j - z_k)\, \exp\{-\sum_{j=1}^{N} \frac{(x_2^j)^2}{2l_0^2}\} \tag{9.7.9}$$

where x_2^j is the x_2-coordinate of the j^{th} particle and $\mathcal{N}$ is a normalization constant. The wave functions Ψ_{CM} and $f(z)$ for the center of mass and relative coordinates are determined by demanding Ψ_N to satisfy the GPBC's. The "pair-wave-functions" $f(z_j - z_k)$ do not change if *all* particles are (magnetically) translated simultaneously. Only the CM wave function Ψ_{CM} is sensitive to a uniform translation of the system as a whole. On the other hand if a particle (say the j^{th} particle) is transported around the torus exactly once, the wave function must change by a sign determined by its antisymmetry property. These conditions can be met by requiring $\Psi_{CM}(z)$ and $f(z)$ to satisfy

$$\begin{aligned} f(z+L_1) &= f(z) \\ f(z+iL_2) &= f(z)\exp[i\pi(\frac{2z}{L_1}+\tau)] \end{aligned} \tag{9.7.10}$$

$$\begin{aligned} \Psi_{CM}(z+L_1) &= e^{i\theta_1}(-1)^{N-1}\Psi_{CM}(z) \\ \Psi_{CM}(z+iL_2) &= e^{i\theta_2}(-1)^{N-1}\exp[-i\pi(\frac{2z}{L_1}+\tau)]\Psi_{CM}(z). \end{aligned} \tag{9.7.11}$$

These conditions imply that both $f(z)$ and $\Psi_{CM}(z)$ are entire (doubly) periodic functions with just one zero in the principal region. The solution is again an odd elliptic theta function

$$f(z_j - z_k) = \vartheta_1(\frac{z_j - z_k}{L_1}|\tau). \tag{9.7.12}$$

The wave function $\Psi_{CM}(z)$ can also be written in terms of a theta function

$$\Psi_{CM}(z) = e^{ikz}\vartheta_1(\frac{z - z_0}{L_1}|\tau). \tag{9.7.13}$$

This solution has three parameters (k and z_0, the coordinates of the zero of Ψ_{CM}) which are determined by the set of consistency conditions

$$\begin{aligned} e^{ikL_1} &= (-1)^N e^{i\theta_1} \\ e^{i2\pi\frac{z_0}{L_1}} &= (-1)^N e^{i\theta_2 + kL_2} \end{aligned} \tag{9.7.14}$$

which has the unique solution $k = \frac{\pi N}{L_1} + \frac{\theta_1}{L_1}$ and $z_0 = L_1(\frac{\theta_2}{2\pi} + \frac{N}{2}) - iL_2(\frac{N}{2} + \frac{\theta_1}{2\pi})$. Thus the wave-function for one filled Landau level is unique. Notice that, in contrast, the single particle states have an N-fold degeneracy. We will encounter a similar phenomenon when we consider the case of fractional filling.

One important feature of the wave function Ψ_N is the fact that the twist angles θ_1 and θ_2 only affect the dynamics of the CM through $\Psi_{CM}(z)$. The wave function $\Psi_{CM}(z)$ can be viewed as the wave function for a single particle located at z with charge $-Ne$ moving on a torus in the presence of an uniform external magnetic field with $N_s = W$ units of flux. Thus, the center of mass carries the full current. The Niu-Thouless-Wu formula can now be used to yield the result

$$\begin{aligned} < \sigma_{xy} > &= \frac{e^2}{i\hbar}\oint d\theta_j < \Psi_N|\frac{\partial}{\partial\theta_j}|\Psi_N > \\ &\equiv \frac{e^2}{i\hbar}\oint d\theta_j \int_0^{L_1} dx_1 \int_0^{L_1} dx_2 |\Psi_N|^2 \frac{\partial}{\partial\theta_j} \ln \Psi_{CM}(z, \vec{\theta}). \end{aligned} \tag{9.7.15}$$

The average Hall conductance is thus determined by *average* change of the phase of the wave function for the center of mass on a closed loop of the torus of boundary conditions. Since $\Psi_{CM}(z, \vec{\theta})$ is an entire function with exactly one zero in the principal region of the elliptic theta function, the theory of functions of complex variables tells us that the integral has the value

$$\oint d\theta_j \frac{\partial}{\partial\theta_j} \ln \Psi_{CM}(z, \vec{\theta}) = 2\pi i. \tag{9.7.16}$$

Thus we get

$$< \sigma_{xy} > = \frac{e^2}{h} \cdot 1 \tag{9.7.17}$$

which is indeed the answer we expected to get.

Let us finallly remark that the wave function Ψ_N is unique because we have an exactly filled Landau level. In contrast, the single-particle states are

N_s-fold degenerate. In the more general situation of fractional filling, where $\frac{N_s}{N_e}$ is not an integer, the wave function has more than one component [Haldane 85a].

9.8 The Quantized Hall Conductance of Filled Hofstadter Bands

We now turn to the far less trivial problem of computing the value of σ_{xy} for the problem of charged particles moving on a square lattice in the presence of an uniform commensurate magnetic field, the Hofstadter problem. In section (9.2) we presented a description of its single-particle states. Let us recall that, if the flux-per-plaquette is $\frac{p}{q}\phi_0$, there are q single-particle Landau bands each with $\frac{L_1 L_2}{q}$ degenerate states. In principle, if we solve the Schrodinger Equation, we can construct all the wave functions and, from there, we can compute anything we wish. These equations are very complicated and yield only to numerical solutions. However, the computation of σ_{xy} is considerably simplified by the fact that, here too, it is related to a topological invariant. Thus, we can calculate σ_{xy} within some approximate scheme and still get the exact answer.

Let us first derive an expression for σ_{xy} for a lattice system with periodic boundary conditions. Unlike the continuum problem case of the last section, the lattice problem is considerably simpler since the main effect of the magnetic field is to generate a sublattice structure. Indeed, in section (9.2) we saw that the requirement that there should be an integer number of flux quanta piercing the lattice means that either L_1 or L_2 have to be integer multiples of q. Since the magnetic unit cell has q plaquettes, there are $\frac{L_1 L_2}{q}$ magnetic unit cells. We have q sublattices and a Schrodinger Equation satisfied by the q sublattices. Hence, unlike the continuum case, we can apply periodic boundary conditions directly. The reason is that for this lattice problem what matters is not the vector potential $A_j(\vec{r})$ on the link $(\vec{r}, \vec{r}+\hat{e}_j)$ but on the phase $e^{iA_j(\vec{r})}$ which is invariant under the shifts $A_j(\vec{r}) \rightarrow A_j(\vec{r}) + 2\pi l_j(\vec{r})$, where $\{l_j(\vec{r})\}$ is a set of arbitrary integers. Furthermore, in section (9.2) we saw that even though the discrete magnetic translations do not commute with each other, there is a subset of discrete magnetic translations (i.e. those generated by $\hat{T}_1$ and $\hat{T}_2^q$) which commute among themselves and with the Hamiltonian. This subset, which defines the magnetic Brillouin zone, consists of the set of translations by integer numbers of magnetic unit cells. Thus, in units of the magnetic unit cell, the Hamiltonian is translationally invariant. It is then perfectly consistent to impose conventional periodic boundary conditions since the wave functions in *real space* are globally defined. However, they are not globally defined on the *momentum space torus* $(-\pi \leq k_1 < \pi, -\frac{\pi}{q} \leq k_2 < \frac{\pi}{q})$.

Let us derive a version of the Niu-Thouless-Wu formula for the case of a tight-binding system. This formula was derived by Kohmoto [Kohmoto 85], [Kohmoto 87].

In the case of a tight-binding system the current operator $\hat{J}_k(\vec{r})$ flowing on the link $(\vec{r}, \vec{r}+\hat{e}_j)$, can be obtained by differentation of the Hamiltonian

$$\hat{J}_k(\vec{r}) = \frac{\delta H}{\delta A_k(\vec{r})} \tag{9.8.1}$$

where H is an arbitrary (many-body) tight-binding Hamiltonian. We will assume here that the external vector potential $A_k(\vec{r})$ only enters in the kinetic energy term of H, which has the form $H_{\rm kin}$

$$H_{\rm kin} = \int_{\vec{k},\vec{k}'} C^\dagger(\vec{k})\, h_{\rm kin}(\vec{k},\vec{k}')\, C(\vec{k}') \tag{9.8.2}$$

where $h_{\rm kin}(\vec{k},\vec{k}')$ is the (hermitian) one-particle non-interacting Hamiltonian. In the particular case of a system coupled to an external electric field $\vec{E}$, the vector potential $\vec{A}$ gets shifted by $\vec{E}t$. It is easy to show that, when $\vec{E}$ is not zero, the kinetic part of the one-particle Hamiltonian $h_{\rm kin}$ takes the form

$$h_{\rm kin}(\vec{k},\vec{k}';\vec{E}) \equiv h_{\rm kin}(\vec{k}+\frac{e}{\hbar c}\vec{E}t, \vec{k}'+\frac{e}{\hbar c}\vec{E}t). \tag{9.8.3}$$

Thus, the external uniform electric field $\vec{E}$ (or a twist $\vec{\theta} \equiv \frac{et}{\hbar c}\vec{E}$) is equivalent to a shift of the momentum of each particle by $\frac{et}{\hbar c}\vec{E}$.
The Kubo formula can be written in the following simple form $(j,k = 1,2)$

$$(\sigma_{xy})_\alpha = -i\hbar L_1 L_2\, \epsilon_{kl}\, \frac{\delta}{\delta A_j} < \alpha | \frac{\delta}{\delta A_l} | \alpha > . \tag{9.8.4}$$

For the case of a non-interacting system this expression reduces to a sum over all the occupied one-particle states $\{|n>\}$ (i.e $\epsilon_n < E_p$)

$$(\sigma_{xy})_\alpha = \frac{e^2}{\hbar}\epsilon_{jl} \sum_{\{n\}} (\frac{\partial}{\partial k_j} < n|)(\frac{\partial}{\partial k_l}|n>). \tag{9.8.5}$$

The one-particle states $\{|n>\}$ are labelled by a band index r $(1 \le r \le q-1)$ and by a momentum label $\vec{k}$, where $\vec{k}$ in the magnetic Brillouin zone, $\Psi_r(\vec{k})$ are the eigenstates of the Schroedinger equation which satisfy the boundary condition $\Psi_{r+q}(\vec{k}) = \Psi_r(\vec{k})$. Let λ be a small parameter $(\lambda \to 1)$ which we will use to define (formally) a perturbation theory. This parameter enters in the Schroedinger equation in the form

$$-\lambda t\, [e^{ik_1}\Psi_{r+1}(k_1,k_2) + e^{-ik_1}\Psi_{r-1}(k_1,k_2)] + \\ -2t\cos(k_2 + 2\pi\frac{p}{q}r)\Psi_r(k_1,k_2) = E(k_1,k_2)\Psi_r(k_1,k_2). \tag{9.8.6}$$

This equation has a set of q linearly independent solutions $\{\Psi_r^{(j)}\}(\vec{k})(j = 1,\ldots,q)$. Each solution $\Psi_r^{(j)}(\vec{k})$ has an eigenvalue $E_j(\vec{k})$. These are the Landau-Hofstadter bands.

Let us now consider the case in which the number of particles N is such that there are an integer number r of exactly filled Landau-Hofstadter bands.

This requirement defines the state $|\alpha>$. The Hall conductance $(\sigma_{xy})_\alpha$ is then a sum of contributions, one from each filled band, of the form

$$(\sigma_{xy})_\alpha = \frac{e^2}{i\hbar}\sum_{n=1}^{r}\int_{-\pi}^{\pi}\frac{dk_1}{2\pi}\int_{-\frac{\pi}{q}}^{\frac{\pi}{q}}\frac{dk_2}{2\pi}\sum_{p=1}^{q}\epsilon_{jl}\partial_{k_j}\Psi_p^{(n)*}(\vec{k})\,\partial_{k_l}\Psi_p^{(n)}(\vec{k}). \qquad (9.8.7)$$

We can define a vector field $\mathcal{A}_j^{(n)}(\vec{k})$, for $\vec{k}$ on the magnetic Brillouin zone, to be

$$\mathcal{A}_j^{(n)}(\vec{k}) = \sum_{p=1}^{q}\Psi_p^{(n)*}(\vec{k})(-i)\partial_{k_j}\Psi_p^{(n)}(\vec{k}). \qquad (9.8.8)$$

The Hall conductance is essentially the flux of $\mathcal{A}_j^{(n)}$ through the magnetic Brillouin zone

$$(\sigma_{xy})_\alpha = \frac{e}{\hbar}\sum_{n=1}^{r}\int_{-\pi}^{\pi}\frac{dk_1}{2\pi}\int_{-\frac{\pi}{q}}^{\frac{\pi}{q}}\frac{dk_2}{2\pi}\epsilon_{jl}\,\partial_{k_j}\mathcal{A}_l^{(n)}(\vec{k}). \qquad (9.8.9)$$

Once again, $(\sigma_{xy})_\alpha$ is identified with a Chern number which counts the winding number of the phase of the wave functions as $\vec{k}$ traces the boundary of the magnetic Brillouin zone. Let us denote with I_n the Chern number for the n-th band. I_n is given by

$$I_n = \int_{-\pi}^{\pi}\frac{dk_1}{2\pi}\int_{-\frac{\pi}{q}}^{\frac{\pi}{q}}\frac{dk_2}{2\pi}\epsilon_{jl}\sum_{p=1}^{q}\partial_{k_j}\Psi_p^{(n)*}(\vec{k})\,\partial_{k_l}\Psi_p^{(n)}(\vec{k}). \qquad (9.8.10)$$

Since the numbers I_n are topological invariants, we can compute their exact value by considering a *smooth* deformation of the Schrodinger equation. For instance, we can compute the integers I_n in the limit $\lambda \to 0$ (or rather a perturbative expansion in powers of λ). If, as λ is varied from $\lambda = 0$ to $\lambda = 1$, there are no band crossing, the integers I_n will not change.

Let us now discuss the qualitative features of a (degenerate) perturbation theory in λ. At $\lambda = 0$ the eigenstates $\Psi_p^{(n)}(\vec{k})$ are $(n = 1, \ldots, q)$

$$\Psi_p^{(n)}(\vec{k}) = \delta_{pn} \qquad (9.8.11)$$

with eigenvalues $E_n^{(0)}(\vec{k})$

$$E_n^{(0)}(\vec{k}) = -2t\cos(k_2 + 2\pi\frac{p}{q}n). \qquad (9.8.12)$$

The spectrum then has q generally non-degenerate bands with dispersion laws $E_n^{(0)}(\vec{k})$. On the magnetic Brillouin zone $(-\pi \le k_1 < \pi, -\frac{\pi}{q} \le k_2 < \frac{\pi}{q})$, these bands cross at $\vec{k} = (k, 0)$ and $\vec{k} = (k_1, \frac{\pi}{q})$. For example, the lowest band $(n = 1)$ crosses the next $(n = 2)$ band at $k_2 = \frac{\pi}{q}$. The second band crosses the third one $(n = 3)$ at $k_2 = 0$, etc. In general, the n-th band (for n *even*) crosses with the $n-1$-st band at $k_2 = \frac{\pi}{q}$ (the bottom of the n-th band) and with the $n+1$-st band at $k_2 = 0$ (the top of the n-th band). Conversely, for n *odd*, the top of the n-th band is at $k_2 = 0$ (where it crosses with the $n+1$-st band) while the bottom is at $k_2 = \frac{\pi}{q}$ (where it crosses with the $n-1$-st band). The

integer n labels the bands as well as the gaps. The top band ($n = q$) has only one crossing (with the band $n = q - 1$ at $k_2 = 0$ (q even) or $k_2 = \frac{\pi}{q}$ (q odd).

The integers I_n are determined by the changes of the phases of the wave function as $\vec{k}$ passes through the degeneracy points. We can determine these phases by using Brillouin-Wigner pertubation theory (see section 2.3). The n-th band (for p and q fixed) crosses with the m-th band if $m = n - l_n$, where the integer l ($|l_n| \leq \frac{q}{2}$) is the solution of the Diophantine equation

$$n = qs_n + pl_n \tag{9.8.13}$$

a result first derived by Thouless and collaborators [Thouless 82].

The Schrodinger equation only mixes $\Psi^{(n)}$ with $\Psi^{(n\pm 1)}$. Thus, it takes l_n orders on perturbation theory to mix $\Psi^{(n)}$ and $\Psi^{(n-l)}$. For $\vec{k}$ close to the degeneracy points, the eigenstates will have almost all of their weight in $\Psi^{(n)}$ and $\Psi^{(n-l)}$. Thus, we get an effective Schroedinger equation of the form

$$\begin{aligned} \epsilon_n \Psi_n + V_{n,n-l}\Psi_{n-l} &= E\Psi_n \\ V_{n-l,n}\Psi_n + \epsilon_{n-l}\Psi_{n-l} &= E\Psi_{n-l}. \end{aligned} \tag{9.8.14}$$

The matrix element $V_{n,n-l}$ is (approximately) equal to

$$V_{n,n-l} = V^*_{n-l,n} \simeq (-\lambda t e^{-ik_1})\Pi_{r=n-l+1}^{n-1}[\frac{-\lambda t e^{-ik_1}}{\frac{1}{2}(\epsilon_n + \epsilon_{n-l}) - \epsilon_r}] \tag{9.8.15}$$

where $\epsilon_n(\vec{k}) = -2t\cos(k_2 + 2\pi\frac{p}{q}n)$. The eigenvalues of Eq.(9.8.14) are

$$E^{\pm}(\vec{k}) = \frac{1}{2}(\epsilon_n + \epsilon_{n-l}) \pm \sqrt{\frac{(\epsilon_n - \epsilon_{n-l})^2}{2} + |V_{n,n-l}|^2}. \tag{9.8.16}$$

The eigenstates have amplitudes $(\Psi_n^{(\pm)}, \Psi^{\pm}_{n-l_n})$ which are given by

$$\Psi_n^{(\pm)} = |\Psi_n^{(\pm)}|\exp(i\theta_n^{(\pm)}) \tag{9.8.17}$$

with a similar expression for $\Psi^{(\pm)}_{n-l_n}$. The amplitudes are

$$\begin{aligned} |\Psi_n^{(\pm)}| &= \frac{|V_{n,n_l}|}{\sqrt{|E^{(\pm)} - \epsilon_n|^2 + |V_{n,n-l_n}|^2}} \\ |\Psi^{(\pm)}_{n-l_n}| &= \frac{|E^{(\pm)} - \epsilon_n|}{\sqrt{|E^{(\pm)} - \epsilon_n|^2 + |V_{n,n-l_n}|^2}} \end{aligned} \tag{9.8.18}$$

with the phases

$$\begin{aligned} \theta_n^{(+)} - \theta^{(+)}_{n-l_n} &= \arg(V_{n,n-l_n}) + \pi = -k_1 l_n - (l_n - 1)\pi \\ \theta_n^{(-)} - \theta^{(-)}_{n-l_n} &= \arg(V_{n,n-l_n}) = -k_1 l_n - l_n \pi. \end{aligned} \tag{9.8.19}$$

Let us consider the n-th band with n even. The result is the same for n odd. At $k_2 = \frac{\pi}{q}$ it crosses with the $n + 1$-th band. At this degeneracy we have to choose the solution $E^{(-)}$ for the top of n-th band. Conversely, at $k_2 = 0$, the n-th band crosses with the $n - 1$-th band. Thus, we have to choose the solution $E^{(+)}$ for the bottom of the n-th band. Let us compute the circulation

of the vector field $\mathcal{A}_j^{(n)}(\vec{k})$ for the n-th band for $\vec{k} = (k_1, k_2)$ along the contour γ $\{(0,0) \rightarrow (\pi,0) \rightarrow (\pi, \frac{\pi}{q}) \rightarrow (0,0)\}$. On the first and third segments of the contour, k_2 is constant while k_2 changes from zero to π and from π to zero respectively. The component $\mathcal{A}_1^{(n)}$ is then equal to

$$\begin{aligned} \mathcal{A}_1^{(n)}|_{k_2=\frac{\pi}{q}} &= \frac{\partial}{\partial k_1} \arg V_{n,n-l_n}|_{k_2=\frac{\pi}{q}} = -l_n \\ \mathcal{A}_1^{(n)}|_{k_2=0} &= \frac{\partial}{\partial k_1} \arg(V_{n-1,n-1-l_{n-1}})|_{k_2=0} = -l_{n-1}. \end{aligned} \tag{9.8.20}$$

For the second and fourth segments we need to compute $\mathcal{A}_2^{(n)}$. But the phases have no essential dependence on k_2. Thus, we get

$$\mathcal{A}_2^{(n)}|_{k_1=0,\pi} = 0. \tag{9.8.21}$$

The results summarized by equations Eq.(9.8.19) and Eq.(9.8.20) shows that the circulation of $\mathcal{A}_j^{(n)}$ on γ is

$$I_n = \frac{1}{2\pi} \oint_\gamma \mathcal{A}_j^{(n)} dk_j = \int_0^\pi \frac{dk_1}{2\pi} [\mathcal{A}_1^{(n)}(k_1, 0) - \mathcal{A}_1^{(n)}(k_1, \frac{\pi}{2})]. \tag{9.8.22}$$

Thus, we get

$$I_n = \frac{1}{2\pi}(l_n - l_{n-1}). \tag{9.8.23}$$

Hence, the contribution from the n-th band to the conductance is

$$(\sigma_{xy})^{(n)} = \frac{e^2}{h}(l_n - l_{n-1}). \tag{9.8.24}$$

For a problem with r filled bands we have

$$(\sigma_{xy}) = \frac{e^2}{h} \sum_{n=1}^{r} (l_n - l_{n-1}) = \frac{e^2}{h}(l_r - l_0) \equiv \frac{e^2}{h} l_r \tag{9.8.25}$$

where we have used the definition $l_0 = 0$.

This result, originally derived by Thouless, Kohmoto, den Nijs and Nightingale [Thouless 82], shows that σ_{xy} is determined by the topological invariant I_n which characterizes the Landau-Hofstadter bands. This integer is the solution of the Diophantine equation. The integers l_n may be positive or negative and are restricted to be in the range $|l| \leq \frac{q}{2}$. Thus, in contrast to the continuum result, the quantized Hall conductance of a filled Landau-Hofstadter band may be positive or negative. This surprising result is a Bragg scattering effect due to the magnetic unit cells. Let us consider an example with $p = 11$ and $q = 7$. There are seven bands. Let us use the notation (s_n, l_n) for the two integers which solve the Diophantine equation. The solutions are $(-3,2), (-6,4), (2,-1), (-1,1), (7,-4), (4,-2)$ and $(1,0)$ for n ranging from $n = 1$ up to $n = 7$. Notice that the bands with $n = 3, 5$ and 6 have $l = -1, -4$ and -2 respectively and carry *negative* Hall conductance.

The Diophantine equation has a unique solution for q odd. For q even, the band with index $n = \frac{q}{2}$ has two possible solutions $(\frac{1-p}{2}, \frac{q}{2})$ and $(\frac{1+p}{2}, -\frac{q}{2})$. What happens here is that for q even and $n = \frac{q}{2}$ the Landau-Hofstadter bands

have a degeneracy, which we already dicussed in section (9.2). Depending on how this degeneracy is removed, the conductance is $+\frac{q}{2}$, $-\frac{q}{2}$ or even zero. This observation is important to the physics of flux phases. Let us finally remark that the $\sum_{n=1}^{q} s_n$ and $\sum_{n=1}^{q} l_n$ obey the sum rules

$$\sum_{n=1}^{q} s_n = \frac{q+1}{2} \qquad \sum_{n=1}^{q} l_n = 0 \tag{9.8.26}$$

for q odd, and

$$\sum_{n=1}^{q} s_n = \frac{q+1 \mp p}{2} \qquad \sum_{n=1}^{q} l_n = \pm\frac{q}{2} \tag{9.8.27}$$

for q even. The ambiguity in the sum rule is due precisely to the double solution at $n = \frac{q}{2}$ (q even).

References

[Abrahams 79] E. Abrahams, P.W. Anderson, D.C. Licciardello and T.V. Ramakrishnan. Phys. Rev. Lett. 42, 673 (1979).

[Anderson 58] P.W. Anderson. Phys. Rev. 112, 1900 (1958).

[Berry 84] M.V. Berry. Proc. Roy. Soc. London 392, 45 (1984).

[Ederlyi 53] Ederlyi et al. *Higher Transcendental Functions*, Vol. 2, page 355, Mc Graw Hill, New York (1953).

[Eguchi 80] T. Eguchi, P.B. Gilkey and A.J. Hanson. Phys. Repts. C66, 213 (1980).

[Fetter 71] A. Fetter and J.D. Walecka. *Quantum Theory of Many Particle Systems*, Mc Graw-Hill, New York (1971).

[Haldane 83] F.D.M. Haldane. Phys. Rev. Lett. 51, 605 (1983).

[Haldane 85a] F.D.M.Haldane and E.Rezayi. Phys. Rev. B31, 2529 (1985).

[Haldane 85b] F.D.M.Haldane. Phys. Rev. Lett. 55, 2095 (1985).

[Halperin 82] B.I Halperin. Phys. Rev. B25, 2185 (1982).

[Hofstadter 76] S. Hofstadter . Phys. Rev. 14, 2239 (1976).

[Kadanoff 61] L.P.Kadanoff and P.C. Martin. Phys. Rev. 124, 677 (1961).

[Klitzing 80] K. von Klitzing, G. Dorda and M. Pepper. Phys. Rev. Lett. 45, 494 (1980).

[Kohmoto 83] M. Kohmoto. Phys. Rev. Lett. 51, 1198 (1983).

[Kohmoto 85] M.Kohmoto. Ann. Phys. (New York) 160, 343 (1985).

[Kohmoto 87] M.Kohmoto. Phys. Rev. B35, 6017 (1987).

[Laughlin 83] R. Laughlin. Phys. Rev. Lett. 50, 1395 (1983).

[Levine 83] H. Levine, S. Libby and A. Pruisken. Phys. Rev. Lett. 51, 1915 (1983).

[Mahan 90] G. Mahan. *Many-Particle Physics* Second Edition, Plenum Press (1990).

[Nash 83] C. Nash and S. Sen. *Topology and Geometry for Physicists*, Academic Press, New York (1983).

[Nielsen 81] H. Nielsen and M. Ninomiya. Nucl. Phys. B185, 20 (1981).

[Niu 85] Q. Niu, D. Thouless and Y.S. Wu. Phys. Rev. B31, 3372 (1985).

[Pines 66] D. Pines and P. Nozieres. *The Theory of Quantum Fluids*, Benjamin (1966).

[Prange 90] R. Prange and S. Girvin. *The Quantum Hall Effect*, Springer-Verlag , New York (1990).

[Pruisken 84] A. Pruisken. Nucl. Phys. B235 (FS11), 277 (1984).

[Simon 83] B. Simon. Phys. Rev. Lett. 51, 2167 (1983).

[Thouless 82] D.J.Thouless, M.Kohmoto, M.P.Nightingale and M. den Nijs. Phys. Rev. Lett. 46, 405 (1982).

[Trugman 83] S. Trugman. Phys. Rev. B27, 7539 (1983).

[Tsui 82] D. Tsui, H. Stormer and A. Gossard. Phys. Rev. Lett 48, 1559 (1982).

[Wegner 79] F. Wegner. Z. Physik B36, 209 (1979).

[Wegner 83] F.Wegner. Z. Physik B51, 579 (1983).

[Wen 89] X.G. Wen and A. Zee. Nucl. Phys. B316, 641 (1989).

[Wen 90a] X.G. Wen. Int. Jour. Mod. Phys. 4, 239 (1990).

[Wen 90b] X.G. Wen. Phys. Rev. B40, 7387 (1990).

[Wen 90c] X.C. Wen. unpublished (1990).

[Wu 75] T.T. Wu and C.N. Yang. Phys. Rev. D12, 3845 (1975).

[Zak 64] J.Zak. Phys. Rev. 134, A1602 (1964).

CHAPTER 10

The Fractional Quantum Hall Effect

In this chapter we discuss the theory of the Fractional Quantum Hall Effect (FQHE). The explanation of this phenomenon has required the development of completely new ideas and methods. The concept of fractional statistics has become a crucial element of the theory.

The physical system involves fermions in strong correlation in the absence of time reversal symmetry. The treatment of systems with these features cannot be achieved succesfully within the conventional Hartree-Fock approach to correlations in Condensed Matter Physics. A new condensed state of matter, the Laughlin state, had to be discovered.

The Chern-Simons gauge theories, already discussed in chapter (7), have come to play an essential role in the theory of the FQHE, both as a way to describe the low energy phenomena (Landau-Ginzburg theory) and as a theoretical tool to explain the most important features of the problem.

We begin with a detailed description of the theory of the Laughlin wave function, which is followed by the field theory approaches to the FQHE.

10.1 The Fractional Quantum Hall Effect and the Laughlin Wave Function

In the past two sections we considered the problem of electrons moving on a two dimensional surface in the presence of a perpendicular magnetic field. We assumed that the electron density was such that an integer number of Landau levels (or bands) were completely filled. Because the system has an energy gap, the interactions do not play a very important role. In fact, a perturbative expansion (in powers of the coupling constant) around a state with one (or more) filled Landau level is likely to be well behaved. Since all processes involve exciting one (or several) electrons across the gap, the energy denominators are always different from zero. The ground state wave function for the interacting system is smoothly connected to the ground state wave function of the non-interacting system. The arguments of the past three sections indicate that the topological properties of the wave function will then

be the same for both the interacting and the non-interacting systems. In other words, naive perturbation theory is a good approximation in this case.

However, if one Landau level (or band) is partially filled, perturbation theory breaks down. Consider for simplicity the case of N particles in a magnetic field B with N_ϕ quanta of flux piercing the surface. The filling fraction $\nu = \frac{N}{N_\phi}$ is not an integer. We will consider the simpler (and popular) case of $\nu = \frac{1}{m}$, where m is an odd integer and, for each electron there are m quanta of flux. We further assume that the magnetic field is sufficiently large so that all the Zeeman energies are so large that the system is completely spin polarized. This is the case for most, but not all, the experimentally accesible systems. In this limit, the electrons behave as charged spinless fermions, each carrying an electric charge of $-e$.

In section 9.7 we saw that if just one Landau level is filled ($m = 1$), the ground state is non-degenerate and its wave function is a Slater determinant. For $m > 1$ only a fraction $\frac{1}{m}$ of the states in the first Landau level are occupied. The remaining $\frac{m-1}{m}$ states are empty. However, occupied and unoccupied states have exactly the same energy. The actual ground state has then to be determined through some sort of degenerate perturbation theory scheme. This procedure is bound to be very complex due to the macroscopic degeneracy of the Landau level . The resulting state is likely to have properties which are completely different from those of the unperturbed state.

The observed phenomenology of FQHE also suggests the need for a completely different state. A non-interacting fractionally filled state would still exhibit a fractional Hall conductance σ_{xy} since, at least for a Galilean invariant system, the conductance is determined by the amount of charge present. But such a state would not support the very precise plateaus which are seen in experiments, since additional particles can be added at almost no energy cost. The fact that the FQHE is seen only in the purest samples indicates that the effect is the result of electron correlations due to the Coulomb interactions. Moreover, the "quenching" of the single-particle kinetic energies by the magnetic field is telling us that the interactions play a dominant role. The FQHE is the result of the competition between degeneracy and interactions. In this sense, the FQHE is an example of strongly correlated electron systems.

The model which naturally describes the essential features of the physical system consists of an assembly of N electrons which occupy a fraction ν of the N_ϕ states of the lowest Landau level and interact with each other via Coulomb interactions. The ground state of this system must be such that it should not support any gapless excitations (otherwise the plateaus of σ_{xy} could not be so sharp) and it should be essentially insensitive to the presence of impurities. The wave function should be a complex function of the electron coordinates. This requirement follows from the fact that, if a magnetic field is present, time-reversal invariance is broken explicitly. Finally, Fermi statistics demands that the wave function $\Psi_N(\vec{r}_1, \ldots, \vec{r}_N)$ should be antisymmetric under the permutation of the positions of any pair of particles. Thus, Ψ_N vanishes as the positions of two particles approach each other.

We will now construct a wave function which satisfies all these requirements. Here we follow closely Laughlin's construction [Laughlin 83] [Laughlin 87]. Let us consider first the *low-density* limit $\nu \ll 1$ ($m \gg 1$). In this limit, the average separation between two electrons is much larger than the single-particle cyclotron radius $l_0 (a_0 \gg l_0)$. The electrons do not venture very far away and interactions further restrict their motion. The natural ground state in this limit is an electron crystal, known as a Wigner crystal. The electrons are able to minimize the total energy by arranging themselves on a triangular lattice. Actually the "guiding center coordinates" form a triangular lattice. A Hartree approximation yields a Wigner crystal state $\Psi_W(z_1, \ldots, z_N)$ of the form [Laughlin 87]

$$\Psi_W(z_1, \ldots, z_N) = \sum_P (-1)^P \phi_{j_1 l_1}(z_{P_1}) \ldots \phi_{j_N l_N}(z_{P_N}) \tag{10.1.1}$$

where the single particle states $\phi_{jl}(z)$ are

$$\phi_{jl}(z) \approx \exp[-\frac{1}{4l_0^2}|z_{jl}^{(0)}|^2 + \frac{1}{2l_0^2}\bar{z}z_{jl}^{(0)} - \frac{1}{4l_0^2}|z|^2] \tag{10.1.2}$$

and $z_{jl}^{(0)}$ are the (complex) coordinates of the (j,l) site of a triangular lattice,

$$z_{jl}^{(0)} = l_0\sqrt{\frac{4\pi m}{\sqrt{3}}}(j + (\frac{1}{2} + i\frac{\sqrt{3}}{2})l). \tag{10.1.3}$$

The Wigner crystal state Ψ_W does satisfy a number of the requirements listed above but not all of them. First, it *does* support elementary excitations with arbitrarily low excitation energy, the sound waves of the Wigner crystal. Since the state is a periodic array of charges, the charge density is not uniform and it is *strongly* affected by the presence of impurities which can, and do, pin the crystal at the impurity sites. Thus, this *pinned* state does not support any charge current unless the electric field is larger than some critical threshold value E_0 determined by the local pinning forces. This behavior is commonly seen in other charge crystals, such as incommesurate Charge Density waves. The best known examples are the quasi-one-dimensional system N_bSe_3 and the quasi-two-dimensional system N_bSe_2.

As the electron density increases (i.e. m grows smaller) the interparticle separation a_0 decreases. For a triangular lattice, a_0 is related to the filling fraction ν and the cyclotron length l_0 through the relation $\nu = \frac{4\pi}{\sqrt{3}}(\frac{l_0}{a_0})^2$. As ν approaches unity, the ratio $\frac{l_0}{a_0}$ is also a number of order one. Thus, as ν grows larger, there should be a phase transition from a Wigner crystal to a state which supports a Hall current. Indeed, as ν grows larger and the cyclotron length approaches the interparticle spacing, the quantum fluctuations should increase. The leading fluctuations should involve exchanges of a small number of nearby particles. In particular, there are processes which involve three-particle exchanges around an elementary triangle (or "ring"). Such processes spoil the long range positional order of the Wigner crystal. If these ring-exchanges are able to proliferate, the Wigner crystal melts and there is a

transition to a liquid state [Kivelson 86]. This phase transition is most likely to be first order but, depending on microscopic properties, it can also be second order. The resulting *liquid* state is expected to have uniform density. What is more important, and far less trivial to see from this point of view, is that it should have a gap to all excitations. The phonon of the Wigner crystal should dissapear from the physical spectrum. This phenomenon is strongly reminiscent of the Anderson-Higgs mechanism in a superconductor coupled to a *dynamical* gauge field: the phase mode of the superconductor gets "eaten" by the gauge field which, in the process, becomes massive. We will see below that FQHE *has a hidden, dynamically generated, gauge field* which is responsible for the most striking features of this phenomenon.

The liquid state should be regarded as a new condensed state of matter. Laughlin was the first to realize that this state is fundamentally different from other known condensed states, such as magnetism or superconductivity. Drawing on intuition he gained by studying systems with small numbers of particles, Laughlin proposed the following class of wave functions [Laughlin 83]

$$\Psi_N(\vec{r}_1,\ldots,\vec{r}_N) = \prod_{1\le j<k\le N} f(z_j - z_k)\ \exp\{-\sum_{j=1}^{N}\frac{|z_j|^2}{4l_0^2}\} \qquad (10.1.4)$$

where $f(z)$ is a suitably chosen *analytic* function of the complex coordinates $\{z_1,\ldots,z_N\}$, i.e. single-particle states only from the lowest Landau level. Fermi statistics demands that $f(z_j - z_k)$ be an *odd* function of $z_j - z_k$ which vanishes as $z_j \to z_k$. These requirements, together with the demand that Ψ_N should be an eigenstate of the total L_z orbital angular momentum, can be met by the simple choice of $f(z) \sim z^m$, where m is an odd integer. We thus arrive to the celebrated Laughlin wave-function Ψ_m,

$$\Psi_m(\vec{r}_1,\ldots,\vec{r}_N) = \prod_{1\le j<k\le N} (z_j - z_k)^m\ \exp\{-\sum_{j=1}^{N}\frac{|z_j|^2}{4l_0^2}\} \qquad (10.1.5)$$

This wave function is remarkable in several ways. Laughlin has computed the overlap between Ψ_m and the *exact* wave function of a small cluster of electrons (with $N \le 3$) and interaction pair potentials $u(r) = \frac{1}{r}$, $-\ln r$, $\exp(-\frac{r^2}{2})$. He found that in all cases, the overlap was better than 99 percent. For a special potential, namely $u(r) = u_0\nabla^2\delta(\vec{r})$, Trugman and Kivelson [Trugman 85]showed that $\Psi_m^{(N)}$ is the *exact* ground state wave function for all F.D.M. Haldane [Haldane 83] has constructed a class of Hamiltonians for which Laughlin-like states are the exact ground states (see below). Laughlin originally thought of Ψ_m as a *variational* wave function, with a Jastrow form which is commonly used to construct variational states for superfluid liquid Helium [E.Feenberg 69].

However, Ψ_m does not contain any variational parameters!. The ground state is determined by just finding the values of m which minimize the energy. But m is in fact determined by the total angular momentum! It is remarkable

that this guess works so well. It is an important problem for theorists to explain why is this such a good state. The Laughlin wave function also admits a number of generalizations which describe other filling fractions. These are the Hierarchical wave functions of Haldane [Haldane 83] and Halperin [Halperin 84]. We will consider mostly the $\frac{1}{m}$ Laughlin states.

We now follow Laughlin and determine the optimal value of m, as well as the nature of the correlations present in Ψ_m, by using the *plasma analogy*. Let $\rho(z_1, \ldots, z_N)$ be the *joint probability distribution* function

$$\rho(z_1, \ldots, z_N) = |\Psi_m(z_1, \ldots, z_N)|^2 \tag{10.1.6}$$

which can be thought of as a classical probability distribution for one-component plasma with N particles located at $\{z_1, \ldots, z_N\}$. Let $U(z_1, \ldots, z_N)$ be the classical potential energy and β an *effective inverse temperature* ($\beta = m$). The potential U is defined by demanding that ρ should have the Gibbs form

$$|\Psi_m(z_1, \ldots, z_N)|^2 = \exp\{-\beta U(z_1, \ldots, z_N)\} \tag{10.1.7}$$

The classical potential energy $U(z_1, \ldots, z_N)$ is given by

$$U(z_1, \ldots, z_N) = -2 \sum_{1 \le j < k \le N} \ln|z_j - z_k| + \frac{1}{2m} \sum_{j=1}^{N} |z_j|^2 \tag{10.1.8}$$

where we have used units of length such that $l_0 = 1$. The potential $U(\{z_j\})$ is equal to the total energy of a gas of classical particles each carrying charge $q = 1$ which interact with each other via the two-dimensional Coulomb pair potential, $V_c(z_j - z_k) = -\ln|z_j - z_k|$, and with a uniform neutralizing background charge of density $\rho_0 = \frac{1}{2\pi m}$. The interaction with the background charge is represented in $U(z)$ by the last term. This can be checked by noting that $\nabla^2 \frac{1}{2m}|z|^2 = \frac{2}{m}$, which agrees with the density being uniform and equal to $\frac{1}{2\pi m}$. This is the one-component classical plasma.

The plasma analogy is a very powerful tool for the investigation of the properties of the Laughlin wave function. All expectation values of local operators in the Laughlin state can be represented as an ensemble average in the plasma. There is a well developped body of knowledge on this subject. For instance, the average electron density at point z, $< \rho(z) >$, is

$$< \rho(z) > = \frac{\int d^2z_1 \ldots d^2z_N \, \rho(z) \, |\Psi_m(z_1, \ldots, z_N)|^2}{\int d^2z_1 \ldots d^2z_N |\Psi_m(z_1, \ldots, z_N)|^2} \tag{10.1.9}$$

where the local charge density $\rho(z)$ is equal to

$$\rho(z) = \sum_{j=1}^{N} \delta(z - z_j). \tag{10.1.10}$$

In the plasma analogy, we write the average charge density $< \rho(z) >$ in the form of a weighted average over the positions of the classical charges

$$< \rho(z) > = \frac{1}{Z_{\text{plasma}}} \int d^2z_1 \ldots d^2z_N \, \rho(z) \exp\{-\beta U(z_1, \ldots, z_N)\} \tag{10.1.11}$$

where Z_{plasma} is the partition function for a classical one-component plasma. The potential energy $U(z_1, \ldots, z_N)$ has a simple form in terms of the density variable $\rho(z)$

$$U[\rho(z)] = \int d^2z \int d^2z' \, (\rho(z) - \rho_0) V(z - z')(\rho(z') - \rho_0) \tag{10.1.12}$$

where $V(z - z')$ is the Coulomb pair-potential

$$V(z - z') = -\ln|z - z'| \tag{10.1.13}$$

and ρ_0 is the background charge.

If the density is low, the quantization of the charge of the individual electrons is very important. The dominant configuration in this limit is a Wigner crystal. But, as the density increases, the local density has larger fluctuations. As a result, the local average charge is not equal to the electron charge. In other words, at high densities, the local averarge density $\rho(z)$ becomes a continuous variable. In this limit, any additional local charge will be rapidly screened, and the local average density should become equal to the background charge. Conversely, at low-densities, screening is very poor and the local density can deviate significantly from the value of the background charge density. Thus, the electron liquid corresponds to the (high-density) plasma phase of the one-component Coulomb gas. The approximation in which the local density becomes a continuous variable is known as the Debye-Hückel theory. It is straightforward to verify that in this limit $< \rho(z) >= \rho_0$. This result is also seen to hold in Monte-Carlo simulations, at least for $m \leq 5$. More details on how are the plasma methods applied to the theory of the Laughlin state can be found in Laughlin's article in the book of Girvin [Laughlin 87] where he uses extensively the methods described by G. Stell [Stell 64].

Let us now discuss Haldane's construction of a class of Hamiltonians which have the Laughlin state as their exact ground state. Haldane begins by noticing that a system with a disk geometry with wave functions which vanish on the boundary (in the thermodynamic limit) is equivalent to a (large) sphere of radius R. A uniform magnetic field flows outwards from the sphere. The Laughlin states are then isotropic on the sphere. If the magnetic field is normal to the sphere, then it has to be that one of a magnetic monopole with magnetic charge equal to the total flux. Let $2s$ be the total flux (in units of the flux quantum $\frac{hc}{e}$). The single particle states for particles of charge e moving on the surface of the sphere have to be smooth and single valued. This demand forces the magnetic charge $2s$ of the monopole to be an integer. This is the famous Dirac quantization condition [Dirac 31]. We have already encountered this problem in chapter (5) when we described the path-integral formalism for spin.

The single particle Hamiltonian H now becomes

$$H = \frac{\omega_c}{2\hbar s}[\vec{r} \wedge (\vec{p} + \frac{e}{c}\vec{A})]^2 \tag{10.1.14}$$

where ω_c is the cyclotron frequency. Let $\vec{n}$ be a unit vector normal to the surface of the sphere (i.e. $\vec{n} = \frac{\vec{r}}{R}$). The magnetic field of the monopole is

$\vec{\nabla} \wedge \vec{A} = B\vec{n}$ where $B = (\hbar cs/eR^2)$. The vector $\vec{\Lambda} = \vec{n} \wedge (\vec{p} + \frac{e}{c}\vec{A})$ satisfies the algebra

$$[\Lambda_a, \Lambda_b] = i\hbar\epsilon_{abc}(\Lambda_c - \hbar s n_c) \tag{10.1.15}$$

and $\vec{\Lambda} \cdot \vec{n} = \vec{n} \cdot \vec{\Lambda} = 0$. Of course, this is the same problem with the gauge-covariant momentum which we discussed in section (9.1). Here too we should define another operator which should generate the magnetic translations. For the spherical geometry, this is just rotations. The generators of rotations are $\vec{L} = \vec{\Lambda} + \hbar s\vec{n}$ and satisfy the algebra

$$[L_a, L_b] = i\hbar\epsilon_{abc}L_c \tag{10.1.16}$$

$$[L_a, \vec{L}^2] = 0 \tag{10.1.17}$$

$$[L_a, n_b] = i\hbar\epsilon_{abc}n_c \tag{10.1.18}$$

$$[L_a, \Lambda_b] = i\hbar\epsilon_{abc}\Lambda_c. \tag{10.1.19}$$

The last condition implies that L_a commutes with $\vec{\Lambda}^2$ and hence with H. Thus L_a and H can be diagonalized simultaneously. The first two equations are telling us that the operators L_a satisfy the algebra of angular momentum. The eigenvalues of $\vec{L}^2$ are $\hbar^2 l(l+1)$, where $l = s+n$, n is a possitive integer (or zero) and $2s$ is an integer. This is just the Dirac quantization condition. Thus, $\vec{\Lambda}^2$ is equal to $\vec{L}^2 - \hbar^2 s^2$. We conclude that the single particle Hamiltonian has eigenstates $|m, l>$ such that

$$L_3|m, l> = \hbar m\ |m, l> \tag{10.1.20}$$

$$\vec{L}^2|m, l> = \hbar^2 l(l+1)\ |m, l> \tag{10.1.21}$$

$$H|m, l> = \hbar\omega_c \frac{(l(l+1) - s)}{2s}\ |m, l> \tag{10.1.22}$$

where $|m| \leq l$. Thus, each level is $2l+1$-fold degenerate. In terms of n and s the degeneracy is $2n+1+2s$. The lowest energy level, which corresponds to the lowest Landau level, has $n = 0$, $(l = 0)$ and it is $2s+1$-fold degenerate. If we represent the unit vector $\vec{n}$ in terms of a two-component spinor $\vec{u} = (u, v)$, as $\vec{n} = u^*_\alpha \vec{\tau}_{\alpha\beta} u_\beta$ ($\vec{\tau}_{\alpha\beta}$ are the Pauli matrices), then the Hilbert space of the lowest Landau level is spanned by the coherent states $\Psi^{(s)}_{(\alpha,\beta)}(u, v) = (\alpha^* u + \beta^* v)^{2s}$,with $|\alpha|^2 + |\beta|^2 = 1$, which are polynomials of degree $2s$.

In this notation the Laughlin states Ψ_m are

$$\Psi_m = \prod_{1 \leq j < k \leq N} (u_j v_k - u_k v_k)^m \tag{10.1.23}$$

with $S = \frac{1}{2}m(N-1)$ for states with N particles. It can be readily checked that this state is also an eigenstate of $\vec{L}^2 = (\sum_{j=1}^N \vec{L}_j)^2$ with zero eigenvalue since the three operators $L^+ = \hbar\sum_{j=1}^N u_j \frac{\partial}{\partial v_j}$, $L^- = \hbar\sum_{j=1}^N v_j \frac{\partial}{\partial u_j}$ and $L_3 = \hbar\sum_{j=1}^N (u_j \frac{\partial}{\partial u_j} - v_j \frac{\partial}{\partial v_j})$ anhilate Ψ_m. The state Ψ_m is thus rotationally and translationally invariant on the sphere.

Haldane further remarked that the states Ψ_m are *exact* eigenstates of a class of Hamiltonians constructed in the following manner. Let $P_J(L)$ be a

projection operator on states with $\vec{L}^2$ eigenvalue equal to $\hbar^2 J(J+1)$ and Π_s be the projection operator onto the Hilbert space of the lowest Landau level. Haldane proposed to write a projected Hamiltonian as

$$\Pi_s H \Pi_s = \sum_{1 \le j < k \le N} \{ \sum_{j > 2s-m} P_J(\vec{L}_j + \vec{L}_k) V_J \} \tag{10.1.24}$$

which, by construction, anihilates the state Ψ_m.

10.2 The Elementary Excitations of the Laughlin State

The Laughlin wave function is an accurate approximation for the ground state of the system only if the electron density ρ and the magnetic field B are such that the filling fraction ν is exactly equal to $\frac{1}{m}$. For densities and fields for which ν is close to (but not equal) $\frac{1}{m}$, it is no longer a good approximation. As we will see below, the states with $\nu \approx \frac{1}{m}$ are excitated states of the $\nu = \frac{1}{m}$ state. It is an essential feature of the Laughlin state that these states are not degenerate with the ground state even in the thermodynamic limit. The Laughlin state is found to have non-zero gap for *all* elementary excitations. The Laughlin state thus represents a uniform *incompressible* fluid.

Several excited states are possible. We may change the magnetic field locally without changing the total number of particles. This can be achieved by inserting an infinitesimally thin solenoid, carrying exactly one flux quantum, at one point of the sample (say, the origin $z = 0$). Or, we may add (or subtract) an electron without changing the external field. Furthermore, we may imagine local density fluctuations which do not change either the field or the total particle number. Among these excitations, there are density fluctuationes involving states only in the lowest Landau levels (phonons) or states in the first (or higher) excitated Landau levels (plasmons). For the sake of simplicity, in this section, I will only consider the state obtained by the addition of a solenoid. This state is a *Laughlin's quasihole*. We will briefly discuss the collective modes in a later section in which we will discuss the field theory picture of the Laughlin state.

The Laughlin state Ψ_m for $\nu = \frac{1}{m}$ is the product of a polynomial in the particle coordinates times an exponential factor. We can expand Ψ_m in a series of the form

$$\Psi_m(z_1, \ldots, z_N) = \sum_{\{k_1, \ldots, k_N\}} C_{k_1, \ldots, k_N} z_1^{mk_1} \ldots z_N^{mk_N} \exp\{-\sum_{j=1}^{N} \frac{|z_j|^2}{4l_0^2}\}. \tag{10.2.1}$$

The integers $\{k_1, \ldots, k_N\}$ run from 0 to N with the restriction

$$\sum_{j=1}^{N} k_j = \frac{1}{2} N(N-1). \tag{10.2.2}$$

The coefficients $C_{k_1},\ldots,k_N$ must be antisymmetric under the permutation of the indices.

Under a rigid rotation of the system as a whole by an angle θ about the origin, the coordinate z_j of each particle gets multiplied by a phase factor $e^{i\theta}$. Thus, Ψ_m transforms like

$$\Psi_m(e^{i\theta}z_1,\ldots,e^{i\theta}z_n) = e^{im\frac{N}{2}(N-1)}\Psi_m(z_1,\ldots,z_N) \tag{10.2.3}$$

which means that the total L_z angular momentum of Ψ_m, is equal to

$$M_m = \frac{1}{2}mN(N-1). \tag{10.2.4}$$

Let us now imagine that an infinitesimally thin solenoid which carries one unit of flux, is introduced adiabatically into the system and it pierces the disk at the origin, $z=0$. For flux ϕ, the single particle state changes from $z^n e^{-|z|^2/4l_0^2}$ to $z^{n+\alpha}e^{-|z|^2/4l_0^2}$, where $\alpha = \frac{\phi}{\phi_0}$, ϕ_0 being the flux quantum $\frac{hc}{e}$. Thus, if $\phi=\phi_0$, the n-th state in the first Landau level becomes the n+1-th state in the same Landau level.

The Laughlin state reacts very much in the same way, by shifting each $z_j^{mk_j}$ to $z_j^{mk_j+1}$ and a change in the coefficients. This process does not alter the exponential factor. If we ignore the change in the coefficients $C_{k_1},\ldots,k_N$ the shift can be seen to be the same as a multiplication of Ψ_m by a factor of the form $\Pi_{j=1}^N z_j$. This observation, which was also made first by Laughlin [Laughlin 83], motivates the choice of the following ansatz for the wave function $\Psi_m^{(+)}(z_o;\{z_j\})$ of the quasihole state, created by the adiabatic insertion of a solenoid

$$\Psi_m^{(+)}(z_o;z_1,\ldots,z_N) = \prod_{j=1}^{N}(z_j - z_0)\ \Psi_m(z_1,\ldots,z_N). \tag{10.2.5}$$

This state has angular momentum $M_m^{(+)} = M_m + N$. Furthermore, the amplitude $\Psi_m^{(+)}$ vanishes whenever the coordinate z_j of any of the N electrons approaches z_0. Thus, at z_0 the effect of the solenoid is to *deplete* the charge density. Hence, this state can be regarded as a *quasihole*. Naturally, since the total charge is the same as in the Laughlin state and since the charge density away from z_0 should be uniform, the only place where the charge missing from z_0 could have gone to is infinity. Or, rather, the physical boundary of the system. Thus, the solenoid causes the electron liquid to swell and to spill over the region it occupied before the solenoid was introduced.

The quasihole excitation energy ϵ_0 can be calculated using the plasma analogy. I will not describe this calculation here since it demands getting into a very technical plasma calculation which is better described elsewhere. The computation is given in considerable detail by Laughlin in his excellent review on the FQHE [Prange 87].

What will matter for the purposes of our discussion, is that the excitation energy is finite and has a finite limit as $N\to\infty$. Thus, the spectrum of quasiholes has an energy gap ϵ_0.

The *charge* q_0 of the quasihole can also be determined using the plasma analogy. It turns out that q_0 is a fraction of the electron charge, $q_0 = +\frac{e}{m}$. The argument goes as follows. The normalization of the quasihole wave function is

$$|\Psi_m^{(+)}(z_0; z_1, \ldots, z_N)|^2 = \Pi_{j=1}^N |z_j - z_0|^2 \, |\Psi_m(z_1, \ldots, z_N)|^2. \tag{10.2.6}$$

We can rewrite this expression in terms of a modified classical potential energy $U(z_0; z_1, \ldots, z_N)$ which has the simple form

$$U(z_0; z_1, \ldots, z_N) = U(z_1, \ldots, z_N) - \frac{2}{m} \sum_{j=1}^{N} \ln |z_j - z_0| \tag{10.2.7}$$

where $U(z_1, \ldots, z_N)$ is the classical potential energy for the one-component plasma.

The potential energy $U(z_0; z_1, \ldots, z_N)$ represents a classical one-component plasma interacting with a charge $-\frac{1}{m}$ which is held fixed at $z = z_0$. The most important properties of a plasma are its uniform density (in the absence of external probes) and the exact screening of all external probes. Since the external probe has charge $-\frac{1}{m}$, it repells the charges of the plasma within a distance ξ which is the plasma screening length. For $|z - z_0| < \xi$, the plasma density is suppressed by the repulsive force due to the probe. The amount of charge expelled from the vecinity of z_0 is equal to $-\frac{1}{m}$ so that there is a *missing charge* of $+\frac{1}{m}$ which neutralizes the charge of the probe. This behavior is indeed seen in detailed calculations, such as the ones reported by Laughlin [Laughlin 87].Thus, the quasihole behaves like a *positive* charge $q_0 = +\frac{e}{m}$. Away from the quasihole, the charge density is uniformand equal to its value in the absence of the quasihole. Where has the missing charge gone? To the boundary, of course!. Indeed, if the N-particle system occupies an area of radius R in the absence of the quasihole, its presence forces the liquid to expand from R to $R + \delta R$. The extra area occupied by the deformed liquid is $\pi(R + \delta R)^2 - \pi R^2$. Since R is large, the density is uniform and equal to $\frac{1}{2\pi m l_0^2}$. The radius R has to grow just enough to accomodate the extra charge $\frac{1}{m}$. Thus, we get the relation

$$[\pi(R + \delta R)^2 - \pi R] \frac{1}{2\pi m l_0^2} = \frac{1}{m} \tag{10.2.8}$$

where $\frac{R}{l_0}$ is given by

$$\frac{R}{l_0} = \sqrt{2mN}. \tag{10.2.9}$$

The total change δR of the radius is

$$\frac{\delta R}{2l_0} = \sqrt{mN + 1} - \sqrt{mN}. \tag{10.2.10}$$

By inspecting the expansions in single particle wave functions of both the Laughlin state and of the quasihole, we see that the highest single particle angular momentum which enters in the Laughlin state has angular momentum

equal to mN. For the quasihole the highest occupied state has angular momentum $mN+1$. Indeed, the change δR of the radius is exactly the amount necessary to include the $mN+1$-th state inside the region occupied by the liquid. On the other hand, had we added or extracted a whole *particle* from the liquid ($N \to \pm 1$), the change in the area would have been m-times the amount we just calculated. This can be seen quite easily in the expansion of the Laughlin state in single particle Landau states. We conclude that the quasihole has *fractional charge* $+\frac{e}{m}$.

The quasi-electron can be constructed in a similar manner. Instead of adiabatically introducing a solenoid which increases the local magnetic field, the solenoid now carries a flux which decreases the field by exactly one flux quantum. An argument along the lines of what we did above for the quasihole, shows that a solenoid carrying a negative flux decreases the angular momentum of each single particle state by one unit. Except for the state with angular momentum zero, which gets shifted to a state on the first excited Landau level, the addition of a solenoid with negative flux is equivalent to a downwards shift of the angular momentum of all single particle states by one unit. At the level of the Laughlin wave function, this is accomplished by a derivative operator which acts on the polynomial factor in the wave function [Laughlin 83],

$$\Psi_m^{(-)}(z_0;\{z_j\}) = \exp\{-\sum_{j=1}^{N}\frac{|z_j|^2}{4l_0^2}\}\prod_{j=1}^{N}(2\frac{\partial}{\partial z_j}-\frac{\bar{z}_0}{l_0^2})\prod_{1\le j<k\le N}(z_j-z_k)^m. \tag{10.2.11}$$

The same line of argument used above on the quasihole shows that the charge q_0 of the quasielectron is also fractional but negative, $q_0 = -\frac{e}{m}$.

The construction of the quasihole, as well as the quasielectron, has a strong resemblance with the construction of soliton states in one-dimensional systems in Quantum Field Theory [Jackiw 76] and in one-dimensional condensed matter systems [Su 81]. However, these two problems are qualitatively different. In fact, the Laughlin states either are non-degenerate, as in the case of a spherical geometry, or have a degeneracy of topological origin, as in the case of a torus. In contrast, the one-dimensional systems which have solitons, have ground states which spontaneously brake a (discrete) global symmetry. The degeneracy of their ground states is a consequence of this phenomenon. Nevertheless, the operator which introduces an extra solenoid has some of the characteristic features of a soliton operator. While the short distance details are unimportant, the topological properties of the extra vector potential (i.e. the line integral on a non-contractible loop) is the only essential property of the "solenoid" or quasihole operator. In fact, the addition of the solenoid changes the value of the circulation of the vector potential around the physical boundary of the system. In turn, this change determines the amount of charge "spilled over the edge." This extra charge becomes an excitation of the states at the edge of the system.

The quasiholes and quasielectrons cannot be made in isolation directly by just adding or subtracting electrons. As a matter of fact a *hole* (not a quasihole) requires to remove a full electron which carries integer charge.

Thus, electrons and holes are equivalent to bound states of m fractionally charged quasiparticles. For certain definite electron densities, the excess electrons which cannot be accommodated into a $\frac{1}{m}$ Laughlin state, can be placed into a generalized Laughlin state. The excess electrons can be regarded as bound states of quasiholes or quasielectrons which, if their number is right, can form a Laughlin state. But this is a Laughlin state for anyons not electrons. This mechanism is known as the Hierarchy scheme of Haldane and Halperin.

The construction of the quasihole also suggests a different interpretation of the Laughlin wave function as well as generalizations valid for other filling fractions. Let us write the Laughlin wave function Ψ_m in the following suggestive form due to J. Jain [Jain 89] [Jain 90].

$$\Psi_m(z_1,\ldots,z_N) = \prod_{1\le j<k\le N} (z_j - z_k)^{m-1}\Psi_1(z_1,\ldots,z_N). \qquad (10.2.12)$$

The factor $\Psi_1(z_1,\ldots,z_N)$ is just the wave function for N particles exactly filling up the lowest Landau level. Following the construction of the quasihole, the factor in front of Ψ_1 is interpreted as the result of having attached a solenoid to each particle. The flux carried for each solenoid is equal to $(m-1)$ flux quanta. Unlike the quasihole construction, the solenoids are physically attached to the particles which fill up the Landau level and move around with them. This factorization, which appears to be quite innocent, has the virtue (and the beauty) of bringing the Fractional and Integer Hall states together. It is also telling us that the Laughlin state can be viewed as the result of a dynamical generation of a local gauge field which generates the solenoids which partially screens the external magnetic flux. In fact, the amount of screening is sufficient to turn the fractional filling of a Landau level of the bare field into the complete filling of a Landau level of the unscreened part of the field. We will see later on this chapter, that this is the starting point of the field theoretic description of the FQHE.

In summary, the $\frac{1}{m}$ Laughlin states are seen to have quasihole and quasielectron excitations which have fractional charge $\pm\frac{e}{m}$ and fractional statistics$\pm\frac{\pi}{m}$. These quasiparticles are obtained by the *adiabatic* addition or removal of infinitesimally thin solenoids carrying one flux quantum. The adiabaticity of this process is essential to this construction, since it is necessary to make the fluid swell enough to include one additional Landau orbit and without promoting electrons to higher Landau levels or producing ripples in the fluid. All these bulk excitations have finite energy gaps. This is required by the incompresibility of the fluid, which guarantees the accuracy of the adiabatic process.

10.3 Physics at the Edge

In an incompressible quantum fluid, such as the Laughlin state, the fluctuations in the bulk induce fluctuations at the boundary. While the local fluctuations in the bulk are associated with local changes in the density, the fluctuations of the states at the boundary are associated with changes in the shape of the "droplet". These "edge waves" are the only gapless excitations of the system. It may seem surprising that an incompressible fluid may have gapless modes at the surface, although this is quite common in conventional fluids such as water!. In the FQHE the gaplessness arises from the fact that the geometric edge of the fluid coincides with the locus of points in which the Fermi energy crosses the external potential which confines the fluid. Thus, the boundary of the fluid behaves like a "Fermi surface" and, as we move from the edge and into the bulk, we get deeper and deeper in the Fermi sea of occupied states. Because of the presence of the magnetic field, the edge waves are *chiral* excitations which move at a velocity which is the drift velocity of the particles at that point. Thus, edge states move only in one direction, which is specified by the magnetic field. The importance of the edge states to the observability of the Quantum Hall effect was first emphasized by Halperin [Halperin 82]. The description of the chiral quantum dynamics of the edge states is due to X.G.Wen [unpublished ,1990] and M.Stone [unpublished, 1990].

Let us discuss the physics of the edge waves in the context of the simplest system: non interacting electrons filling up the lowest Landau level ($\nu = 1$). Strictly speaking we are discussing the behavior of the edge states in a system with an *Integer* Quantum Hall Effect (IQHE). However, at least within a mean (or average) field approximation , the *Fractional* Quantum Hall Effect can also be regarded as an IQHE of an equivalent system of fermions. We will discuss this point of view (originally due to Jain) at the end of this chapter where we discuss the Chern-Simons approach to the FQHE. In this section we follow the methods of M.Stone [M.Stone, unpublished, 1990].

Let us, once again, consider a set of N electrons which are filling up the lowest Landau level of a system with $N_\phi = N$ flux quanta piercing the surface. In the absence of any other forces, the system has uniform density $\langle\rho\rangle = B/2\pi$, if the units are such that $\hbar = c = e = 1$. But, if no external forces are present, a system with N electrons in an *infinite* plane cannot have a fixed density. Furthermore, since we are interested in the physics at the edges, we must assume that the N electrons are constrained to remain within some region of the sample by the action of some extermal force. In the experimental setting, even in the purest samples there are forces as we examine the system close to the edges. So, we sould assume that, in addition to experiencing the uniform magnetic field B, the electrons also feel an *electrostatic* potential $V(\vec{x})$ which keeps them inside the sample. We will consider the simple geometry of an infinite strip. The system has finite width L_1 along the axis x_1 and length L_2 along the axis x_2, with $L_2 \gg L_1$. We can also assume *periodic boundary conditions* along x_2. I will also assume that the potential V only varies along

x_1 and that its variation is so slow that we locally can always be approximated by the linear function of x_1, $V(x_1) \approx Ex_1$. In this geometry, it is natural to use the axial-Landau gauge $A_1 = 0, A_2 = Bx_1$.

Let us expand the second quantized electron field operators $\psi(\vec{x})$ as a sum over states of the lowest Landau level, namely

$$\psi(x_1, x_2) = \sqrt{\frac{B}{\pi L_2}} \sum_{n=-\infty}^{+\infty} a_n \, e^{ik_n x_2} \, e^{-\frac{B}{2}(x_1 - k_n/B)^2} \tag{10.3.1}$$

which satisfies the boundary conditions. The allowed momenta k_n are $k_n = 2\pi n/L_2$. The creation and destruction operators , $a_n^\dagger$ and a_n, obey the anti-commutation relations

$$\{a_n, a_m^\dagger\} = \delta_{nm} \tag{10.3.2}$$

In the presence of an external potential, the degeneracy of the Landau level is lifted. For the particular case of the linear potential, the wave functions are the same as the wave functions in the absence of the potential but the single particle $\epsilon(k)$ energies become

$$\epsilon(k) = \frac{E}{B}k \tag{10.3.3}$$

with a sign determined by the sign of B (for $V(x_1)$ fixed). This expresion is accurate for those states whose energy is close to the Fermi energy, which I have set to zero. Away from the boundaries, the potential is essentially constant and the Landau level effectively has a degeneracy. The origin of my coordinate system is at the point where the potential crosses the Fermi energy. Thus, far to the left of the crossing point, the density is constant and to the right of the crossing point there are no particles. It is clear from this picture that it takes a negligible amount of energy to add a particle to the system, but the particle is added to the surface, not to the bulk. The low energy excitations of the system are local changes of density at the surface, the edge waves. Notice that since the number of particles is fixed and since the next Landau level is separated from the ground state by a very large energy gap, a lower density at a point on the surface means that there should be an excess density at some other point of the same surface.

As usual, we are only interested in the excitations with low energy. Here, to be close to the Fermi energy means to be close to the surface. Let $j(x_2)$ be the operator which measures the amount of charge localized within some region of size Λ of the edge,

$$j(x_2) = \int_{-\infty}^{+\infty} dx_1 \, f_\Lambda(x_1) \psi^\dagger(x_1, x_2) \, \psi(x_1, x_2). \tag{10.3.4}$$

The cutoff function $f_\Lambda(x_1)$ must be chosen in such a way that it is vanishingly small in the region $|x_1| \gg \Lambda$ and the cutoff Λ must be larger than the typical amplitude fluctuation of the low energy states. We will choose the cutoff function to be a gaussian, $f_\Lambda(x_1) = \frac{1}{\sqrt{2\pi}\Lambda} \exp(-x_1^2/2\Lambda^2)$.

Since we are using periodic boundary conditions in x_2, it is convenient to consider the Fourier transform of the operator $j(x_2)$ *i.e.*,

$$j(x_2) = \sum_n e^{-ik_n x_2} j_n \tag{10.3.5}$$

Conversely, we can write

$$j_n = \sum_{m=-\infty}^{+\infty} a^\dagger_{m+n} a_m \, e^{-\frac{B}{4}k_n^2} \tag{10.3.6}$$

It is apparent that the gaussian factor $\exp(-Bk_n^2/4)$ is negligibly small away from the Fermi surface.

If we ignore the gaussian factor,the density operator j_n coincides with the density operator for a system of fermions in one space dimension which are only allowed to move in one direction. In this case the direction is specified by the sign of the magnetic field B. Indeed, the dispersion law $\epsilon(k) = v_F k$, with a Fermi velocity $v_F = E/B$, follows from the Hamiltonian

$$H = \int dx_2 \, \psi^\dagger_R(x_2) \, (-iv_F \partial_2) \psi(x_2) \tag{10.3.7}$$

which governs the dynamics of right-moving chiral fermions in one dimension. Notice that this is precisely the same Hamiltonian that we found in chapter (4) when we discussed bosonization.

The results of chapter (4) enable us to write down the commutation relations obeyed by the operators j_n. There we found that the commutator of the Fourier transformed density operators is different from zero due to the presence of a Schwinger term

$$[j_n, j_m] = -n\delta_{n+m,0} \tag{10.3.8}$$

Alternatively, in position space we can write

$$[j(x_2), j(x_2')] = -\frac{i}{2\pi}\partial_2 \delta(x_2 - x_2') \tag{10.3.9}$$

This algebra is known as the level one $U(1)$ Kac-Moody algebra. Wen has shown that the spectrum of the edge states is always determined by an appropriate Kac-Moody algebra. For instance, if the fermions were not fully polarized, spin would have to be included in the dynamics. In that case the relevant algebra is the (level one) $SU(2)$ Kac-Moody algebra.

Stone has also given the following explicit construction of the edge density waves. Let $|0\rangle$ denote the ground state, which has an undisturbed droplet. Let us define the family of coherent states $\{|\theta(x_2)\rangle\}$, where

$$|\theta(x_2)\rangle = e^{i\int dx_2 \, \theta(x_2) j(x_2)} \, |0\rangle \tag{10.3.10}$$

which represent *coherent excitations* of the edge states. Throughout it is assumed that the density operator has been *normal-ordered* relative to the undisturbed state, namely $j(x_2)|\,0\rangle \equiv 0$.

We now show that the states $|\theta(x_2)\rangle$ are eigenstates of the (normal ordered) density operator $j(x_2)$:

$$j(x_2)\,|\theta(x_2)\rangle = \frac{1}{2\pi}\partial_2\theta(x_2)|\theta(x_2)\rangle \tag{10.3.11}$$

This property can be derived by using the identity

$$e^{-i\int dx_2'\,\theta(x_2')j(x_2')}\,j(x_2)\,e^{+i\int dx_2'\,\theta(x_2')j(x_2')} = j(x_2) + \frac{1}{2\pi}\partial_2\theta(x_2) \tag{10.3.12}$$

These states represent local changes in the density. This can be seen by the following argument: the state $|\theta(x_2)\rangle$ has local excess of charge equal to $\frac{1}{2\pi}\partial_2\theta(x_2)$. From the linearity of the energy-momentum relation we know that an extra number of particles means that the local position of the Fermi level has gone from zero to $\partial_2\theta$, which is still much less than the Landau gap. Likewise, the momentum k has changed by the same amount. Since we also saw that, for these states, there is a precise relation between the energy of the state and its location on the axis x_1, we conclude that this state is in fact a local change of the *shape* of the droplet. Moreover, at least within the accuracy of the linear approximation for the dispersion relation, these states propagate without deformation since all the excitations propagate at the same speed v_F.

Throughout this discussion we have focused on the states close to the edge. But, as we have already warned the reader, the bulk cannot be decoupled from the edge. In fact, theories of chiral fermions, such as the one we are discussing here, are intrinsically sick. The reason is that, if the linear spectrum is taken literally this system would not be able to keep track of the conservation of charge at the moment it is coupled to a fluctuating vector potential. Indeed, in one dimension, all the components of the vector potential are longitudinal since there is no way to "enclose flux inside a line". But it is possible to do it if the line closes on itself forming a closed curve. This is precisely the case of interest to us. For example, in the gauge $A_0 = 0$, the only component we are left with is the component $A_\parallel$ tangent to the curve (the edge). By general arguments of gauge invariance we know that the Hamiltonian for the chiral fermions coupled to the gauge field is obtained by the minimal coupling procedure, which replaces the derivative ∂_2 by the covariant derivative $D_2 = \partial_2 - ieA_\parallel(x_2)$. Thus, the Hamiltonian picks up an extra term H_{gauge} of the form

$$H_{\text{gauge}} = \int dx_2\; eA_\parallel(x_2)\;\psi_R^\dagger(x_2)\;\psi(x_2) \tag{10.3.13}$$

This term shows that the local fluctuations of $A_\parallel(x_2)$ will cause the Fermi level to move up and down. Thus, charge has to "leak-in" or "leak-out" through the bottom of the Fermi sea. For a theory "without a bottom", such as in a relativistic field theory, this is a disaster. The chiral theories are then said to be sick and to break gauge invariance and to have a gauge anomaly. But in the problem that we are considering, the Fermi sea does have a bottom. It is determined by the Landau level, which acts like a reservoir of particles and it redistributes the particles from one point of the edge to another.

The configurations with a non-zero circulation of $A_{\|}$ have a very interesting meaning: the circulation of $A_{\|}$ on a closed curve such as the edge is just the amount of flux enclosed inside the curve. Thus, the uniform field causes the electrons on edge states to move around the system. A change in the circulation means that flux has been added or taken from the system. Thus, the addition of one quasihole should cause a jump in the circulation by exactly one flux quantum. The edge states see this extra flux as a change of the position of the Fermi level. This is then interpreted as the generation of a net charge at the edge. For a non-interacting problem, the net charge is equal to e. But for Laughlin state, it is equal to e/m. The extra charge accumulated at the edge is interpreted as a lack of charge conservation, as a gauge anomaly of the theory of the edge states.

10.4 The Statistics of the Quasiparticles within Laughlin's Theory

In this section we will discuss the statistics of the quasiholes within the first quantized picture of the FQHE. In the last section of this chapter we will return to this problem and derive the main results directly from the field theory . The statistics of the quasielectron can also be discussed along a very similar line of argument.

The quasihole wave function discussed in section (10.2), is given up to a normalization factor. For a single quasihole, the amplitude of this wave function is not very important. However, at the moment we wish to construct a wave function for two or more quasiholes, the normalization begins to play a rather subtle but importand role. During a process which involves dragging a single quasihole very slowly around a closed loop, the phase of the quasihole wave function becomes very important. Indeed, since the quasihole carries an electric charge of $-\frac{e}{m}$, we should expect an Aharonov-Bohm effect(AB) $\frac{1}{m}$ times smaller than the value for electrons. In fact, the AB effect is perhaps the "operationaly correct" way of measuring the charge of a quasiparticle.

The quasihole wave function is physically appealing but it has several drawbacks. Consider for example a naively constructed wave function for two quasiholes located at $z = u$ and $z = w$ respectively

$$\Psi^{(+)}(u, w, z_1, \ldots, z_N) = N(u, w) \prod_{j=1}^{N} (z_j - u)(z_j - w)\, \Psi_m(z_1, \ldots, z_N). \tag{10.4.1}$$

The factor $N(u, w)$ has a subtle origin. On the one hand, it can be regarded as the normalization constant for the state with two quasiholes. However, if that was indeed the case, N would have to be not only a function of u and w, but also of $\bar{u}$ and $\bar{w}$ and it would not be analytic. More importantly, this amplitude has to be determined from the requirement that it represents the *physical* process of adiabatic insertion of two thin solenoids. In section (10.2)

we saw that the form of the wave function for one quasihole was *suggested* by the observation that the *adiabatic* insertion of a solenoid carrying one flux quantum implied an increase of the angular momentum *relative to the location of the solenoid* of one unit per particle. We also argued that the quasihole carries charge e/m. Below in this section we will give a path integral argument to support this picture.

But, let us assume that we have already manufactured one quasihole, which is sitting at $z = u$. We now want to create another quasihole but, this time at $z = w$. The *adiabatic* addition of the extra solenoid must change the angular momentum of the particle also by one unit, but this time the angular momentum is measured *relative* to w, not to u. Furthermore, since the quasihole carries electric charge equal to e/m, as we drag one quasihole slowly around the other, we should pick up an extra AB phase factor. This phase factor should correspond to an AB effect for a charge equal to e/m [Kivelson 85].

We are going to determine the amplitude (or "normalization constant") $N(u, w)$ by demanding that the following conditions are met: (a) the wave function should be an analytic function of the coordinates of the electrons $\{z_1, \ldots, z_N\}$ and of the quasiholes u, w up to exponential factors; (b) the normalization of this wave function should be invariant under translations, i.e. a function of differences of the coordinates $\{z_1, \ldots, z_N, u, w\}$. The analyticity condition is just the requirement that the wave function should only have contributions from the lowest Landau level. This conditions, as well as the solution, were first proposed by Halperin [Halperin 84][Halperin 83].The normalization of the state is

$$|\Psi_m^{(+)}(u, w, z_1, \ldots, z_N)|^2 = \exp\{-\beta U_{\text{eff}}(u, w, z_1, \ldots, z_N)\} \tag{10.4.2}$$

where U_{eff} is given by

$$U_{\text{eff}}(u, w, z_1, \ldots, z_N) = U(z_1, \ldots, z_N) - \frac{2}{m}\sum_{j=1}^{N}(\ln|u - z_j| + \ln|w - z_j|) + \\ + \frac{2}{m^2}\ln|N(u, w)|. \tag{10.4.3}$$

The translation invariance and analyticity requirements are met by the choice

$$N(u, w) = N_0(u - w)^{1/m}\exp\{-\frac{|u|^2 + |w|^2}{4l_0^2 m}\}. \tag{10.4.4}$$

Thus, the Halperin wave function for two quasiholes is

$$\Psi_m^{(+)}(u, w; \{z_j\}) = N_0(u - w)^{\frac{1}{m}}\prod_{j=1}^{N}[(u - z_j)(w - z_j)] \\ \exp\{-\frac{1}{4l_0^2}\sum_{j=1}^{N}|z_j|^2 - \frac{1}{4ml_0^2}(|u|^2 + |w|^2)\}. \tag{10.4.5}$$

With this choice, the effective potential U_{eff} is

$$U_{\text{eff}}(u,w;\{z_j\}) =$$
$$-2\sum_{1\le j<k\le N}\ln|z_j-z_k| - \frac{2}{m}\sum_{j=1}^{N}(\ln|z_j-u|+\ln|z_j-w|)+$$
$$-\frac{2}{m^2}\ln|u-w| + \frac{1}{2ml_0^2}\sum_{j=1}^{N}|z_j|^2 + \frac{1}{2m^2l_0^2}(|u|^2+|w|^2). \tag{10.4.6}$$

In plasma language, this is the potential energy of a set of N classical particles (each carrying charge (-1)) at sites $\{z_1,\ldots,z_N\}$ interacting with two extra particles (each with charge $-\frac{1}{m}$) at u and w. All $N+2$ charges are coupled to a neutralizing background charge of density $\frac{1}{2\pi ml_0}$. The manifest translation invariance of U_{eff} takes care of the translation invariance requirement.

The wave function for two quasiholes is a multivalued function of the complex coordinates of the two quasiholes. As a result, if the quasiholes undergo a counterclockwise exchange process, defined as a counterclockwise rotation by π of one quasihole around the other followed by a translation which restores the relative position of the quasiholes, *the phase of this wave function changes by* $\frac{\pi}{m}$

$$\Psi_m^{(+)}(u,w;\{z_j\}) = e^{+\frac{i\pi}{m}}\Psi_m^{(+)}(w,u;\{z_j\}). \tag{10.4.7}$$

Thus, the quasiholes are *anyons* with statistical angle $\delta=\frac{\pi}{m}$ relative to bosons or $\delta=\frac{m-1}{m}\pi$ relative to fermions. This remarkable result suggests that the Fractional Quantum Hall Effect can be described in terms of a theory of either bosons or fermions coupled to a hidden (or dynamically generated) Chern-Simons gauge field. In the next section we will describe both a "Landau-Ginzburg" approach to the FQHE and a field theory which are based on this idea.

We conclude that the quasiholes of the Laughlin state carry fractional charge $+\frac{e}{m}$ and fractional statistics $\frac{\pi}{m}$. This is a very striking result. Arovas, Schrieffer and Wilczek [Arovas 84] have given an alternative derivation of both results using an argument based on the concept of Berry phases [Berry 84][Simon 83]. Rather than following that path, we will now construct a path-integral to represent the motion of the quasiholes. The key ingredient of our construction is the observation that the quasihole wave functions are *coherent states* [Kivelson 87]. Thus, we can adapt the formalism described in chapter (5) to construct the path integral for spin s particles, to treat the quantum dynamics of the quasiholes. The reader should keep in mind that the following arguments are heuristic at best. In the last section of this chapter I give a different derivation of the same result, based on the field theory approach to the FQHE. Fortunately, the results agree!

Let us begin with the wave function for a *single* quasihole. Let us define

the state $|z>$ as

$$|z>=\exp\{-\frac{1}{4ml_0^2}|z|^2\}\prod_{j=1}^{N}(z_j-z)|m> \tag{10.4.8}$$

where $|m>$ is the Laughlin state. The set of states $\{|z>\}$ is overcomplete [Laughlin 87]. The overlap between two states $|z>$ and $|w>$ is

$$<z|w>=\exp\{-\frac{1}{4ml_0^2}(|z|^2+|w|^2)\}<m|\prod_{j=1}^{N}[(\bar{z}_j-\bar{z})(z_j-w)]|m> . \tag{10.4.9}$$

Except for the exponential factor, $<z|w>$ is an analytic function of $\bar{z}$ and w separately. Thus, $<z|w>$ can be related to $<z|z>$ by analytic continuation [Laughlin 87]. The result is

$$<z|w>=\exp\{-\frac{1}{4ml_0^2}(|z|^2+|w|^2)+\frac{1}{2m}\bar{z}w\}<z|z> . \tag{10.4.10}$$

Indeed, the translation invariance of the 2D one-component plasma guarantee that the overlap $<z|z>$ is just a constant independent of z. Also, up to a normalization constant we can write the resolution of the identity

$$1=\mathcal{N}\int|z><z|\,d^2z. \tag{10.4.11}$$

We consider now a process in which we prepare the quasihole in a coherent state $|z_0>$ at time $t=t_0$. We now ask for the quantum mechanical amplitude $<z_0,t_0+T|z_0,t_0>$ for the quasihole to return to $|z_0>$ after a very long time T. By inserting the resolution of identity at $\mathcal{N}_\tau$ intermediate times $t_n=t_0+n\Delta t$ in the limit $\mathcal{N}_\tau\to\infty$ and $\Delta\to 0$ with $T=\mathcal{N}_\tau\Delta t$ fixed we can write

$$<z_0,T+t_0|z_0,t_0>=\mathcal{N}\int\prod_{n=1}^{\mathcal{N}_\tau}d^2z_n\prod_{n=1}^{\mathcal{N}_\tau}<z_n|z_{n+1}> \tag{10.4.12}$$

where $z_n=z(t_0+n\Delta t)$ and $z_{N_\tau}=z_0$.

In the limit $\Delta t\to 0$ we can approximate the overlaps by the expression

$$<z_{n+1}|z_n>\approx<z_n|z_n>\ \exp\{\frac{1}{4ml_0^2}(\bar{z}_n\frac{dz_n}{dt}-z_n\frac{d\bar{z}_n}{dt})\Delta t\}. \tag{10.4.13}$$

Thus, the path integral is

$$<z_0,T+t_0|z_0,t_0>=\mathcal{N}\int\mathcal{D}z\ \exp\{\frac{1}{2ml_0^2}\int_0^T dt\,\bar{z}\frac{dz}{dt}\}. \tag{10.4.14}$$

By expanding the exponent in its real and imaginary components we get the identity

$$\frac{1}{2ml_0^2}\int_0^T dt\,\bar{z}\,\frac{dz}{dt}=\frac{(e/m)}{\hbar c}\oint_\Gamma\vec{A}(\vec{x})\cdot d\vec{x} \tag{10.4.15}$$

where $\vec{A}$ is the vector potential for the field B in the isotropic gauge and Γ is the path. Thus, the amplitude is given by the path integral

$$< z_0, T + t_0 | z_0, t_0 >= \mathcal{N} \int \mathcal{D}\vec{x} \, \exp\{i \frac{(e/m)}{\hbar c} \oint_\Gamma d\vec{x} \cdot \vec{A}(\vec{x})\} \qquad (10.4.16)$$

which is just the path integral for a particle of mass M and charge $\frac{e}{m}$ moving in the field $B = \vec{\nabla} \times \vec{A} \cdot \hat{z}$ in the limit $M \to 0$. This limit is just the projection onto the lowest Landau level. Notice that the normalization constants $< z|z >$ have been absorbed into the uninteresting factor $\mathcal{N}$. The amplitude for the path Γ of this path integral is just the Aharonov-Bohm phase factor [Arovas 84]. At the end of this chapter we give a derivation of this result based on the field theory approach, which does not require the choice of a set of wave functions with a specific form.

Let us briefly discuss the generalization of this result for the problem of two quasiholes. Let us assume that at some initial time t_0 the quasiholes are prepared in the state $|z_0, w_0 >$. Once again we ask for the amplitude $< z_0', w_0'; t_0 + T | z_0, w_0; t >$ after a very long time T. The normalized two-quasiholes states $|z, w >$ will be taken to be of the Halperin form. The derivation for two quasiholes follows quite closely the arguments given for one quasihole. However, the two results differ in two important aspects :(a) the multivalued phase factors $(z - w)^{1/m}$ lead to an "induced" gauge interaction and (b) the diagonal overlaps are no longer constant but functions of $|z - w|$. The final result is

$$< \vec{z}_0', \vec{w}_0' | \vec{z}_0, \vec{w}_0 >= \mathcal{N} \int \mathcal{D}\vec{z} \mathcal{D}\vec{w} \, \exp\{\frac{i}{\hbar} S_{\text{eff}}^{(2)}(\vec{z}, \vec{w})\} \qquad (10.4.17)$$

where $S_{\text{eff}}^{(2)}(\vec{z}, \vec{w})$ is the effective action for two quasiholes. The integration measure, denoted here by " $\mathcal{D}\vec{z}\,\mathcal{D}\vec{w}$", has absorbed the diagonal overlaps $\prod_n < \vec{z}_n, \vec{w}_n | \vec{z}_n, \vec{w}_n >$. Laughlin [Laughlin 87] has shown that these factors are constant at long distances but vanish at short distances like $|\vec{z} - \vec{w}|^{\frac{2}{m}}$. Thus, their main effect is to remove from the path-integral the paths in which the particles get to be too close to each other. This feature of the integration measure is essential since fractional statistics cannot be defined if the paths of the particles are allowed to cross.

The effective action $S_{\text{eff}}^{(2)}$ for two holes is

$$\begin{aligned} S_{\text{eff}}^{(2)}(\vec{z}, \vec{w}) = \int_{t_0}^{t_0+T} dt \, \{ & \frac{d\vec{z}}{dt} \cdot (\frac{e/m}{c} \vec{A}(\vec{z}) + \frac{\hbar}{m} \vec{\mathcal{A}}(\vec{z} - \vec{w})) + \\ & + \frac{d\vec{w}}{dt} \cdot (\frac{e/m}{c} \vec{A}(\vec{w}) + \frac{\hbar}{m} \vec{\mathcal{A}}(\vec{w} - \vec{z}))\} \end{aligned} \qquad (10.4.18)$$

where m is the index of the Laughlin state (not to be confused with a mass!) and $\vec{A}$ is the electromagnetic vector potential. The "induced" vector potential $\vec{\mathcal{A}}$ arises from the multivalued factors. It is given by the total change of phase

accumulated during the process, i.e.

$$\frac{1}{m}\int_{t_0}^{t_0+T}[\vec{\mathcal{A}}(\vec{z}-\vec{w})\cdot\frac{d\vec{z}}{dt}+\vec{\mathcal{A}}(\vec{w}-\vec{z})\cdot\frac{d\vec{w}}{dt}]= \\ =\frac{1}{m}[\arg(z_0'-w_0')-\arg(z_0-w_0)]. \tag{10.4.19}$$

This equation only requires that the "induced" vector potential $\vec{\mathcal{A}}$ give the correct winding number. It is clear that $\vec{\mathcal{A}}$ can be represented by an effective Chern-Simons gauge field with an appropiately chosen coupling constant. One possible choice for $\vec{\mathcal{A}}$ was given by Arovas, Schrieffer and Wilczek [Arovas 84] (in the isotropic gauge) to be

$$\mathcal{A}_j(\vec{z}-\vec{w})=\frac{\epsilon_{jk}(z-w)_k}{|\vec{z}-\vec{w}|^2} \tag{10.4.20}$$

which has the quantized circulation

$$\oint_{C[\vec{w}]}\mathcal{A}_j(\vec{z}-\vec{w})\,dz_j=2\pi \tag{10.4.21}$$

for any closed path $C[\vec{w}]$ which encloses the point $\vec{w}$.

Hence, each quasihole carries a solenoid with just one flux quantum. In agreement with our discussion of section (7.2), these *"induced " or statistical gauge fields* change the statistics of the quasiparticles. In the problem of spinons in the Chiral Spin State (see section 7.1) the quasiparticles are semions or half-fermions. The quasiholes of the FQHE have statistical angle equal to $\frac{\pi}{m}$. This property can be seen very directly from the coherent-state path integral. Let us consider a process in which two quasiholes undergo a counterclockwise exchange, during which $\Delta\arg(z_0-w_0)=\pi$. The amplitude of the path integral picks up a phase of $e^{i\pi/m}$. Below, when we derive the Laughlin theory from a field theory, we will see that these phase factors arise directly from a Chern-Simons gauge field.

10.5 The Field Theory Approach to The Fractional Quantum Hall Effect

In the past sections we discussed the first quantization approach to the FQHE. Here we will discuss an alternative approach which is based on a special form of field theory, the Chern-Simons theory which we discussed extensively in chapter (7) in the context of theories of anyons. Here we will show that the Chern-Simons theory is quite useful from two different points of view: (a) as a Landau-Ginzburg theory for the long distance phenomenology and (b) as a way to derive the Laughlin state from a microscopic theory. For reasons of space and conciseness I will only discuss the simplest case of fully polarized (*i.e.*, "spinless") electrons. Also I will restrict myself to the theory of the

Laughlin sequence and to the first level of the hierarchy. The methods that I will describe below can be used (and have been used) to study the problem of *unpolarized* electrons, the *singlet* Hall effect [A. Balatsky and E.Fradkin, unpublished, 1990; G.Moore and N.Read, unpublished, 1990]. They are also very useful for the study of the structure of the hierarchical states [V.Blok and X.G.Wen, unpublished, 1990].

In section 8.10 we saw that the construction of the state for the quasihole suggested a different interpretation of the Laughlin wave-function first proposed by Jain. This structure of the state for the quasihole gave rise to the picture of the FQHE as a ground state of "electrons bound to fluxes". From this point of view, all what the long range correlations do is to make it possible for the electrons to "nucleate" flux. Jain [Jain 89] realized that, in the Laughlin state, the electrons nucleate enough flux so that the bound states *exactly* fill up an integer number of the Landau levels of the *unscreened* part of the field. In this formulation, the FQHE is an Integer Quantum Hall Effect of the bound states. Jain proposed to write the Laughlin wave function in the suggestive factorized form

$$\Psi(z_1,\ldots,z_N)=\prod_{i<j}(z_i-z_j)^{m-1}\ \chi_1(z_1,\ldots,z_N). \tag{10.5.1}$$

where χ_1 is the wave function for a completely filled lowest Landau level

$$\chi_1(z_1,\ldots,z_N)=\prod_{i<j}(z_i-z_j)\ \exp(-\sum_{i=1}^{N}\frac{|z_i|^2}{4\ell^2}). \tag{10.5.2}$$

The phases associated with the factor multiplying χ_1 represent an even number $(m-1)$ of fluxes which are attached to each coordinate z_i where an electron is present. It is a crucial feature of this picture that the electrons bind to an *even* number of flux quanta and, in this way, they retain their fermion character. We will also see below that this approach has allowed for a simple description of the so-called Hierarchy states in terms of wave functions which have a factorized structure.

In chapter (7) we saw that there is a natural and local way to attach particles and fluxes together: the Chern-Simons gauge theory. The relevance of the Chern-Simons theory to the physics of the FQHE actually predates its application to problems (presumably) connected to High Temperature Superconductors, such as the Chiral Spin Liquid. Girvin and MacDonald [Girvin 87] were the first to realize that the Laughlin state had a hidden form of Off-Diagonal Long Range Order (ODLRO). They further suggested an order parameter for the Laughlin state but it turned out to be non-local. As a matter of fact, the Girvin-MacDonald order parameter is closely related to the anyon operators constructed in chapter (7). We also saw that it is always possible to map *any* two-dimensional fermion system into an equivalent problem with arbitrarily chosen statistics. We are going to use this mapping in two different ways: (a) as a mapping to a theory of bosons and (b) to a theory of fermions (each coupled to a Chern-Simons gauge field with a suitably chosen coupling constant).

10.6 Off-Diagonal Long Range Order

The Girvin-MacDonald argument that the Laughlin state has a hidden form of off-diagonal long range order goes as follows. The ground state correlation function $\rho(z,z')$ for the electron operator (also called the one-particle density matrix) in the m^{th} Laughlin state for a system with N particles $|0_m;N\rangle$ is given by the expansion

$$\begin{aligned}\rho(z,z') &\equiv \langle 0_m;N|\hat{\psi}^\dagger(z)\hat{\psi}(z')|0_m;N\rangle \\ &= \sum_{n,k}\varphi_n^*(z)\varphi_k(z')\langle 0_m;N|\hat{\psi}_n^\dagger\hat{\psi}_k|0_m;N\rangle \end{aligned} \tag{10.6.1}$$

where $\{\varphi_n(z)\}$ is the set of one-particle wave functions of the lowest Landau level (see section 8.1) and n and k run over all the occupied states. Since the states $\{\varphi_n(z)\}$ all have different angular momentum, the expectation value $\langle 0_m|\hat{\psi}_n^\dagger\hat{\psi}_n|0_m;N\rangle$ in an isotropic uniform state, such as the Laughlin state, takes the very simple form

$$\langle 0_m;N|\hat{\psi}_n^\dagger\hat{\psi}_k|0_m;N\rangle = \nu\delta_{nk} \tag{10.6.2}$$

where ν is the filling fraction. The correlation function can be shown to be given by [Girvin 84]

$$\rho(z,z') = \frac{\nu}{2\pi}\exp\left(-\frac{|z-z'|^2}{4\ell^2}\right)\exp\left[\frac{1}{4\ell^2}(z^*z'-z'^*z)\right] \tag{10.6.3}$$

This identity shows that the one-particle electron correlation function decays exponentially fast in a Laughlin ground state.

Consider now the *composite operator* $\hat{K}(z)$, introduced by Read [Read 89](see also [Rezayi 88]), who refined the arguments of Girvin and MacDonald. The operator $\hat{K}(z)$ that creates one electron, together with a solenoid carrying m flux quanta, at point (z) is

$$\hat{K}(z) = \hat{\psi}^\dagger(z)\hat{U}^m(z) \tag{10.6.4}$$

where $\hat{U}(z)$ is the second quantized operator which creates a quasihole at z.

Each quasihole has charge $\frac{1}{m}$ and fractional statistics $\frac{\pi}{m}$ and m quasiholes have charge 1 and statistics π. Thus m holes have the same quantum numbers as a *missing electron*. Furthermore, the operator $\hat{K}(z)$ obeys *bosonic* commutation relations. This implies that the operator $\hat{K}(z)$ must have a non-vanishing expectation value in a ground state with an indefinite number of particles. This property is indeed strongly reminiscent to Bose condensation. More precisely, Read showed that the following identity holds

$$\langle 0_m;N|\hat{K}^\dagger(z)\hat{K}(z')|0_m;N\rangle = \frac{1}{\rho_0}\langle 0_m;N+1|\hat{\rho}(z)\hat{\rho}(z')|0_m;N+1\rangle \to \rho_0 \tag{10.6.5}$$

where $\hat{\rho}(z)$ is the density operator and its expectation value is $\rho_0 = \frac{1}{2\pi m}$. Thus, there is ODLRO in the Laughlin state.

Since ODLRO is the hallmark of superfluidity, its existence suggested the idea that there should be a Landau-Ginzburg theory for the FQHE. But, unlike superfluids, the Laughlin state is an incompressible state and it does not have excitations with arbitrarily low energy. So, whatever the Landau-Ginzburg theory happened to be it could not have any Goldstone modes. Now, a system with an order parameter which is complex as the Girvin-MacDonald order parameter is, in principle should have Goldstone modes unless the order parameter is coupled to a fluctuating gauge field. In this case the gauge field would "eat" the Goldstone mode and, at the same time, it would become massive. In this case , there would not be any gapless modes left. This phenomenon, usually called the Anderson-Higgs mechanism, does take place in charged superfluids *i.e.,* superconductors. This is the Meissner state of the superconductor.

The problem with this picture is that the FQHE is not a superconductor! There is no flux expulsion in the Laughlin state. Moreover, the non-locality of the Girvin-MacDonald order parameter is clearly indicating that a naive application of the Anderson-Higgs mechanism is not possible. So the gauge fields have to arise from the fluctuations about the Laughlin ground state rather than come from "honest-to-god" electromagnetism. In other words, the gauge field has to be self-generated. Furthermore, since the Laughlin state is not a superconductor, the mechanism for generation of mass (or gaps) to all excitations should be gauge invariant. This fact suggested to Girvin and MacDonald that the gauge field should have a Chern-Simons form.

10.7 The Landau-Ginzburg Theory of the Fractional Quantum Hall Effect

The methods that we have discussed for the field theoretic treatment of anyons can also be used to study the FQHE. As a matter of fact, Zhang, Hansson and Kivelson (ZHK) [Zhang 89] used a mapping to *bosons* in terms of a Chern-Simons gauge field. This procedure allowed them to derive the qualitative features of a Landau-Ginzburg theory for the FQHE. Their Landau-Ginzburg approach, which is valid at low energies and long distances, qualitatively confirmed the idea that the FQHE had a hidden form of ODLRO without Goldstone bosons. Read [Read 89] has given a careful derivation of the Landau-Ginzburg theory directly from the Laughlin wave function.

Let us use the methods of chapter (7) to derive the Landau-Ginzburg theory. Consider once again a system of N electrons moving on a plane in the presence of an external uniform magnetic field B perpendicular to the plane. The electrons will be assumed to have an interparticle interaction governed by a pair potential $V(|\vec{r}|)$, for two electrons separated a distance $|\vec{r}|$ on the plane. The magnetic filed will be assumed to be so large that the system is completely polarized and that we can ignore the spin degrees of freedom. The eigenstates

$\Psi(\vec{x}_1, \ldots, \vec{x}_N)$ are eigenfunctions of the (first quantized) Hamiltonian $\hat{H}$

$$\hat{H} = \sum_{i=1}^{N} \{ \frac{1}{2m} \left(\vec{p}_j - \frac{e}{c} \vec{A}_j(\vec{x}_j) \right)^2 + eA_0(\vec{x}_j) \} + \sum_{i<j} V(|\vec{x}_i - \vec{x}_j|). \qquad (10.7.1)$$

where we have included the coupling to both the electromagnetic vector potential $\vec{A}$ and scalar potential A_0. Hence, we are dealing with N spinless fermions of charge $-e$ and mass m. In second quantized notation, the electron operator is $\psi(x)$ and the dynamics of the system is governed by the action $\mathcal{S}$

$$\begin{aligned} \mathcal{S} = & \int d^3z \left\{ \psi^*(z)[iD_0 + \mu]\psi(z) + \frac{1}{2m} |\vec{D}\psi(z)|^2 \right\} + \\ & -\frac{1}{2} \int d^3z \int d^3z' \, (|\psi(z)|^2 - \rho_0) V(|\vec{z} - \vec{z}\,'|)(|\psi(z')|^2 - \rho_0). \end{aligned} \qquad (10.7.2)$$

where ρ_0 is the average density. The quantum partition function $\mathcal{Z}$ for this system is (at zero temperature and in real time)

$$\mathcal{Z} = \int \mathcal{D}\psi^* \mathcal{D}\psi \exp\{\frac{i}{\hbar}\mathcal{S}\} \qquad (10.7.3)$$

10.8 Mapping to a Chern-Simons Theory with Bosons

In chapter (7) we showed that a system of fermions in two dimensions is equivalent to a system of bosons coupled to a Chern-Simons gauge field $\mathcal{A}_\mu$. The action $\mathcal{S}_B$ for the Bose system is

$$\begin{aligned} \mathcal{S}_B = & \int d^3z \left\{ \phi^*(z)[iD_0 + \mu]\phi(z) + \frac{1}{2m} |\vec{D}\phi(z)|^2 + \frac{\theta}{4} \epsilon_{\mu\nu\lambda} \mathcal{A}^\mu \mathcal{F}^{\nu\lambda} \right\} \\ & -\frac{1}{2} \int d^3z \int d^3z' \, (|\phi(z)|^2 - \rho_0) V(|\vec{z} - \vec{z}\,'|)(|\phi(z')|^2 - \rho_0). \end{aligned} \qquad (10.8.1)$$

where $\phi(z)$ is the Bose field and $\theta = 1/2\pi n$ and n, for the moment, is an arbitrary *odd* integer. The covariant derivatives D_μ in this action contain both electromagnetic and Chern-Simons gauge fields *i.e.*,

$$D_\mu = \partial_\mu + i\frac{e}{c} A_\mu + i\mathcal{A}_\mu. \qquad (10.8.2)$$

It is an implicit assumption of this theory that the bosons must have a hard-core since, otherwise, the fractional statistics transformation does not make sense. It is very difficult to keep track of this constraint in the continuum. On a lattice the hard-core constraint does not pose any serious problem. However, if we are interested *only* in the long distance and low energy behavior, we can replace the hard-core by an effective short distance repulsive force. This change amounts to add an extra term $\mathcal{S}_{hc}$ to the action of the form

$$\mathcal{S}_{hc} = \int d^3z (-\lambda |\phi(z)|^4). \qquad (10.8.3)$$

The total action is $\mathcal{S}_{\text{eff}} = \mathcal{S}_{\text{B}} + \mathcal{S}_{\text{hc}}$ and we have now a bosonic functional integral

$$\mathcal{Z} = \int \mathcal{D}\phi^* \mathcal{D}\phi \exp\{\frac{i}{\hbar}\mathcal{S}_{\text{eff}}\} \tag{10.8.4}$$

This functional integral can be regarded as a Landau-Ginzburg theory and was first proposed by Zhang, Hansson and Kivelson [Zhang 89]. The parameter λ cannot be calculated directly from this theory. In fact, ZHK dropped the repulsive pair potential term altogether and replaced it by the $|\phi|^4$ term. Actually, the effective action has to be derived directly from the Laughlin wave function [Read 89]. We now follow ZHK and extract the low energy behavior.

Landau-Ginzburg Theory

The effective theory looks like a theory of bosons coupled to a gauge field. In the absence of the gauge field, the bosons condense and spontaneously break the *global* phase symmetry $U(1)$

$$\phi(z) \to e^{i\alpha}\, \phi(z). \tag{10.8.5}$$

The system is then a superfluid and its spectrum has a massless excitation, the phase ω of ϕ, which is the Goldstone boson associated with the broken $U(1)$ symmetry. We will see now that this Goldstone boson disappears from the spectrum once the system is coupled to the statistical gauge field.

Let us consider the behavior of the system in the semiclassical (mean-field) limit. In that limit, the fluctuations of the amplitude of the Bose field ϕ are small. Let us write ϕ in the form

$$\phi(z) = \sqrt{\rho(z)}\, \exp[i\omega(z)]. \tag{10.8.6}$$

The classical equations of motion of the Bose theory are, for a configuration ϕ with constant amplitude (the ground state)

$$\begin{aligned} \rho + \theta\langle\mathcal{B}\rangle &= 0 \\ \mu\rho - \lambda\rho^2 &= 0 \\ \langle\mathcal{A}_\mu\rangle + \frac{e}{\hbar c}A_\mu &= 0 \\ \rho - \frac{1}{2\pi m} &= 0 \end{aligned} \tag{10.8.7}$$

Thus, the *average* statistical gauge field $\langle\mathcal{A}_\mu\rangle$ exactly cancels, or *screens*, the electromagnetic field A_μ. We get $\langle\mathcal{B}\rangle = -\frac{e}{\hbar c}B$. However the first equation requires the average statistical magnetic field to be proportional to the average particle density. Hence, the density and the field are not independent from each other but satisfy $\rho = \theta\frac{e}{\hbar c}B$. Recall the definition of the filling fraction ν as $\nu = \phi_0(\rho/B)$, where ϕ_0 is the flux quantum $\phi_0 = \frac{hc}{e}$. Thus, the classical equations of motion only have uniform solutions if the filling fraction ν is $\nu = \frac{\theta}{2\pi} = \frac{1}{n}$, with n odd. We can then identify the odd integer n with the index m of the Laughlin wave function which is also odd. Thus, the Landau

theory suggests the picture of the FQHE as a problem of bosons in an average magnetic field which is determined by the number of bosons!. Notice that, with the identification of n as the index m of the Laughlin wave function, the constraint implies that each boson is made of a fermion and $n = m$ flux quanta. This is precisely what the arguments of Girvin and MacDonald, and Read told us.

10.9 The Low Energy Fluctuations of the Landau Theory

However, this story does not end at the level of mean-field-theory. The fluctuations play a very important role in this problem. Mean-field-theory told us that the average particle density and average statistical magnetic field are fixed. But the fluctuations of the phase ω appear to be completely unconstrained. In order to investigate this problem we need an effective action for the slow modes of the phase field. This effective action can be obtained by integrating out the amplitude fluctuations. Indeed, we can write the field ϕ in the form

$$\phi(z) = [\rho_0 + \delta\rho(z)]^{\frac{1}{2}} \exp[i\omega(z)]. \tag{10.9.1}$$

We now substitute this expression back into the Landau-Ginzburg action and expand it in powers of the density fluctuation $\delta\rho(z)$ up to second order. By integrating out the density fluctuations we get the effective Lagrangian for the fluctuations of the phase and statistical gauge fields

$$\mathcal{L}_{\text{eff}} = \frac{\kappa}{2}\left[\frac{1}{v^2}\left(\partial_0\omega + \mathcal{A}_0\right)^2 - \left(\vec{\nabla}\omega + \vec{\mathcal{A}}\right)^2\right] + \frac{\theta}{4}\epsilon_{\mu\nu\lambda}\mathcal{A}_\mu\mathcal{F}_{\nu\lambda} \tag{10.9.2}$$

where the rigidity κ and the velocity v have to be determined from the microscopic theory. This effective Lagrangian has the same form as the one we derived for the anyon superconductor in chapter (7) except for the very important difference that the gauge field here is the statistical one whereas there it was the electromagnetic field. Nevertheless, the phase field still disappears from the spectrum. Indeed, the phase field ω can be eliminated by a gauge transformation $\mathcal{A}_\mu = \mathcal{A}'_\mu - \partial_\mu\omega$. The resulting theory is that of a gauge field which has just two massive modes. The masses were also calculated in chapter (7). Thus, this is an incompressible ground state. The two massive modes represent the magnetophonon and magnetoplasmon which were derived directly from Laughlin's theory by Girvin, MacDonald and Platzman [Girvin 86]. The magnetoroton cannot be studied within the Landau-Ginzburg approach.

The Order Parameter of the Fractional Quantum Hall Effect

In hindsight, we can construct the order parameter directly in the theory of bosons, without having to rely on the Landau-Ginzburg theory. The first

guess is that the order parameter is the Bose field ϕ itself. However, ϕ is not invariant under gauge transformations of the statistical gauge field. Thus, its expectation value, as well as the expectation values of any product of ϕ's, is zero when averaged over *all* configurations of the gauge field. It may be argued that this is not much of a problem since one always has to fix the gauge. Since this gauge theory is abelian and non-compact, all small gauge transformations (*i.e.*, those which do not wind around the system) are connected to the identity and it is possible to fix the gauge completely. Now, the expectation value of products of ϕ fields will depend on the gauge in which it is evaluated. Thus, it does not represent a physical observable. However, all we need is *an* operator which in some convenient *gauge* reduces to a product of ϕ fields. Fortunately, it is quite easy to construct such operators.

Let us consider the case of the boson correlation function, which is the expectation value of the product $\phi^\dagger(x)\phi(y)$, where x and y are two arbitrary points in 2+1-dimensional space-time. Under a gauge transformation $\phi(x) \rightarrow \exp(i\Lambda(x))\phi(x)$, the product transforms like

$$\phi^\dagger(x)\phi(y) \rightarrow \exp[i(-\Lambda(x) + \Lambda(y))]\ \phi^\dagger(x)\phi(y). \tag{10.9.3}$$

Thus, we need to find an operator which transforms in the opposite way and cancels the unwanted phase factor. One possibility is the exponential of the line integral $\int_\Gamma \mathcal{A}_\mu dx_\mu$, where Γ is a path that goes from x to y. But this is just an Aharonov-Bohm phase factor, which fluctuates very rapidly and it does not vanish in any gauge. It can be shown that the expectation value of the product

$$\phi^\dagger(x) \exp\left(i \int_\Gamma \mathcal{A}_\mu dx_\mu\right)\ \phi(y) \tag{10.9.4}$$

decays rapidly as $|x - y| \rightarrow \infty$.

Let us consider the operator $\mathcal{O}^\dagger(x)\mathcal{O}(y)$

$$\mathcal{O}^\dagger(x)\mathcal{O}(y) \equiv \exp\left(i \int d^3z \mathcal{A}_\mu(z) B^c_\mu(z)\right)\ \phi^\dagger(x)\phi(y) \tag{10.9.5}$$

where A^c_μ is some suitably chosen, fixed classical configuration. We will choose B^c_μ in such a way that the product $\mathcal{O}^\dagger(x)\mathcal{O}(y)$ is gauge invariant and that, in the Landau-Lorentz gauge ($\partial_\mu \mathcal{A}_\mu = 0$), it reduces to the product $\phi^\dagger(x)\phi(y)$. Under a gauge transformation which vanishes at infinity $\lim_{|x|\rightarrow\infty} \Lambda(x) = 0$,

$$\begin{aligned} \phi(x) &= \exp(i\Lambda(x))\ \phi'(x) \\ \mathcal{A}_\mu(x) &= \mathcal{A}'_\mu(x) - \partial_\mu \Lambda(x) \end{aligned} \tag{10.9.6}$$

the operator $\mathcal{O}^\dagger(x)\mathcal{O}(y)$ transforms like

$$\mathcal{O}^\dagger(x)\mathcal{O}(y) = \exp(i\Phi)\ \mathcal{O}^\dagger(x)\mathcal{O}(y) \tag{10.9.7}$$

where Φ is given by

$$\Phi = \Lambda(y) - \Lambda(x) + \int d^3z\ \Lambda(z)\ \partial_\mu B^c_\mu(z) \tag{10.9.8}$$

Gauge invariance demands that $\Phi \equiv 0$ for *all* gauge transformations $\Lambda(z)$ and for *all* points x and y. The only way to meet these requirements is for $B^c_\mu(z)$ to satisfy the equation

$$\partial_\mu B^c_\mu(z) = \delta(z-x) - \delta(z-y) \tag{10.9.9}$$

We can think of $B^c_\mu(z)$ as being the *classical* magnetic field of two magnetic monopoles of (opposite) unit magnetic charge located at x and y. If we denote by $U(z)$ the "potential", we get

$$\begin{aligned} B^c_\mu(z) &= \partial_\mu U(z) \\ \nabla^2 U(z) &= \delta(z-x) - \delta(z-y) \end{aligned} \tag{10.9.10}$$

the solution of which is just the electrostatic potential for two unit and opposite charges.

Having checked that it is gauge invariant, we now want to see what this operator is in the Landau-Lorentz gauge ($\partial_\mu \mathcal{A}_\mu = 0$). In this gauge, the exponential part of the operator vanishes

$$\int d^3z\ \mathcal{A}_\mu(z) B^c_\mu(z) = \int d^3z\ \mathcal{A}_\mu(z)\partial_\mu U(z) = -\int d^3z\ \partial_\mu \mathcal{A}_\mu(z) U(z) = 0 \tag{10.9.11}$$

Thus, in the Landau-Lorentz gauge, we get

$$\mathcal{O}^\dagger(x)\mathcal{O}(y) \equiv \phi^\dagger(x)\phi(y) \tag{10.9.12}$$

Therefore, the operator $\mathcal{O}(x)$, defined by

$$\mathcal{O}(x) \equiv \phi(x) \exp\left(i \int d^3z\ \mathcal{A}_\mu(z) B^c_\mu(z) \right) \tag{10.9.13}$$

where $B^c_\mu(z) = \partial_\mu U(z)$ is the field created by a single charge at x, is the gauge invariant order parameter operator for this problem in the boson description since, in this gauge, it becomes identical to the field operator of the bosons. Thus, the correlation functions of this operators exhibit long-range order.

In an arbitrary gauge, this operator is highly non-local. But, in the Landau-Lorentz gauge, it becomes local andjust simple. This is not a surprise since, for instance the order parameter of an ordinary BCS-like superconductor is only local in this gauge. Indeed, it is possible to define an order parameter for a superconductor in the same way. For practical purposes, in the case of a superconductor, this is not very useful since the electromagnetic field is not usually treated as a dynamical field. In the problem of the FQHE, the gauge field is dynamically generated and it plays an essential role.

10.10 The Fermion Field Theory Picture of the FQHE

In this section we derive a field theory for the FQHE based on the fermion picture. These methods, which have been so succesful in the treatment of

Anyon Superfluidity (see chapter (7)), are also very useful for the study of the FQHE. They have a great advantage over the boson theories in that there is no difficulty in handling the short distance behavior unlike the case of bosons. It is quite easy to derive an effective action for the fluctuations which explicitly involves Chern-Simons gauge fields. The Landau-Ginzburg theory can be seen to be the *dual* of the fermion theory in very much the same way as in the case of the anyon superconductor. The fermion field theory has been developed by Lopez and myself [A.Lopez and E.Fradkin, unpublished, 1990].

Let us go back to the second quantized form of the problem of electrons in a magnetic field. In its standard form, the dynamics is governed by the action

$$\begin{aligned} \mathcal{S} = & \int d^3z \left\{ \psi^*(z)[iD_0 + \mu]\psi(z) + \frac{1}{2m}|\vec{D}\psi(z)|^2 \right\} + \\ & -\frac{1}{2}\int d^3z \int d^3z' \, (|\psi(z)|^2 - \rho_0)\, V(|\vec{z} - \vec{z}\,'|)\, (|\psi(z')|^2 - \rho_0). \end{aligned} \tag{10.10.1}$$

Since we are dealing with a problem in which one Landau level is fractionally filled, we do not expect that the semiclassical approximation for this problem will, in general, be very reliable. Unless, of course, the ground state of the system is such that there is a gap in the energy spectrum. For example, in the low-density limit, the system can lower its energy by modulating the electron density and forming a Wigner crystal. Wigner crystals can also be studied with a path-integral of this section, but we will not do it here.

Let us recall Jain's interpretation of the Laughlin state as a state in which the electrons "nucleate" flux to screen enough of the external magnetic field, so that the bound states of electrons plus fluxes exactly filled an integer number of Landau levels. In this section we are going to use the *periodicity property* of theories of fermions coupled to Chern-Simons gauge fields, derived in chapter (7), to make this nucleation picture more explicit.

In chapter (7) we saw that a system of fermion could be mapped into a system of fermions coupled to Chern-Simons gauge fields if the Chern-Simons coupling constant is chosen to be equal to $\theta = 1/2\pi n$, where n is an *even* integer. Thus, the problem becomes equivalent to a theory with fermions and gauge fields with an action given by

$$\begin{aligned} \mathcal{S}_\theta = & \int d^3z \left\{ \psi^*(z)[iD_0 + \mu]\psi(z) + \frac{1}{2m}|\vec{D}\psi(z)|^2 + \frac{\theta}{4}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu\mathcal{F}^{\nu\lambda} \right\} \\ & -\frac{1}{2}\int d^3z \int d^3z' \, (|\psi(z)|^2 - \bar{\rho})V(|\vec{z} - \vec{z}\,'|)(|\psi(z')|^2 - \bar{\rho}). \end{aligned} \tag{10.10.2}$$

where $\psi(z)$ is a second quantized Fermi field, μ is the chemical potential and D_μ is the covariant derivative which couples the fermions to both the external electromagnetic field A_μ and to the statistical gauge field $\mathcal{A}_\mu$

$$D_\mu = \partial_\mu + i\frac{e}{c}A_\mu + i\mathcal{A}_\mu. \tag{10.10.3}$$

We are going to see below, that the *even* integer n has to be identified with $m-1$, where m is the index of the Laughlin state.

10.11 The Semiclassical Limit and The Laughlin Ground State

In this section we will show that the *semiclassical limit* of the theory described by the action S_θ, with $\frac{1}{\theta} = 2\pi(m-1)$, yields the same physics as the Laughlin state. In order to prove this statement we will develop a semiclassical approach to this problem. As a by-product, this formalism provides for a systematic procedure to compute the corrections to the Laughlin approximation. This is, to the best of my knowledge, the first formalism for which the Laughlin *ansatz* arises as the first of a series of approximations.

The action $\mathcal{S}_\theta$ governs the dynamics of a system of spinless *fermions* interacting through a pair interaction potential $V(|\vec{x}-\vec{x}'|)$ coupled to both electromagnetic and statistical gauge fields. The starting point of the semiclassical approximation maps this FQHE problem into an *equivalent* IQHE system. This mapping is made possible by the statistical or Chern-Simons gauge fields which screen-out enough of the external magnetic field to the point that the number of flux quanta of the effective magnetic field which is left, is an exact factor of the total number of particles. Naturally, this perfect screening is not possible for arbitrary values of the external magnetic field for a fixed number of electrons. The values of the filling fraction for which this perfect screening can be accomplished happens to be the same as the Laughlin sequence $\frac{1}{m}$ and the first level of the hierarchy. For all other cases, there will be some partially filled level leftover. As we discussed in section (8.11), these quasiparticles are anyons.

Consider the quantum partition function for this problem (at $T=0$)

$$\mathcal{Z} = \int \mathcal{D}\psi^* \mathcal{D}\psi \mathcal{D}\mathcal{A}_\mu \exp(iS_\theta). \tag{10.11.1}$$

We will treat this path-integral in the semi-classical approximation. In order to do that we will first integrate-out the fermions and treat the resulting theory within the saddle-point-expansion (SPA). The procedure is almost identical to the theory of Anyon Superconductivity discussed in chapter (7). In the absence of electron-electron interactions the fermions can be integrated-out immediately since the action becames quadratic in Fermi fields. In the presence of interactions, this is no longer possible since the interaction term makes the action quartic in the Fermi fields. This problem can be sidestepped by means of a Hubbard-Stratonovich transformation by which we trade a quartic form in fermions for a quadratic action coupled to a new Bose field , the density fluctuation. This procedure will allow us to give a full description of the spectrum of collective modes of the FQHE states. Note that, since we are dealing

with a gauge theory, a gauge has to be specified in order to make the functional integral well defined. We will assume that a gauge fixing condition has been imposed but, for the moment, we will not make any specific choice of gauge.

Before we proceed to integrate the Fermi degrees of freedom out, we perform the Hubbard-Stratonovich transformation in terms of a scalar Bose field $\lambda(x)$. Let F be the weight in the path-integral amplitude which contains in its exponent the terms in the action which are quartic in ψ *i.e.*,

$$F = \exp\left\{-i\int d^3z \int d^3z' \, \frac{1}{2}(|\psi(z)|^2 - \bar{\rho})\, V(z-z')\, (|\psi(z')|^2 - \bar{\rho})\right\} \tag{10.11.2}$$

The Hubbard-Stratonovich transformation allows us to write F as a gaussian functional integral over the Bose field $\lambda(x)$ in the form

$$F = \mathcal{N}\int \mathcal{D}\,\lambda\, \exp\left\{i\int d^3z\, \lambda(z)\, (|\psi(z)|^2 - \bar{\rho})\right\} \times \\ \times \exp\left\{\frac{i}{2}\int d^3z \int d^3z'\, \lambda(z)\, V^{-1}(z-z')\, \lambda(z')\right\} \tag{10.11.3}$$

where $\mathcal{N}$ is a normalization constant and $V(z-z')$ represents the instantaneous pair interaction *i.e.*,

$$V(z-z') = V(|\vec{z}-\vec{z}'|)\delta(t-t'). \tag{10.11.4}$$

$V^{-1}(z-z')$ is the inverse of $V(z-z')$ in an operator sense. I will assume that the physics of the FQHE can be studied in a model system in which the pair potential is reasonably local, such as

$$V(|\vec{z}-\vec{z}\,'|) = V_2 \nabla^2_{\vec{z}}\, \delta(\vec{z}-\vec{z}\,'). \tag{10.11.5}$$

For this case, the inverse potential $V^{-1}(z-z')$ takes the particularly simple form

$$V^{-1}(z-z') = \frac{1}{V_2} G_0(|\vec{z}-\vec{z}'|)\delta(t-t'). \tag{10.11.6}$$

Here $G_0(R)$ is the two-dimensional Coulomb green function. Hence,

$$V^{-1}(R) = \frac{1}{2\pi V_2}\ln(R/a). \tag{10.11.7}$$

where a is a constant with dimensions of length which represents the range of the interaction.

After the Hubbard-Stratonovich transformation is performed, the partition function $\mathcal{Z}$ can be written in the form of a functional integral involving the fermi fields ψ, the statistical gauge fields $\mathcal{A}_\mu$ and the collective modes λ. The action for the system is now given by

$$S = \int d^3z \left\{\psi^*(z)\,(iD_0 + \mu + \lambda(z))\,\psi(z) + \frac{1}{2m}|\vec{D}\psi(z)|^2 - \lambda(z)\bar{\rho}\right\} + \\ + \int d^3z\, \frac{\theta}{4}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu \mathcal{F}^{\nu\lambda} + \frac{1}{2}\int d^3z \int d^3z'\, \lambda(z)\, V^{-1}(z-z')\, \lambda(z'). \tag{10.11.8}$$

The Fermi fields can be integrated out without any difficulty yielding a fermion determinant. The resulting partition function can thus be written in terms of an effective action $S_{\rm eff}$ given by

$$S_{\rm eff} = -i \ {\rm Tr} \ \log[iD_0 + \mu + \lambda + \frac{1}{2m}\vec{D}^2] + \theta S_{\rm CS}(\mathcal{A}_\mu - \tilde{A}_\mu) + \\ - \int d^3z \ \lambda(z)\bar{\rho} + \frac{1}{2}\int d^3z \int d^3z' \ \lambda(z) \ V^{-1}(z-z') \ \lambda(z'). \tag{10.11.9}$$

where D_0 and $\vec{D}$ are the covariant derivatives and $S_{\rm CS}$ is the Chern-Simons action for $\theta = 1$. The field $\tilde{A}_\mu$ represents a small fluctuating electromagnetic field, with vanishing average everywhere, which will be used to probe the system. The electromagnetic currents will be calculated as first derivatives of $\mathcal{Z}$ with respect to $\tilde{A}_\mu$. The full electromagnetic response can be obtained in this way. Notice that since the collective mode $\lambda(z)$ can be thought of as a locally fluctuating electrostatic potential, it couples to the fermions in the same way as the time component of the statistical and electromagnetic vector potentials. From this point of view, it does not describe an independent excitation.

We are now ready to proceed with the semiclassical approximation. The path-integral $\mathcal{Z}$ will be approximated by expanding its degrees of freedom around stationary configurations of the effective action $S_{\rm eff}$ in powers of the fluctuations. This is the conventional WKB approximation. The classical configurations ($\bar{\mathcal{A}}_\mu(z)$ and $\bar{\lambda}(z)$)can be obtained by demanding that $S_{\rm eff}$ be stationary under small fluctuations. This requirement yields the classical equations of motion

$$\frac{\delta S_{\rm eff}}{\delta \mathcal{A}_\mu(z)}|_{\bar{\mathcal{A}},\bar{\lambda}} = 0; \qquad \frac{\delta S_{\rm eff}}{\delta \lambda(z)}|_{\bar{\mathcal{A}},\bar{\lambda}} = 0. \tag{10.11.10}$$

By varying $S_{\rm eff}$ with respect to $\mathcal{A}_\mu(z)$ and $\lambda(z)$ we get

$$\begin{aligned} \langle j^F_\mu(z)\rangle + \frac{\theta}{2}\epsilon_{\mu\nu\lambda}[\langle \mathcal{F}^{\nu\lambda}(z) - e\langle F^{\nu\lambda}\rangle] = 0 \\ \langle j_0(z)\rangle - \bar{\rho} + \int d^3z' \ V^{-1}(z-z') \ \langle \lambda(z')\rangle = 0 \end{aligned} \tag{10.11.11}$$

In addition, we must fix the particle density to be uniform and equal to $\bar{\rho}$ *i.e.*,

$$\langle j_0(z)\rangle = \bar{\rho} \tag{10.11.12}$$

This condition, which implies that we are in a liquid phase, is only consistent with a solution of the equation of motion with

$$\langle \lambda(z)\rangle = 0. \tag{10.11.13}$$

If the external electromagnetic fluctuation is assumed to have zero average, the only time-independent uniform solutions have uniform average statistical flux $\langle \mathcal{B}\rangle$ and vanishing average statistical electric field $\langle \vec{\mathcal{E}}\rangle$ (unless there is a

non-zero current in the ground state) and satisfy

$$\begin{aligned} \langle \mathcal{B} \rangle &= -\frac{\bar{\rho}}{\theta} \\ \langle \vec{\mathcal{E}} \rangle &= 0 \end{aligned} \tag{10.11.14}$$

The non-uniform solutions have $\langle \lambda(z) \rangle$ a periodic function which induces a periodic modulation of the electron density. This is the Wigner crystal state. Notice that, in principle, the crystalline solutions have a modulation in both the charge density and in the local statistical flux.

The equations of motion show that, for a translationally invariant ground state, the effect of the statistical gauge fields, at the level of the saddle-point-approximation, is to *reduce* the effective flux experienced by the fermions. The total effective field is thus reduced from the value of the external field B down to $B_{\text{eff}} = B + \langle \mathcal{B} \rangle = B - \frac{\bar{\rho}}{\theta}$. Let us assume that we have a situation in which we are trying to find the ground state of N (interacting) electrons in the presence of an external magnetic field of strength B. We will further assume that the linear size L of the sample is such that a total of N_ϕ quanta of the magnetic flux are piercing the surface. In general, the filling fraction $\nu = \frac{N}{N_\phi}$ is not an integer. Thus, a perturbative approach based on a Slater determinant wave function of the occupied single particle states does not yield a stable answer. This is so because there is a macroscopic number of essentially degenerate states which will mix with this trial state. On the other hand, a Laughlin state is known to represent a state with an energy gap. Thus, the correlations have removed the massive degeneracy of the free electrons. Since this gap is not equal to the Landau gap of the non-interacting electrons, we can expect our saddle-point-expansion to succeed only if the *effective* theory ends up with a non-zero gap.

It is easy to check that the uniform saddle-point state has a gap *only if* the effective field B_{eff} experienced by the N fermions is such that the *fermions fill exactly an integer number p of the effective Landau levels.* This is precisely the point of view advocated by Jain: the FQHE is an IQHE of a system of electrons dressed by an even number of flux quanta. However, this condition cannot be met for arbitrary values of the filling fraction ν at fixed field (or at fixed density). Let N_ϕ^{eff} denote the effective number of flux quanta piercing the surface after screening. It is given by

$$2\pi N_\phi^{\text{eff}} = 2\pi N_\phi - \frac{\bar{\rho}}{\theta} L^2 = 2\pi N_\phi - 2\pi 2sN. \tag{10.11.15}$$

where $2s$ is an even integer, which before we had denoted by n. The spectrum supported by this state has an energy gap if the N fermions fill exactly p of the Landau levels created by the effective field B_{eff}. In other words, the *effective* filling fraction is $\nu_{\text{eff}} \equiv \frac{N}{N_\phi^{\text{eff}}} = p$. Using these results, we find that the filling fraction ν and the external magnetic field B must satisfy

$$\frac{N}{p} = \frac{N}{\nu} - 2sN, \tag{10.11.16}$$

or, equivalently,

$$\frac{1}{\nu} = \frac{1}{p} + 2s. \tag{10.11.17}$$

Since the filling fraction ν is in general equal to the ratio of two integers, $\nu = \frac{n}{m}$, a solution exists if the integers n and m satisfy

$$\frac{m}{n} = \frac{1}{p} + 2s. \tag{10.11.18}$$

The states are thus parametrized by two integers p (the number of filled Landau levels of the effective field) and $2s$ (the number of flux quanta carried by each fermion). The Laughlin sequence is an obvious solution since, for $n = 1$ and m odd, we get the unique solution $p = 1$ and $2s = m - 1$. The effective fermions thus fill up exactly one Landau level and θ has to be chosen to be $\frac{1}{\theta} = 2\pi(m-1)$. This result agrees with Jain's theory. At this mean-filled level the wave function is the Slater determinant for one filled Landau level χ_1 . The additional factor, $\prod_{i<j}(z_i - z_j)^{m-1}$ is due to the fluctuations of the statistical gauge fields.

In addition to the Laughlin sequence ($p = 1, 2s = m - 1$), there is a host of other solutions to the equations of motion. For $n < m$, we can use the division algorithm to find a pair of integers r and q ($0 \leq q < n$) such that $m = nr + q$. The solution is consistent only if $r = 2s$ and q is a factor of n such that $\frac{q}{n} = \frac{1}{p}$. For instance, the sequence $m = 2sn + 1$ is a solution if $\frac{m}{n}$ is an odd integer ($q = 1$). Clearly, this case has $p = n$ filled Landau levels. This is the first level of the hierarchy. Here too, the semiclassical theory yields the same answers as in Jain's approach.

10.12 The Excitation Spectrum in the Semiclassical Limit

In this section we consider the role of the Gaussian fluctuations around the classical solutions. This is equivalent to a WKB approximation ~~of the functional~~ integral. We begin by considering the effective action. We showed that the saddle-point approximation has a uniform liquid-like solution. Let $\mathcal{A}_\mu(x)$ and $\lambda(x)$ denote the *fluctuations* of the statistical vector potential $\mathcal{A}_\mu$ and of the collective mode $\lambda(x)$ respectively *i.e.,* we set $\mathcal{A}_\mu \to \langle \mathcal{A}_\mu \rangle + \mathcal{A}_\mu$ and $\lambda \to \langle \lambda \rangle + \lambda$. The effective action can be expanded in a series in powers of the fluctuations. We will be interested only in keeping the terms up to quadratic order in the fluctuations. As usual, the linear terms are cancelled if the saddle-point equations are satisfied. It will be convenient to shift the component $\mathcal{A}_0$ of the statistical vector potential by $\mathcal{A}_0 \to \mathcal{A}_0 + \lambda$. In this way, the collective mode λ disappears from the fermion determinant. Naturally, this means that the Chern Simons piece of the action now has the form $S_{\rm CS}(\mathcal{A}_\mu - \tilde{\mathcal{A}}_\mu - \delta_{\mu 0}\lambda)$.

At the quadratic (gaussian) level the effective action has the form

$$S^{(2)} = \frac{1}{2}\int d^3x d^3y\, \mathcal{A}_\mu(x)\Pi^{\mu\nu}(x,y)\mathcal{A}_\nu(y) + \\ + S_{\rm CS}(\mathcal{A}_\mu - \tilde{\mathcal{A}}_\mu - \delta_{\mu 0}\lambda) + S_\lambda(\lambda) \tag{10.12.1}$$

where $S_\lambda(\lambda)$ is the part of the effective action which only depends on λ *i.e.*,

$$S_\lambda(\lambda) = \frac{1}{2}\int d^3z \int d^3z'\, \lambda(z)V^{-1}(z-z')\lambda(z') - \int d^3z\, \lambda(z)\bar{\rho} \tag{10.12.2}$$

The polarization tensor $\Pi_{\mu\nu}$ was derived explicitly in section (8.6). For a system with an integer number of Landau levels, the most important properties of $\Pi_{\mu\nu}$ are its transversality (*i.e.*, $\partial_\mu \Pi_{\mu\nu} = 0$) and that it can be expanded in powers of gradients. This last property is a consequence of the fact that the system has an energy gap. Thus, in close analogy with the problem of Anyon Superconductivity of chapter (7), gauge invariance and locality are sufficient to fix the form of the effective action for the low-energy fluctuations.

To leading order in fluctuations and in gradients, we get the following effective action

$$S_{\rm eff} = \int d^3z \left(\frac{\epsilon}{2}\vec{\mathcal{E}}^2 - \frac{\chi}{2}\mathcal{B}^2 + \frac{\sigma^0_{xy}}{4}\epsilon_{\mu\nu\lambda}\mathcal{A}_\mu\mathcal{F}_{\nu\lambda}\right) + \theta\mathcal{S}_{\rm cs}(\mathcal{A}_\mu - \tilde{A}_\mu - \lambda\delta_{\mu 0}) + \mathcal{S}_\lambda(\lambda) \tag{10.12.3}$$

Once again, we find that the effective action is parametrized in terms of the three quantities ϵ, χ and σ^0_{xy} which we already discussed in chapter (7). For exactly the same arguments, we expect that ϵ and χ will have significant finite renormalizations but the Hall conductance σ^0_{xy} will remain unrenormalized at the value predicted by mean-field-theory. Thus, we know that, for a state with an integer number p of filled Landau levels, $\sigma^0_{xy} = \frac{p}{2\pi}$ (in units of $\frac{e^2}{\hbar}$).

The Chern-Simons term in the effective action $S_{\rm eff}$ can be expanded to give

$$\mathcal{S}_{\rm cs}(\mathcal{A}_\mu - \tilde{A}_\mu - \lambda\delta_{\mu 0}) = \mathcal{S}_{\rm cs}(\mathcal{A}_\mu - \tilde{A}_\mu) - \int d^3z\, \lambda(z)(\mathcal{B}(z) - \tilde{B}(z)). \tag{10.12.4}$$

Since the term in the action $\mathcal{S}_\lambda(\lambda)$, which is determined by the interaction potential V, is quadratic in the collective mode λ, we can further simplify the effective action by doing the gaussian integral over λ. After some straightforward algebra, the final form of the effective action is

$$S_{\rm eff} = \int d^3z \left(\frac{\epsilon}{2}\vec{\mathcal{E}}^2 - \frac{\chi}{2}\mathcal{B}^2\right) - \int d^3z \int d^3z'\, \frac{\theta^2}{2}\bar{B}(z)\, V(z-z')\, \bar{B}(z') + \\ + (\sigma^0_{xy} + \theta)\mathcal{S}_{\rm cs}(\mathcal{A}_\mu) + \theta\mathcal{S}_{\rm cs}(\tilde{A}_\mu) - \int d^3z \frac{\theta}{2}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu \tilde{F}^{\nu\lambda} \tag{10.12.5}$$

where I used the definition $\bar{B}(z) \equiv (\mathcal{B}(z) - \tilde{B}(z))$. This effective action is sufficient to derive the behavior of the collective modes, their spectrum and dynamics, as well as the electromagnetic response of the system and the statistics of the quasiparticles. Notice that, if the electromagnetic fluctuation $\tilde{A}_\nu$

is turned off, the action for the statistical gauge field has a Chern-Simons with a coupling constant equal to the *sum* of the bare (θ) and induced (σ_{xy}) couplings. In the Anyon Superconductor of chapter (7), these two contributions *cancelled* each other out leading to a *compressible* state. In a Fractional Hall state, they *add up* and the state is *incompressible*. It is also worth to note that, except for the "Maxwell-like" first two terms, this expression is exact and independent of the gradient fluctuation. In particular, it contains the *exact* dependence on the interaction pair potential V.

I will show now how this formalism can be used to compute the Hall conductance σ_{xy} and the statistics of the quasiparticles. Let us first note that the quantity σ^0_{xy} is the Hall conductance of the effective fermions in mean-field-theory and it is not equal to σ_{xy}. In particular it is equal to $\sigma^0_{xy} = p/2\pi$, and shows it integer quantization instead of a *fractional* Hall conductance. The full Hall conductance is obtained by calculating the electromagnetic response function.

If we are only interested in the behavior at very low frequency and momentum, we can further approximate S_{eff} by keeping only the terms with the smallest number of derivatives. The Chern-Simons terms have just one derivative while the other terms have at least two. Thus, at long wavelengths and low frequencies, we can use the approximation

$$S_{\text{eff}}[\mathcal{A},\tilde{A}] \approx (\sigma^0_{xy}+\theta)\mathcal{S}_{\text{cs}}(\mathcal{A}_\mu) + \theta\mathcal{S}_{\text{cs}}(\tilde{A}_\mu) - \int d^3z\, \frac{\theta}{2}\epsilon_{\mu\nu\lambda}\mathcal{A}^\mu \tilde{F}^{\nu\lambda} \quad (10.12.6)$$

where only the statistical gauge field $\mathcal{A}_\mu$ is dynamical. This approximation is sufficient for our purposes. We will see below that, this approximate form of the effective action is sufficient to determine the charge and statistics of the quasiparticles as well as the Hall conductance.

10.13 The Hall Conductance in the Chern-Simons Theory

The electromagnetic response is calculated from the partition function $\mathcal{Z}[\tilde{A}]$

$$\begin{aligned}\mathcal{Z}[\tilde{A}] &= \int \mathcal{D}\mathcal{A}_\mu \exp(iS_{\text{eff}}[\mathcal{A},\tilde{A}]) \\ &= \exp\left(\frac{i}{2}\int d^3z \int d^3z'\, \tilde{A}_\mu(z)\, \Pi^{\mu\nu}_{\text{eff}}(z,z')\, \tilde{A}_\nu(z')\right)\end{aligned} \quad (10.13.1)$$

where $\Pi^{\mu\nu}_{\text{eff}}(z,z')$ is the effective polarization tensor (*i.e.*, the current-current correlation function for the full system) in the gaussian (RPA) approximation. The calculation is particularly simple in the infrared limit.

In chapter (7) we showed that a theory with two gauge fields $\tilde{A}_\mu$ and $\mathcal{A}_\mu$ with just Chern-Simons terms in the action, with couplings $\theta_1 = \theta$ and $\theta_2 = \sigma^0_{xy}$ respectively, is equivalent upon integration over $\mathcal{A}_\mu$, to a theory with

a Lagrangian $\mathcal{L}_{\text{eff}}[\tilde{A}] \equiv -i \log \mathcal{Z}[\tilde{A}]$ which has the Chern-Simons form

$$\mathcal{L}_{\text{eff}}[\tilde{A}] \approx \theta_{\text{eff}} \mathcal{L}_{\text{CS}}[\tilde{A}] \tag{10.13.2}$$

The effective coupling θ_{eff} is given by

$$\frac{1}{\theta_{\text{eff}}} = \frac{1}{\theta_1} + \frac{1}{\theta_2} \tag{10.13.3}$$

For the values $(1/\theta) = 2\pi(2s)$ and $\sigma^0_{xy} = (p/2\pi)$, that we found above, we get

$$\frac{1}{\theta_{\text{eff}}} = 2\pi(2s) + \frac{2\pi}{p} \tag{10.13.4}$$

Since in the effective Lagrangian $\mathcal{L}_{\text{eff}}[\tilde{A}]$ we are only keeping the terms with the smallest number of gradients, we are neglecting even the Maxwell terms, coming both from electrodynamics and from their renormalization by the charge fluctuations.

The (induced) current $J_\mu(x)$ is computed by its usual definition

$$J_\mu(x) = -i \frac{\delta \log \mathcal{Z}[\tilde{A}]}{\delta \tilde{A}_\mu(x)} \equiv \frac{\delta \mathcal{L}_{\text{eff}}[\tilde{A}]}{\delta \tilde{A}_\mu(x)} \tag{10.13.5}$$

The current $J_\mu(x)$ is determined by the Chern-Simons term alone

$$J_\mu(x) = \frac{\theta_{\text{eff}}}{2} \epsilon_{\mu\nu\lambda} \tilde{F}^{\nu\lambda}(x) \tag{10.13.6}$$

For a weak external static electric field $\tilde{E}_j(\vec{x})$, we find that the induced charge vanishes and that there is a non-zero Hall current *i.e.*,

$$\begin{aligned} \rho_{\text{ind}}(\vec{x}) &\equiv J_0(\vec{x}) = 0 \\ J^{\text{ind}}_k(\vec{x}) &\equiv \theta_{\text{eff}} \epsilon_{kj} \tilde{E}_j(\vec{x}) \end{aligned} \tag{10.13.7}$$

The form of the Hall current enables us to identify the Hall conductance σ_{xy} with θ_{eff}. Thus, the Hall conductance for this ground state is

$$\sigma_{xy} = \theta_{\text{eff}} = \frac{1}{2\pi} \left(\frac{p}{2sp+1} \right) \tag{10.13.8}$$

For the *odd* integers m, in the sequence $m = 2sp + 1$, we can write the Hall conductance as the *fraction*

$$\sigma_{xy} = \frac{1}{2\pi} \frac{p}{m} \left(\frac{e^2}{\hbar} \right) \tag{10.13.9}$$

where I have restored the factor $\frac{e^2}{\hbar}$. Hence, we get a *fractional* quantum Hall effect. The particular choice $p = 1$ yields the family of Laughlin states Ψ_m, with $m = 2s + 1$.

10.14 The Quantum Numbers of the Quasiparticles in the Chern-Simons Theory

Let us now evaluate the quantum numbers of the quasiparticles within the Chern-Simons theory. In particular, we want to compute their charge and statistics. Much of what follows is a rederivation, directly from the path integral, of results which were obtained before using Berry phase arguments. The path integral methods have the great advantage that they are very general and widely applicable.

We first need to identify the operators which create the quasiparticles in the Chern-Simons theory . Or, at least, we need to find a set of operators whose correlation functions yield information about the spectrum of the quasiparticles.

We have already identified the collective modes. Let us now identify the *quasihole*. From Laughlin's theory we know that the quasihole is an *anyon* which carries fractional charge.

We will now define a gauge invariant operator which creates an excitation at $\vec{x}$ at time x_0 and destroy it at $\vec{x}'$ at time x_0' which behaves like a quasihole. Let us consider the "bilinear"

$$\psi^\dagger(x) \exp\left(i \int_{\Gamma(x,x')} (A_\mu + \mathcal{A}_\mu) dx_\mu \right) \psi(x') \tag{10.14.1}$$

where $\Gamma(x,x')$ is a path in space-time going from x to x'. By construction, this operator is invariant under gauge transformations of the statistical gauge field $\mathcal{A}_\mu$. We will assume for the moment that the fluctuating component $\tilde{A}_\mu$ of the electromagnetic field is switched off and, therefore, this object only feels the uniform magnetic field A_μ (as far as electromagnetism is concerned). In any event, the line integral in the exponent of the bilinear, only depends on the sum of all the vector potentials. According to the procedure we used above, the fields $\tilde{A}_\mu$ and λ have already been shifted away and do not appear explicitly in this operator. Their effect is felt through their coupling to the the vector potential $\mathcal{A}_\mu$.

Let us evaluate the Green function $G_\Gamma(x,x')$ defined by

$$G_\Gamma(x,x') = \langle 0_m |\, T \left[\psi^\dagger(x) \exp\left(i \int_{\Gamma(x,x')} (A_\mu + \mathcal{A}_\mu)\, dx_\mu \right) \psi(x') \right] |0_m\rangle \tag{10.14.2}$$

where T is the time ordering operator. This Green function is gauge-invariant but depends on the choice of path Γ.

In path integral language, this Green function is given by an average over the histories of Fermi and statistical fields, weighed with the amplitude $\exp(iS_\theta)$ defined earlier in this section. We now proceed to integrate out the

Fermi fields, and find that the Green function is given by the average

$$G_\Gamma(x,x') = \langle\, G(x,x'|\{A_\mu + \mathcal{A}_\mu\})\, \exp\left(i\int_{\Gamma(x,x')} (A_\mu + \mathcal{A}_\mu)\, dx_\mu\right)\rangle_{\mathcal{A}} \tag{10.14.3}$$

The function $G(x,x'|\{A_\mu+\mathcal{A}_\mu\})$ is the one-particle Green function for a problem of fermions in *fixed* statistical and electromagnetic gauge fields at finite particle density, determined by the chemical potential μ. It is straightforward to see that $G(x,x'|\{A_\mu+\mathcal{A}_\mu\})$ is the inverse of the Schrodinger operator *i.e.*,

$$G(x,x'|\{A_\mu+\mathcal{A}_\mu\}) = \langle x|\frac{1}{iD_0+\mu+\lambda+\frac{1}{2m}\vec{D}^2}|x'\rangle \tag{10.14.4}$$

From now on we will not write down explicitly in our formulas the constant part of the electromagnetic field, A_μ. Its presence will be assumed throughout the rest of the discussion.

The average of any operator $\mathcal{O}[\{\mathcal{A}\}]$ over all configurations of the fields $\mathcal{A}_\mu$ is given by the path integral

$$\langle\mathcal{O}[\{\mathcal{A}\}]\rangle = \frac{1}{\mathcal{Z}}\int \mathcal{D}\mathcal{A}_\mu \mathcal{O}[\{\mathcal{A}\}]\exp(i\mathcal{S}_{\rm eff}[\mathcal{A}]) \tag{10.14.5}$$

where $\mathcal{Z}$ is the partition function and $\mathcal{S}_{\rm eff}[\mathcal{A}]$ is the effective action, which turns out to be given by

$$\begin{aligned} S_{\rm eff} = &-i\ {\rm Tr}\ \log[iD_0+\mu+\lambda+\frac{1}{2m}\vec{D}^2] + \theta S_{\rm CS}(\mathcal{A}_\mu) + \\ &-\int d^3z\ \lambda(z)\bar{\rho} + \frac{1}{2}\int d^3z \int d^3z'\ \lambda(z)\ V^{-1}(z-z')\ \lambda(z'). \end{aligned} \tag{10.14.6}$$

Let us now represent the one-particle Green function $G(x,x'|\{\mathcal{A}_\mu\})$ in terms of a Feynman path integral [Feynman 65] [Polyakov 86].We first use the representation of the Green function as an integral of a transition matrix element, namely

$$G(x,x'|\{\mathcal{A}_\mu\}) = -i\int_0^{+\infty} dT\ \langle\vec{x},0|\vec{x}',T\rangle\ e^{i\mu T} \tag{10.14.7}$$

where the weight $\exp(i\mu T)$ serves to fix the number of particles. Since the saddle point has p filled Landau levels, the chemical potential has to be set to lie between the levels p and $p+1$. The matrix element $\langle\vec{x},0|\vec{x}',T\rangle$ can be written as a sum over histories by means of the Feynman formula

$$\langle\vec{x},0|\vec{x}',\tau\rangle = \int \mathcal{D}\vec{z}[t]\ e^{iS[\vec{z}(t)]}. \tag{10.14.8}$$

with the boundary conditions

$$\begin{aligned} \lim_{t\to 0}\vec{z}(t) &= \vec{x} \\ \lim_{t\to T}\vec{z}(t) &= \vec{x}' \end{aligned} \tag{10.14.9}$$

Thus, as usual, the matrix element $\langle\vec{x},0|\vec{x}',T\rangle$ is a sum over all paths $\tilde{\Gamma}$ which go from $\vec{x}$ to $\vec{x}'$ in time T. The action S in the path integral is the standard

action for non-relativistic quantum mechanics for particles coupled to a gauge field

$$S = \int_0^\tau dt \left\{ \frac{m}{2} \left(\frac{d\vec{z}}{dt} \right)^2 + \frac{e}{c} \frac{dz^\mu}{dt}(t) A_\mu(\vec{z}(t)) \right\} \tag{10.14.10}$$

where we have used the notation $z_0 \equiv t$. The second term in the integrand is a shorthand notation for the coupling to the electromagnetic and statistical gauge fields,

$$\begin{aligned} \frac{d\vec{z}}{dt}(t) A_\mu(\vec{z}(t)) \equiv & \frac{e}{c} \frac{d\vec{z}}{dt}(t) \cdot \vec{A}(\vec{z}(t)) + eA_0(\vec{z}(t)) + \\ & + \frac{d\vec{z}}{dt}(t) \cdot \vec{\mathcal{A}}(\vec{z}(t)) + \mathcal{A}_0(\vec{z}(t)) + \lambda(\vec{z}(t)) \end{aligned} \tag{10.14.11}$$

Similar looking formulas can be derived for the two-particle and many-particle propagators.

For a problem with an energy gap, the long-distance, long-time limit, $|x - x'| \to \infty$, of the path integral is dominated by paths close to the solution of the classical equations of motion. Thus, in this case, the dominant trajectories are smooth. Thus, it should be a good approximation for our problem to pull the integral over the trajectories $\{\vec{z}(t)\}$ outside of the functional integral over the statistical gauge fields and to over all the configurations of these fields for a fixed path γ. The averaging over the trajectories of the particle is done at a later stage. We should keep in mind that these averages are performed around the saddle-point configuration which has an effective constant uniform magnetic field B_{eff} and a total number p of Landau levels completely filled. Formally, we can write the average in the form

$$\begin{aligned} G_\Gamma(x, x') = \int_0^{+\infty} dT \int \mathcal{D}\vec{z}(t)\, e^{i\mu T} \exp\left(i \int_0^T dt \frac{m}{2} (\frac{d\vec{z}}{dt})^2 \right) \times \\ \times \langle \exp(i \oint_\gamma A^\mu dz_\mu) \rangle_{\mathcal{A}} \end{aligned} \tag{10.14.12}$$

where the set of *closed curves* $\{\gamma\}$, represents paths which are the oriented sum of the path Γ and the histories of the particle $\tilde{\Gamma}$. It is important to keep in mind that this formula is a sum over all trajectories that go from $\vec{x}$ to $\vec{x}'$ with a *fixed* return path Γ. Notice that the particle does not return to $\vec{x}$, only the gauge fields see the *closed* paths γ.

It is straightforward to find a generalization of this formalism for the calculation of the two-particle Green function. The main difference is that, for the two particle case, there are two sets of trajectories to be summed over. The Grassman integral automatically antisymmetrizes the two particle Green function, which comes in the form of a sum over direct and exchange processes with the gauge fields as a fixed background.

In the semiclassical approximation, the exact average is replaced by an expansion around the solutions of the classical equations of motion. Thus, in this approximation, the particle only feels the average of the sum of the electromagnetic and statistical gauge field. The effective field felt by the particle

is equal to $B_{\rm eff} = B - (\rho/\theta)$. Thus, for each closed trajectory γ, there is a constant factor which can be factored out from the funtional integral. This factor corresponds to an Aharonov-Bohm (AB) phase factor for a particle moving in the field $B_{\rm eff}$, not in the external field B. It is easy to show that, as a result of the screening of the external magnetic field, the AB phase factor is that of a particle of charge $\frac{1}{m}$ of the electron charge moving in the unscreened field B.

Indeed, the exponent of AB phase factor is equal to $\frac{2\pi}{\phi_0} B_{\rm eff} A_\perp(\gamma)$, where $A_\perp(\gamma)$ is the (spacial) cross sectional area bounded by the path γ. Since $B_{\rm eff} = B - (\rho/\theta)$, we can define the *effective charge* (in units of e) $q_{\rm eff} \equiv 1 - (\rho/\theta B)$ and write $B_{\rm eff} = q_{\rm eff} B$. The effective charge $q_{\rm eff}$ can also be written in the more useful form

$$q_{\rm eff} = 1 - \frac{\rho}{\theta B} = 1 - \frac{\rho L^2}{\theta B L^2} = 1 - \frac{N}{2\pi\theta N_\phi} \tag{10.14.13}$$

where L is the linear size of the system. Thus, we get

$$q_{\rm eff} = 1 - \frac{\nu}{2\pi\theta} \tag{10.14.14}$$

For a filling fraction $\nu = (p/m) \equiv (p/2sp+1)$ and $\theta = 1/4\pi s$, we find that the effective charge is

$$q_{\rm eff} = 1 - \frac{2sp}{2sp+1} = \frac{1}{2sp+1} \equiv \frac{1}{m} \tag{10.14.15}$$

Hence, the effective charge is $\pm(e/m)$.

The fractional statistics can be studied by considering the two-particle Green function. Recall that now we have to consider two sets of trajectories, one for each particle. We now consider two paths γ_1 and γ_2, such as the ones discussed in section (7.2). Here too, the configurations of paths can be classified according to their *linking number* $\nu_{\rm L}$. The weights of configurations with different linking numbers have different phase factors. Likewise, configurations of paths from direct and exchange processes also have different linking number. While the phase factors themselves depend on the trajectories, and thus on the arbitrarily chosen paths for the two particles, the *relative phase* only depends on the topological properties of the configurations of paths and it is determined entirely by the relative linking number $\Delta\nu_{\rm L}$. In particular we want to compare two paths which form a linked knot with two paths which do not. In this case, the the linking number changes by $\Delta\nu_{\rm L} = 1$.

If the paths are very long and wide, such as the dominant paths for the low energy excitations, the average over the statistical gauge fields can be calculated using the effective action in the infrared approximation. This effective action only contains one Chern-Simons term (if $\tilde{A} = 0$) with coupling constant $\bar{\theta}$ equal to

$$\bar{\theta} = \sigma^0_{xy} + \theta = \frac{p}{2\pi} + \frac{1}{4\pi s} \tag{10.14.16}$$

The arguments of section (7.2) show that these two amplitudes differ by a factor W_{ex}

$$W_{ex} = - \exp(i\frac{\Delta\nu_L}{2\theta}) \tag{10.14.17}$$

Thus the effective statistics (relative to bosons) is given by the angle δ

$$\delta = \pi[\frac{2\pi s}{2sp+1} + 1] \tag{10.14.18}$$

where we have included the effect of the minus sign. For the Laughlin sequence and for the first level of the hierarchy, we got the filling fraction $\nu = (n/m)$ with $n = p$ and $m = 2sp + 1$. Hence, in this case the effective statistics (relative to bosons) is $\delta = \pm(\pi/m)$

We conclude that the operator that we found creates quasiholes (or conversely, quasielectrons) of charge $\pm(e/m)$ and statistics (π/m). This result agrees with the Berry phase arguments of Arovas, Schrieffer and Wilczek and with the calculations based on the Laughlin wave function, which we summarized in section (10.4). The power of the derivation that we just gave lies in the fact that it follows directly from the general principles of quantum mechanics (just as the Berry phase arguments do) but without having to make any specific ansatz for the wave functions for the ground state and for the quasihole. The adiabatic approximation, which is essential to the Berry phase argument, is just as important here since it results from the existence of a non-vanishing energy gap. But the general formula for the path integral is valid even in the absence of a gap.

References

[Arovas 84] D. Arovas, J.R. Schrieffer and F. Wilczek. Phys. Rev. Lett. 53, 722,(1984).

[Berry 84] M.V. Berry. Proc. Roy. Soc. London,Ser.A 392, 45,(1984).

[Dirac 31] P.A.M. Dirac. Proc. Roy. Soc. London, Ser.A 133, 60,(1931).

[Feenberg 69] E. Feenberg. Theory of Quantum Fluids, Pure and Applied Physics Series 31, 166,(1969).

[Feynman 65] R.P. Feynman and A. Hibbs. *Quantum Mechanics and Path Integrals*,McGraw-Hill, New York,1965.

[Girvin 84] S. Girvin and T. Jach. Phys. Rev. B29, 5617,(1984).

[Girvin 86] S. Girvin,A. MacDonald and P. Platzman. Phys. Rev. B33, 2481,(1986).

[Girvin 87] S. Girvin and A. MacDonald. Phys. Rev. Lett. 58, 1252,(1987).

[Haldane 83] F.D.M. Haldane. Phys. Rev. Lett. 51, 605,(1983).

[Halperin 82] B.I. Halperin. Phys. Rev. B. 25, 2185,(1982).

[Halperin 83] B.I. Halperin. Helv. Phys. Acta 56, 75,(1983).

[Halperin 84] B.I. Halperin. Phys. Rev. Lett. 52, 1583,(1984).

[Jackiw 76] R. Jackiw and C. Rebbi. Phys. Rev. D 13,3398,(1976).

[Jain 89] J. Jain. Phys. Rev. B40, 8079,(1989).

[Kivelson 85] S.Kivelson and M.Rocek. Phys. Lett. 156B, 85 (1985).

[Kivelson 86] S. Kivelson, C. Kallin, D. Arovas and J.R. Schrieffer. Phys. Rev. Lett.56, 873,(1986).

[Laughlin 83] R.B. Laughlin. Phys. Rev. Lett. 50, 1395, (1983).

[Laughlin 87] R.B. Laughlin. In *"The Quantum Hall Effect"*.Edited by R. Prange and S. Girvin, Springer-Verlag, New York, 1987.

[Polyakov 86] A. M. Polyakov. it Gauge Fields and Strings,Harwood, London, 1986.

[Prange 87] R.Prange and S.Girvin, Editors. *The Quantum Hall Effect* Springer-Verlag, New York, 1987

[Razayi 88] E. Rezayi and F.D.M. Haldane. Phys. Rev. Lett. 61, 1985,(1988).

[Read 89] N. Read. Phys. Rev. Lett. 62, 86,(1989).

[Simon 83] B. Simon. Phys. Rev. Lett. 51, 2167,(1983).

[Stell 64] G. Stell. in *Equilibrium Theory of Classical Fluids*edited by H. Frisch and J. Lebowitz, Benjamin, New York, 1964.

[Su 81] W.P. Su and J.R. Schrieffer. Phys. Rev.Lett. 46, 738,(1981).

[Trugman 85] S. Trugman and S. Kivelson. Phys. Rev.B 31, 5280,(1985).

[Zhang 89] S.C. Zhang, T. Hansson and S. Kivelson. Phys. Rev. Lett. 62, 82,(1989).

Index