Jeff Greensite

An Introduction to the Confinement Problem

Lecture Notes in Physics

Volume 821

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching—quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research and to serve three purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas
- to be a source of advanced teaching material for specialized seminars, courses and schools

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive being available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic net-works, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg / Germany
christian.caron@springer.com

Jeff Greensite

An Introduction to the Confinement Problem

Jeff Greensite
Department of Physics and Astronomy
San Francisco State University
San Francisco, CA 94132
USA
e-mail: greensit@stars.sfsu.edu

*For my mother, Carol Greensite, and to
the memory of my father, Arthur L. Greensite*

Preface

This contribution to the Springer Lecture Notes in Physics series is intended to provide an overview of the confinement problem in non-abelian gauge theory, with a particular emphasis on the relevance of center symmetry and the lattice formulation, and an introduction to the current research. The book is an expanded and updated version of lectures and review talks on the confinement problem that I have presented over the years, particularly at the Schladming Winter School in 2005, and the Cracow School of Theoretical Physics in 2009.

It is a pleasure to thank Maarten Golterman and Štefan Olejník for volunteering to read a first draft of this manuscript, and for their many helpful suggestions.

San Francisco, April 2010   Jeff Greensite

Contents

Chapter 1
Introduction

The standard model of particle physics, put forward almost forty years ago, has been so successful at explaining observations that there are few experimental hints, at least from traditional sources such as accelerator centers, that anything more is needed. Perhaps this situation will change as new data emerges from the Large Hadron Collider at CERN; it is, after all, unlikely that the standard model is the end of the story. The Higgs sector seems a little unnatural, among other things. There is also a strong CP problem, there are many free parameters, there is no explanation of dark matter (let alone dark energy) or three fermion generations, and the theory does not accommodate gravity. A minor adjustment is also required to allow for neutrino masses. On the other hand, given this adjustment, there is little reason to doubt the validity of the standard model as a description of all the known elementary particles, and their interactions, down to a distance scale of at least 10^{-15} cm, and there is good reason to admire a theory which can describe such a wide range of phenomena so economically.

Because of the great importance of the standard model, and the central role it plays in our understanding of particle physics, it is unfortunate that, in one very important respect, we don't really understand how it works. The problem lies in the sector dealing with the interactions of quarks and gluons, the sector known as Quantum Chromodynamics or QCD. We simply do not know for sure why quarks and gluons, which are the fundamental fields of the theory, don't show up in the actual spectrum of the theory, as asymptotic particle states. There is wide agreement about what must be happening in high energy particle collisions: the formation of color electric flux tubes among quarks and antiquarks, and the eventual fragmentation of those flux tubes into mesons and baryons, rather than free quarks and gluons. But there is no general agreement about *why* this is happening, and that limitation exposes our general ignorance about the workings of non-abelian gauge theories in general, and QCD in particular, at large distance scales.

This book concerns the confinement problem, which deals with the behavior of gauge theories, and the force which is mediated by gauge fields, at large distances.

Historically, the word "confinement" in the context of hadronic physics originally referred to the fact that quarks and gluons appear to be trapped inside mesons and baryons, from which they cannot escape. There are other, and possibly deeper meanings that can be attached to the term, and these will be explored in due course. Although the confinement problem is far from solved, much is now known about the general features of the confining force, and there are a number of very well motivated (although not yet compelling) theories of confinement which are under active investigation. In this volume I will try to give an overview of the main ideas in this field, their attractive features, and, as appropriate, their shortcomings.

Chapter 2
Global Symmetry, Local Symmetry, and the Lattice

Confinement in non-abelian gauge theory involves the idea that the vacuum state is disordered at large scales; our best evidence that this is true comes from Monte Carlo simulations of lattice gauge theories. So to begin with, I need to explain what is meant by

- a disordered state,
- a lattice gauge theory,
- a Monte Carlo simulation.

I also need to explain, since this is a book about the confinement problem, what is meant by the word "confinement." That, however, is a surprisingly subtle, and even controversial issue, which will be deferred to the next chapter.

2.1 Global Symmetry and the Ising Model

Let us begin with the concept of a disordered state, as it appears in the Ising model of ferromagnetism. We know that certain materials (e.g., iron) can be magnetized at low temperatures, but above a certain critical temperature, known as the "*Curie Temperature*," the magnetic moment disappears, at least in the absence of an external field. The tendency to retain a magnetic moment at low temperature is due to an interaction between neighboring atoms in the solid, whose potential energy is lowered if the magnetic moments of the neighboring atoms are aligned. The Ising model is a simplified picture of this situation. We imagine that the solid is a cubic array of atoms, and that each atom can be in one of two physical states: "spin up" or "spin down," with the magnetic moment oriented in the direction of the spin. The Hamiltonian of the Ising model is

$$H = -J \sum_x \sum_{\mu=1}^{D} s(x)s(x+\hat{\mu}), \qquad (2.1)$$

where $s(x) = 1$ represents an atom at point x with spin up, $s(x) = -1$ represents spin down, and J is a positive constant.

At low temperatures, in any dimension $D > 1$, the system tends to be in an "*ordered state*," meaning that most spins tend to point in the same direction. The magnetization (average $s(x)$) is non-zero. At high temperatures the system is in a "*disordered state*," in which the average spin is zero, corresponding to a vanishing magnetization. According to the usual principles of statistical mechanics, the probability of any given spin configuration $\{s(x)\}$ at a particular temperature T is

$$\text{Prob}[\{s(x)\}] = \frac{1}{Z} \exp\left[-\frac{H}{kT}\right] \tag{2.2}$$

and

$$Z = \sum_{\{s(x)\}} \exp\left[\beta \sum_x \sum_{\mu=1}^{D} s(x)s(x + \hat{\mu})\right] \tag{2.3}$$

where $\beta \equiv J/kT$. Now observe that the Hamiltonian $H[\{s(x)\}]$, and the probability distribution $\text{Prob}[\{s(x)\}]$, are left unchanged by the transformation of each spin by

$$s(x) \to s'(x) = zs(x) \quad \text{where } z = \pm 1. \tag{2.4}$$

Although the transformation with $z = +1$ does nothing to the spins, we include it because the two transformations $\{1, \ -1\}$ together form a group, known as Z_2. The operation (2.4) is called a "*global*" transformation, because every spin $s(x)$ at every location x is transformed in the same way, by the same factor z.

Now it is obvious that the probability distribution is invariant with respect to the global Z_2 transformations, i.e., $\text{Prob}[\{s(x)\}] = \text{Prob}[\{zs(x)\}]$, and therefore the average spin

$$\langle s \rangle = \sum_{\{s(x)\}} \frac{1}{N_{spins}} \left(\sum_{x'} s(x')\right) \text{Prob}[\{s(x)\}] \tag{2.5}$$

must equal zero. After all, to any given spin configuration $\{s(x)\}$ contributing to the sum, with some average spin s_{av}, there is another configuration with spins $\{-s(x)\}$, with average spin $-s_{av}$, which contributes with exactly the same probability. From this argument, it would appear that magnets are impossible. And in a sense that's true . . . *permanent* magnets, permanent at a finite temperature for infinite time, *are* impossible! Suppose spins are aligned, at some low temperature, with $s_{av} > 0$. There will be small thermal fluctuations, and the s_{av} will vary a little from configuration to configuration, but in general, at low temperatures, s_{av} will be positive for a very long time. A thermal fluctuation which would flip enough spins so that s_{av} becomes negative is very unlikely. Nevertheless, providing the number of spins N_{spins} is finite, and we wait long enough, at some point one of these vastly unlikely fluctuations will occur, and then s_{av} will be negative, again for a very,

very long, yet finite time. Averaged over sufficiently long time scales, the mean magnetization is zero. But the time between such large fluctuations grows exponentially with the number of spins, and for real ferromagnets of macroscopic size the time between flipping the overall magnetic moment, just by thermal fluctuations, would certainly exceed the age of the universe.

So using (2.5) gives a result which is formally correct at low temperatures, yet wrong for all "practical" purposes. It is therefore useful to introduce an external magnetic field h

$$H_h = -J \sum_x \sum_{\mu=1}^{D} s(x)s(x+\hat{\mu}) - h \sum_x s(x)$$
$$Z_h = \sum_{\{s(x)\}} \exp[-H_h/kT] \tag{2.6}$$

so that $\langle s \rangle \neq 0$ at any temperature, and then consider what happens in the pair of limits where we take first the volume (number of spins) infinite, and then reduce h to zero

$$m = \lim_{h \to 0} \lim_{N_{spins} \to \infty} \frac{1}{Z_h} \sum_{\{s(x)\}} \left(\frac{1}{N_{spins}} \sum_{x'} s(x') \right)$$
$$\times \exp\left[\beta \sum_{x,\mu} s(x)s(x+\hat{\mu}) + \frac{h}{kT} \sum_x s(x) \right]. \tag{2.7}$$

In this pair of limits, done in the order shown, it is possible that $m = \langle s \rangle \neq 0$. When this is the case, we say that the Z_2 global symmetry is "*spontaneously broken.*" The term means that despite the invariance of the Hamiltonian, an observable (such as average magnetization) which is not invariant under the symmetry can nevertheless come out with a non-zero expectation value. At high temperatures, even in the limits shown, the magnetization m vanishes. In that case we say that the Z_2 symmetry is unbroken, and the spin system is in the symmetric phase. In general, in the unbroken symmetry phase, the symmetry of the Hamiltonian is reflected in the expectation values: the expectation value vanishes for any quantity which is not invariant with respect to the symmetry group. In the broken phase, non-invariant observables can have non-zero expectation values in the appropriate infinite volume limit.

We also say that the low-temperature, broken phase is an "*ordered*" phase, while the high-temperature, symmetric phase is the "*disordered*" phase. In the case of the Ising model, the term "order" obviously refers to the fact that the spins are not randomly oriented, but point, on average, in one of the two possible directions. This is not true in the disordered phase, although if a given spin in a typical configuration is pointing up, its immediate neighbors are more likely to point up than down. This correlation, however, falls off exponentially with distance between spins. The quantitative measure is provided by the correlation function of the spins at two sites,

loosely denoted "0" and "R," which are a distance R apart in units of the inter-atomic spacing. The correlation function is defined as

$$G(R) = \langle s(0)s(R) \rangle$$
$$= \frac{1}{Z} \sum_{\{s(x)\}} s(0)s(R) \exp\left[\beta \sum_{x,\mu} s(x)s(x+\hat{\mu})\right]. \tag{2.8}$$

In the disordered phase $G(R) \sim \exp[-R/l]$, where length l is known as the "*correlation length*." In the ordered phase, $G(R) \rightarrow m^2$ as $R \rightarrow \infty$. It turns out that in $D = 1$ dimension, a spin system is in the disordered phase at *any* finite temperature; there is no phase transition between an ordered and a disordered phase, i.e., no non-zero Curie temperature. The symmetry-breaking transition makes its appearance for any dimension greater than one.

The existence of (at least) two phases, broken and unbroken, ordered and disordered, is characteristic of the statistical mechanics of most many body Hamiltonians that are invariant under some global symmetry.

2.2 Gauge Invariance: the Unbreakable Symmetry

The spin system which was just described is an example of a lattice field theory. The number of dynamical degrees of freedom (spins, in the Ising model) is discrete; each field component interacts with only a few neighbors, and they are usually (but not always) arranged on a square, cubic, or hypercubic lattice, in $D = 2, 3, 4$ dimensions, respectively, as illustrated in Fig. 2.1 for $D = 3$ dimensions. The points of the lattice are known as *sites*, the lines joining neighboring sites are *links*, and, on a hypercubic lattice, the little squares joined by four connecting links are known as *plaquettes*.

A *gauge transformation* is a position-dependent transformation of the degrees of freedom, in which the transformation can be chosen independently at each site.

Fig. 2.1 A lattice is a discretization of spacetime. Gauge fields are associated with the links between lattice sites
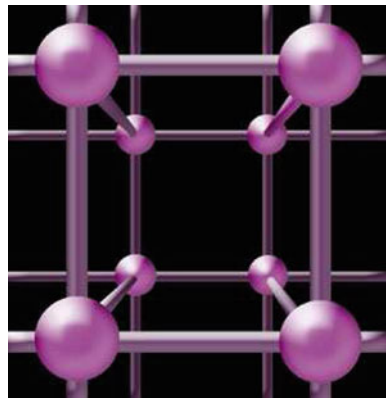
Fig. 2.2 A plaquette is the smallest loop on the lattice. The gauge field action is proportional to a product of link variables around each plaquette, summed over all plaquettes
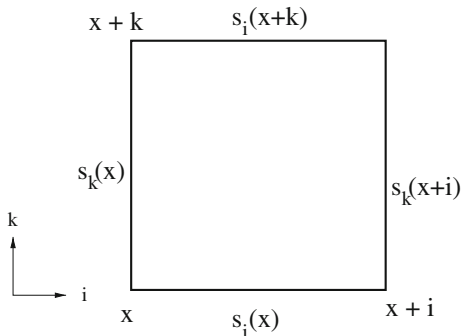
The trick is to associate the dynamical degrees of freedom with the links of the lattice, while the gauge transformation is specified at each site. A link can be uniquely specified by a lattice site x and direction μ; the link then runs between sites x and $x + \hat{\mu}$. The Hamiltonian for the gauge-invariant Ising model, also known as Z_2 lattice gauge theory, is

$$H = -J \sum_x \sum_{\mu=1}^{D-1} \sum_{v > \mu}^{D} s_\mu(x) s_v(x + \hat{\mu}) s_\mu(x + \hat{v}) s_v(x). \tag{2.9}$$

The sum is over all plaquettes (Fig. 2.2), and the quantity which is summed is the product of spin variables on links around each plaquette. This Hamiltonian is invariant under the gauge transformation of each link variable $s_\mu(x)$ by

$$s_\mu(x) \rightarrow z(x) s_\mu(x) z(x + \hat{\mu}), \tag{2.10}$$

where the $z(x) = \pm 1$ can be chosen independently at each site; i.e., gauge transformations are *local* transformations, and gauge symmetry is a local symmetry. The local Z_2 gauge symmetry is vastly larger than the global Z_2 symmetry of the Ising model, and this huge expansion of the symmetry can be shown to have the following consequence [1]:

Elitzur's Theorem *A local gauge symmetry cannot break spontaneously. The expectation value of any gauge non-invariant local observable must vanish.*

This means there is no analog of magnetization; e.g., the expectation value of a spin on a link $\langle s_\mu(x) \rangle$ is zero, even if we introduce an external field h (which explicitly breaks the gauge symmetry) and then carefully take first the infinite volume, and then the $h \rightarrow 0$ limits. We must look, instead, to gauge-invariant observables, which are unaffected by gauge transformations. These can be constructed by taking the product of spins on links around a closed loop C, known as a "*Wilson loop*"[1]

---

[1] The construction was first introduced by Wegner [2].

$$W(C) = \left\langle \prod_{(x,\mu)\in C} s_\mu(x) \right\rangle. \tag{2.11}$$

The product of link variables around a plaquette, found in the gauge-invariant Hamiltonian (2.9), is just a particular example of a Wilson loop.

We can generalize the Z_2 construction to any symmetry group G, just by choosing link variables which are elements of G. Choosing G to be $SU(3)$, or $SU(2) \times U(1)$, we obtain the action of the gauge sector of the strong or electroweak interactions, in a discretized spacetime which allows for numerical simulations. A simple example is $G = U(1)$, and the corresponding gauge theory is known as *compact QED*. In this example the link variables are elements of the $U(1)$ group

$$U_\mu(x) = \exp[iaeA_\mu(x)], \quad A_\mu(x) \in \left[-\frac{\pi}{ae}, \frac{\pi}{ae}\right], \tag{2.12}$$

where a is the lattice spacing, i.e., the distance (in meters, fermis, parsecs, or whatever) between neighboring sites on the lattice, and e is the unit of electric charge. The probability distribution of lattice gauge field configurations is

$$\text{Prob}[\{U_\mu(x)\}] = \frac{1}{Z} e^{-S[U]}, \tag{2.13}$$

where $S[U]$ is known as the "Euclidean action" (to be explained below), and is given by

$$S[U] = -\frac{\beta}{2} \sum_{x,\mu<\nu} U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu^*(x+\hat{\nu}) U_\nu^*(x) + \text{c.c.} \tag{2.14}$$

It is customary to associate the link variable $U_\mu(x)$ with a line running from site x to site $x + \hat{\mu}$ in the positive μ-direction, and the complex conjugate $U_\mu^*(x)$ (or, for matrix-valued variables, the Hermitian conjugate $U_\mu^\dagger(x)$), with a line running from site $x + \hat{\mu}$ to x in the negative μ-direction, as shown in Fig. 2.3. In that case the product of link variables shown in (2.14), and its complex conjugate, can be visualized as Wilson loops running along the same plaquette in opposite directions. The Euclidean action $S[U]$ is invariant under local gauge transformations of the form

$$U_\mu(x) \rightarrow e^{i\theta(x)} U_\mu(x) e^{-i\theta(x+\hat{\mu})}, \tag{2.15}$$

where the set of transformations $g(x) = \exp[i\theta(x)]$ belongs to the compact $U(1)$ gauge group. To see why this theory is called compact QED, make the Taylor expansion

$$U_\mu(x) = 1 + iaeA_\mu(x) - \frac{1}{2}a^2e^2A_\mu^2(x) + \cdots \tag{2.16}$$
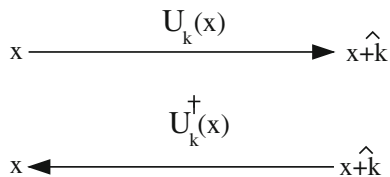
Fig. 2.3 A gauge field variable $U_\mu(x)$ is represented graphically by a directed line joining points x and $x + \hat{\mu}$ in the positive μ-direction. The hermitian conjugate $U_\mu^\dagger(x)$ is represented by a line joining the same points, but in the opposite direction

Keeping terms in the action to lowest order in lattice spacing (and dropping an irrelevant constant)

$$S = \frac{\beta}{2} \sum_x a^4 e^2 \sum_{\mu < \nu} \left[\frac{A_\nu(x + \hat{\mu}) - A_\nu(x)}{a} - \frac{A_\mu(x + \hat{\nu}) - A_\mu(x)}{a} \right]^2. \tag{2.17}$$

Identifying $\beta = 1/e^2$, we obtain in the limit $a \to 0$, known as the "*continuum limit*,"

$$S = \int d^4x \frac{1}{4} F_{\mu\nu} F_{\mu\nu}, \tag{2.18}$$

where $F_{\mu\nu}$ is the standard field strength of electromagnetism. However, in the action S the repeated indices are simply summed over, rather than contracted with a Minkowski metric. That is why S is termed a Euclidean action. It is obtained from the usual action of electromagnetism by rotating the time axis in the complex plane $t \to it$ (it is also necessary to rotate the time component of the vector potential, i.e., $A_0 \to -iA_0$). This is called a "*Wick rotation*," and it essentially takes the metric from Minkowski to Euclidean space. In the continuum limit, the gauge transformation (2.15) is equivalent to the familiar form

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e} \partial_\mu \theta(x). \tag{2.19}$$

These constructions are readily generalized to non-abelian groups of interest, $SU(2)$ and $SU(3)$ in particular. For example, $SU(2)$ is the group of 2×2 unitary matrices of determinant one. We let the link variables be elements of this group

$$\begin{aligned} U_\mu(x) &= e^{iagA_\mu(x)} \\ &= b_0 \mathbb{1}_2 + i\mathbf{b} \cdot \boldsymbol{\sigma}, \quad b_0^2 + \mathbf{b} \cdot \mathbf{b} = 1, \end{aligned} \tag{2.20}$$

where $A_\mu(x)$ is a 2×2 Hermitian matrix

$$A_\mu(x) = A_\mu^a(x)\frac{\sigma_a}{2}, \tag{2.21}$$

and the $\{\sigma_a, a = 1, 2, 3\}$ are the three Pauli spin matrices. The Euclidean action

$$S = -\frac{\beta}{2}\sum_{x,\mu<\nu}\text{Tr}[U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)] \tag{2.22}$$

is invariant under $SU(2)$ gauge transformations

$$U_\mu(x) \rightarrow G(x)U_\mu(x)G^\dagger(x+\hat{\mu}), \quad G \in SU(2). \tag{2.23}$$

$SU(2)$ is a bit special, in the sense that the trace of any group element is real. The generalization to an $SU(N)$ gauge theory is simply to take the real part of the trace, i.e.,

$$S = -\frac{\beta}{2N}\sum_{x,\mu<\nu}\left\{\text{Tr}[U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)] + \text{c.c.}\right\}. \tag{2.24}$$

This is known as the *Wilson action*.

Once again, if we expand $U_\mu(x)$ in powers of $A_\mu(x)$, let $\beta = 2N/g^2$, and keep only terms which survive in the $a \rightarrow 0$ limit, we obtain

$$S = \int d^4x \frac{1}{2}\text{Tr}[F_{\mu\nu}F_{\mu\nu}], \tag{2.25}$$

where this time $F_{\mu\nu}(x)$ is the non-abelian field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu], \tag{2.26}$$

and the lattice gauge transformation (2.23) goes over to the continuum gauge transformation

$$A_\mu(x) \rightarrow G(x)A_\mu(x)G^\dagger(x) - \frac{i}{g}G(x)\partial_\mu G^\dagger(x), \tag{2.27}$$

with $G(x)$ an element of the gauge group.

2.3 The Monte Carlo Method

I have already mentioned that the lattice action S is a Euclidean action; spacetime indices are contracted with a Kronecker delta, rather than a Minkowski space metric. This Euclidean space formulation has a great advantage: the field theory can be regarded as a statistical (rather than quantum) system, and this lends itself to a powerful method of numerical simulation known as the lattice Monte Carlo

method. Before describing that method, it is important to understand how a simulation in Euclidean space can yield results which are relevant to the corresponding quantum field theory in Minkowski space.

The connection between the Euclidean path integral, and the Minkowski space Hamiltonian operator H, is provided by the following relation: Suppose Q_t is some observable which depends on the fields at a fixed time t. Then

$$
\begin{aligned}
\langle Q_{t_2}^\dagger Q_{t_1} \rangle &= \frac{1}{Z} \int DA Q_{t_2}^\dagger Q_{t_1} e^{-S} \\
&= \langle \Psi_0 | Q^\dagger e^{-H(t_2 - t_1)} Q | \Psi_0 \rangle,
\end{aligned}
\tag{2.28}
$$

where the second line is an operator expression, with Ψ_0 the ground state of Hamiltonian H. With the help of this relation, it becomes possible to calculate properties of the Minkowski space theory, e.g., the low-lying spectrum, and the potential energy between static charges, from the lattice theory formulated in Euclidean space.

In a lattice Monte Carlo calculation, the idea is to replace the integral over all configurations $\{U_\mu(x)\}$, weighted by the probability distribution (2.13), by an average over a finite set of sample lattice configurations $\{U^{(n)}, n = 1, 2, \ldots, N_{conf}\}$, i.e.,

$$
\begin{aligned}
\langle Q \rangle &= \int DU Q[U] \frac{1}{Z} e^{-S[U]} \\
&\approx \frac{1}{N_{conf}} \sum_{n=1}^{N_{conf}} Q[U^{(n)}],
\end{aligned}
\tag{2.29}
$$

where the sample configurations are generated stochastically, with probability weighting $\text{Prob}\big[\{U_\mu(x)\}\big] \sim \exp[-S[U]]$.

There are various ways to generate configurations with the desired probability weighting. The simplest and most general is known as the "*Metropolis algorithm*," and the procedure is the following: Start with any convenient lattice configuration, e.g., all link variables set to the unit element. Sweep through the lattice link by link. At each link choose at random an element of the gauge group, denoted U', as a possible replacement for the existing link variable. Calculate the change ΔS in the action, if link variable $U_\mu(x)$ were replaced by U'. Then

1. If $\Delta S \leq 0$, set $U_\mu(x) = U'$.
2. Otherwise, for $\Delta S > 0$, use a standard random number generator to obtain a pseudo-random number x, in the interval $[0, 1]$ with uniform weighting. If $x < \exp[-\Delta S]$, set $U_\mu(x) = U'$. Otherwise, leave the link variable unchanged, and go on to the next link.

It can be shown that by iterating this procedure, one can generate as many lattice configurations as one likes with the desired probability weighting. Then any observable, such as a Wilson loop, can be averaged over the N_{conf} sample configurations to give an estimate of the expectation value. The statistical error

inherent in this procedure goes down as $1/\sqrt{N_{conf}}$. While the Metropolis algorithm is easy to explain and easy to implement, it is often not the most efficient method in practice. For the more sophisticated algorithms which are used in modern, large-scale lattice Monte Carlo calculations, the reader should consult, e.g., the text by DeGrand and DeTar [3].

We can use the expectation value of Wilson loops, computed in the Euclidean, statistical theory, to determine the interaction energy of two static, color-charged sources in the Minkowski space quantum theory. This interaction energy is known as the *static quark potential*, whether or not the sources are massive quarks. Consider, e.g., adding a massive scalar field to the gauge theory in D-dimensions, with action

$$S_{matter} = -\sum_{x,\mu}\big(\phi^\dagger(x)U_\mu(x)\phi(x+\hat\mu)+\text{c.c.}\big) + \sum_x(m^2+2D)\phi^\dagger(x)\phi(x). \quad (2.30)$$

This action is gauge invariant, given that the scalar field transforms as $\phi(x) \to G(x)\phi(x)$. Let Q_t be an operator which creates a gauge invariant particle–antiparticle state at time t

$$Q_t = \phi^\dagger(0,t)U_i(0,t)U_i(\hat\imath,t)U_i(2\hat\imath,t)\ldots U_i((R-1)\hat\imath,t)\phi(R\hat\imath,t). \quad (2.31)$$

For $m^2 \gg 1$ the one-link term in S_{matter} can be treated as a small perturbation. In this limit it is straightforward to integrate over ϕ in the functional integral, and we find, to leading order in $1/m^2$

$$\langle Q_T^\dagger Q_0\rangle = \text{const.} \times \langle\text{Tr}[UUUU\ldots UUU]_C\rangle, \quad (2.32)$$

where $[UUU\ldots U]_C$ represents an ordered product of link variables (either U or $U^\dagger$), running counterclockwise around a rectangular $R \times T$ contour C, with the convention shown in Fig. 2.3. The trace of this quantity is a gauge-invariant Wilson loop, with expectation value denoted $W(R, T)$. The brackets $\langle\ldots\rangle$ in this case denote the expectation value of the observable in the gauge theory with no matter field, and the constant is $O(m^{-4(T+1)})$.[2] On the other hand, in the operator formalism we also have

$$\begin{aligned}
\langle Q_T^\dagger Q_0\rangle &= \frac{\sum_{nm}\langle 0|Q^\dagger|n\rangle\langle n|e^{-HT}|m\rangle\langle m|Q|0\rangle}{\langle 0|e^{-HT}|0\rangle}\\
&= \sum_n |c_n|^2 e^{-\Delta E_n T} \quad\quad\quad\quad\quad (2.33)\\
&\sim e^{-\Delta E_{min}T} \quad \text{as } T \to \infty,
\end{aligned}$$

---

[2] This leading-order (in $1/m^2$) result is obtained by neglecting the one-link term in S_{matter} everywhere except along the time-like links from $t = 0$ to $t = T$, at $x = 0$ and $x = R$. On these links, expand $\exp[\phi^\dagger U_0\phi + \text{h.c.}] \approx 1 + \phi^\dagger U_0\phi + \text{h.c.}$. Integration over the scalar field then yields the result (2.32).

where the sum is over all energy eigenstates of the Hamiltonian containing two static charges at the given locations, and ΔE_n is the difference in energy of the nth excited state and the vacuum. The minimal energy state containing two static charges, with energy difference ΔE_{min}, dominates the expectation value at large T, and this energy difference is also (apart from R-independent self-energy terms) the interaction energy between the static color-charged sources. Therefore, the relationship between the static quark potential $V(R)$ and the expectation value of the rectangular Wilson loop is

$$W(R,T) \sim e^{-V(R)T} \quad (\text{as } T \to \infty). \tag{2.34}$$

2.4 Possible Phases of a Gauge Theory

Elitzur's theorem tells us that gauge symmetries do not break spontaneously, in the sense that any local, gauge non-invariant observable that could serve as an order parameter must have a vanishing expectation value. There is, nonetheless, a qualitative distinction between different phases of a gauge theory, based on the behavior of large Wilson loops. Let us again consider a large rectangular $R \times T$ loop, with $T \gg R$. There are three possibilities:

- **Massive phase**. The potential is of the Yukawa type

$$V(R) = -g^2 \frac{e^{-mR}}{R} + 2V_0, \tag{2.35}$$

where V_0 is a self-energy term, and the Wilson loop falls off at $R > 1/m$ as $\exp(-2V_0T)$. For a generic, non self-intersecting planar loop C, of extension $\gg 1/m$,

$$W(C) \sim \exp[-V_0 P(C)], \tag{2.36}$$

where $P(C)$ is the perimeter of loop C.
- **Massless phases**. The potential has the form [4, 5]

$$V(R) = -\frac{g^2(R)}{R} + 2V_0, \tag{2.37}$$

and the rectangular loop falls off as $\exp[-2V_0T + g^2(R)T/R]$. The effective coupling $g^2(R)$ can go to a constant (Coulomb phase), or, in gauge theories with massless matter fields, it may decrease logarithmically (free electric phase), or increase logarithmically (free magnetic phase), depending on the type of massless particles in the spectrum. For $R \gg V_0^{-1}$, the Wilson loop again has a perimeter-law falloff.
- **Magnetic disorder phase**. The potential is asymptotically linear

$$V(R) = \sigma R + 2V_0. \tag{2.38}$$

Then the Wilson loop falls off asymptotically as $\exp[-\sigma RT - 2V_0 T]$. More generally, for a planar non self-intersecting loop with minimal area $A(C)$ and perimeter $P(C)$,

$$W(C) \sim \exp[-\sigma A(C) - V_0 P(C)], \tag{2.39}$$

and there is an area-law type of falloff.

The term "magnetic disorder" is a terminology inspired by electromagnetism. Consider any lattice (or continuum) gauge theory, based on an abelian gauge group. Let C be a planar loop, whose minimal area $A(C)$ is subdivided into a set $\{C_i\}$ of n smaller loops, each of area A' as shown in Fig. 2.4. If the orientations of the loops are chosen so that neighboring contours run in opposite directions, it is easy to see that

$$U(C) = \prod_{i=1}^{n} U(C_i). \tag{2.40}$$

Now suppose that if the subareas $\{A_i'\}$ are large enough, the expectation value of the product of loops is equal to the product of expectation values, i.e.,

$$\left\langle \prod_{i=1}^{n} U(C_i) \right\rangle = \prod_{i=1}^{n} \langle U(C_i) \rangle. \tag{2.41}$$

In that case,

$$\begin{aligned} W(C) &= \langle U(C) \rangle \\ &= \exp[-\sigma A(C)], \end{aligned} \tag{2.42}$$

Fig. 2.4 Subdivison of the minimal area of a large loop C into smaller areas bounded by loops $\{C_i\}$. In an abelian gauge theory, the Wilson loop around C is the product of Wilson loops around the $\{C_i\}$
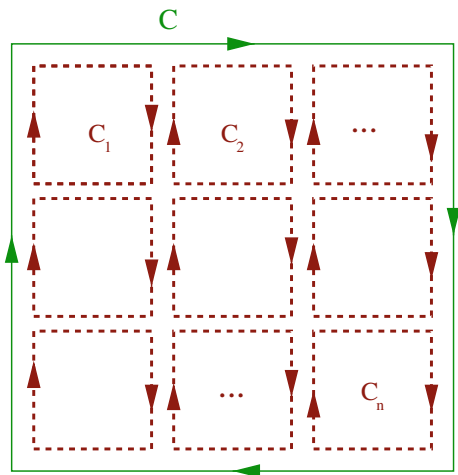
where

$$\sigma = -\frac{\log\langle U(C_i)\rangle}{A'}.$$

(2.43)

If we consider varying the size and number of the subloops C_i, it is clear that, so long as the factorization property (2.41) holds, consistency requires that the string tension σ in (2.43) is independent of the subloop area A'.

The relationship (2.40), for the $U(1)$ gauge group, is simply a consequence of Stokes Law, since, in the continuum limit, for a planar loop in, e.g., the ij plane

$$\begin{aligned} U(C) &= \exp\left[ie \oint_C \mathrm{d}x^k A_k(x)\right] \\ &= \exp\left[ie \int \mathrm{d}S_C F_{ij}\right], \end{aligned}$$

(2.44)

where the second integration runs over the minimal area of the loop. Then (2.40) follows from the fact that

$$\int \mathrm{d}S_C F_{ij} = \sum_n \int \mathrm{d}S_{C_n} F_{ij}.$$

(2.45)

If i, j are spacelike directions, then in $3 + 1$ dimensions the integrals over S_{C_n} just give the magnetic flux passing through the area of the loop C_n.

It is Stoke's Law, in the abelian theories, that tells us that the product of smaller Wilson loops equals the single larger Wilson loop (2.40). The statement that the expectation value of the product of smaller loops is equal to the product of expectation values in (2.41) goes beyond Stoke's Law, and translates into the statement that if the minimal area A' is large enough, then the magnetic fluxes through the loops in the set $\{C_i\}$ are completely uncorrelated. This is the property which leads to an area law falloff for Wilson loops, and which motivates the term "magnetic disorder."

For non-abelian gauge groups, the Stokes law is not so simple (it exists, in fact, in several inequivalent versions, cf. [6–9]), but still boils down to the fact that a large planar Wilson loop can be expressed as the product of Wilson loops over smaller subareas.[3]

In $D = 2$ dimensions it is easy to demonstrate that the only phase that exists is the magnetically disordered phase, for any lattice action of the form (2.24), and for any gauge group. Take the Z_2 gauge theory for simplicity, with

---

[3] One approach, based on the non-abelian Stokes Law, derives an area law for a large Wilson loop from an assumed finite range behavior of field strength correlators, which means that field strengths are uncorrelated, i.e. disordered, at sufficiently large separations. This "field correlator" approach to magnetic disorder has been pursued by Simonov and co-workers [10].

$$S = -\beta \sum_p s(p)$$

$$s(p) = s_1(x)s_2(x+\hat{1})s_1(x+\hat{2})s_2(x), \tag{2.46}$$

where the sum is over all plaquettes p defined on a rectangular lattice of N_p plaquettes, with free, rather than periodic boundary conditions.[4] We use the fact that, for $s(p)$ the product of links around a plaquette p,

$$\exp[\beta s(p)] = \cosh(\beta) + s(p)\sinh(\beta), \tag{2.47}$$

and

$$\sum_{s_\mu=\pm 1} s_\mu = 0, \quad \sum_{s_\mu=\pm 1} 1 = \sum_{s_\mu=\pm 1} s_\mu^2 = 2. \tag{2.48}$$

Then for a loop C enclosing a region S of area $A(C)$

$$W(C) = \frac{\sum_{s_\mu(x)} \left(\prod_{(x',\mu')\in C} s_{\mu'}(x') \right) \prod_p (\cosh(\beta) + s(p)\sinh(\beta))}{\sum_{s_\mu(x)} \prod_p (\cosh(\beta) + s(p)\sinh(\beta))}. \tag{2.49}$$

The only terms which survive in the numerator and denominator have either zero or an even number of link variables at each link, i.e.,

$$\begin{aligned}
\text{numerator} &= \sum_{s_\mu(x)} \left(\prod_{(x',\mu')\in C} s_{\mu'}(x') \right) \prod_p (\cosh(\beta) + s(p)\sinh(\beta)) \\
&= \sum_{s_\mu(x)} \left(\prod_{(x',\mu')\in C} s_{\mu'}(x') \right) \prod_{p'\in S} s(p')\sinh(\beta) \prod_{p\notin S} \cosh(\beta) \\
&= \sum_{s_\mu(x)} (\sinh(\beta))^{A(C)} (\cosh(\beta))^{N_p - A(C)} \\
&= 2^{N_p} (\sinh(\beta))^{A(C)} (\cosh(\beta))^{N_p - A(C)},
\end{aligned} \tag{2.50}$$

$$\begin{aligned}
\text{denominator}(=Z) &= \sum_{s_\mu(x)} \prod_p (\cosh(\beta) + s(p)\sinh(\beta)) \\
&= \sum_{s_\mu(x)} (\cosh(\beta))^{N_p} \\
&= 2^{N_p} (\cosh(\beta))^{N_p},
\end{aligned}$$

so that

---

[4] The Wilson loop calculation is a little easier in two dimensions with free boundary conditions. Periodic boundary conditions introduce a correction which is irrelevant for N_p large.

$$W(C) = (\tanh(\beta))^{A(C)}$$
$$= \exp[-\sigma A(C)] \quad \text{with } \sigma = -\log[\tanh(\beta)]. \tag{2.51}$$

This demonstrates that Wilson loops have an area law falloff, and Z_2 lattice gauge theory is in the magnetically disordered phase, for any value of $\beta > 0$ in $D = 2$ dimensions. The calculation for $U(1)$ and $SU(N)$ gauge groups is similar, with $\cosh(\beta)$ and $\sinh(\beta)$ replaced by other special functions. For abelian gauge groups such as Z_2 and $U(1)$, it is also straightforward to verify the property (2.41); in the case of Z_2 gauge theory both sides of (2.41) equal $\tanh^{A(C)}(\beta)$. For non-abelian groups, an expression of the form (2.41) holds, in $D = 2$ dimensions, for the product of traces $Tr[U(C_i)]$ of loops.

The underlying reason for magnetic disorder in $D = 2$ dimensions is the absence of a Bianchi constraint relating different components of the field strength tensor. In general, the classical gauge field equations in $D = 4$ dimensions are of two sorts. First, there are the genuine equations of motion, which follow from variation of the action with respect to A_μ^a

$$(D^\mu)^{ab} F_{\mu\nu}^b = j_\nu^a, \tag{2.52}$$

where D_μ is the covariant derivative

$$D_\mu^{ab} = \delta^{ab} \partial_\mu - g f^{acb} A_\mu^c, \tag{2.53}$$

and j_μ^a is an external current source. In electromagnetism, where the color indices a, b are absent, and the covariant derivative is just an ordinary derivative, these equations of motion are Gauss's Law and Ampere's Law. But there are also a set of conditions on the field strength tensor which follow as identities from the fact that the field strength is expressed in terms of the 4-vector potential

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c. \tag{2.54}$$

These identities are the Bianchi constraint

$$(D^\mu)^{ab*} F_{\mu\nu}^b = 0, \tag{2.55}$$

where $^*F_{\mu\nu}^b$ is the "dual" field-strength tensor

$$^*F_{\mu\nu}^a = \frac{1}{2} \varepsilon_{\mu\nu\alpha\beta} F^{a\alpha\beta}. \tag{2.56}$$

In electromagnetism, the Bianchi constraint in $D = 4$ dimensions consists of the source-free Maxwell's equations, i.e., Faraday's Law plus the condition $\nabla \cdot B = 0$. In $D = 3$ dimensions there is only the condition $\nabla \cdot B = 0$. In $D = 2$ dimensions the Bianchi constraint is absent.

Many years ago, Halpern [11] showed that it was possible to change integration variables in the functional integral from the gauge potentials $A_\mu(x)$ to the field strength variables $F_{\mu\nu}$, with the result that

$$Z = \int DA_\mu \exp\left[-\int d^D x \frac{1}{2} \text{Tr}[F_{\mu\nu}^2]\right]$$
$$= \int DF_{\mu\nu} \delta[I(F)] \exp\left[-\int d^D x \frac{1}{2} \text{Tr}[F_{\mu\nu}^2]\right], \tag{2.57}$$

where $I(F) = 0$ is the Bianchi constraint, expressed entirely in terms of the (abelian or non-abelian) field strengths. In this formulation it is clear that the only thing which correlates the values of the field strength at neighboring points is the Bianchi constraint. In two dimensions this constraint is absent, and field strengths fluctuate independently from one point to another. A similar change of variables, from link to plaquette variables, can be carried out in lattice gauge theory [12], with the similar result that in two dimensions the plaquette variables fluctuate independently. Again using Z_2 lattice gauge theory as an example, we can use the gauge freedom to transform some of the link variables to +1:

$$s_1(x) = 1, \quad s_2(x_1 = 0, x_2) = 1. \tag{2.58}$$

This is a choice of gauge, analogous to the Coulomb or Landau gauge-fixing conditions in electromagnetism. We may then change variables from the remaining link variables $s_2(x)$ to (an equal number of) plaquette variables $s(p) = s_2(x + \hat{1})s_2(x)$, so that

$$Z \to \prod_p \sum_{s(p)=\pm 1} (\cosh(\beta) + s(p)\sinh(\beta)), \tag{2.59}$$

and we see that the probability distribution factorizes into a product of independent distributions for each plaquette. Using the Stokes Law

$$\prod_{(x',\mu')\in C} s_{\mu'}(x') = \prod_{p\in S} s(p) \tag{2.60}$$

we can quickly rederive the result for Wilson loops

$$W(C) = \frac{1}{Z} \prod_{p'\in S} \sum_{s(p')=\pm 1} s(p')(\cosh(\beta) + s(p')\sinh(\beta))$$
$$\times \prod_{p\notin S} \sum_{s(p)=\pm 1} (\cosh(\beta) + s(p)\sinh(\beta))$$
$$= \frac{(2\sinh(\beta))^{A(C)} (2\cosh(\beta))^{N_p - A(C)}}{(2\cosh(\beta))^{N_p}} \tag{2.61}$$
$$= (\tanh(\beta))^{A(C)}.$$

It is clear in this case that the absence of a correlation among even the smallest loops (i.e., plaquette variables) is the crucial ingredient which leads to the area-law expression for Wilson loops of any size. The possibility of changing from link to plaquette variables, without any additional constraint on the plaquette variables, is unique to $D = 2$ dimensions.

In $D > 2$ dimensions, the area law falloff of Wilson loops holds at strong couplings, i.e., $\beta \ll 1$; the calculation, to leading order in β, gives the same result as in $D = 2$ dimensions. At weak couplings, $\beta \gg 1$, the area law certainly does not hold for small loop areas. This is a good thing; we would not like to end up predicting a linear potential between positive and negative electric charges in $3 + 1$ dimensions, and in fact one finds that $U(1)$ gauge theory is in the massless phase in $D = 4$ dimensions at weak couplings. But the area law falloff has been found, via Monte Carlo simulations, to hold for sufficiently large loop areas in non-abelian gauge theories. In the case of the Ising model, the ordered and disordered phases were distinguished by the breaking of a global Z_2 symmetry. Elitzur's theorem assures us that the magnetically disordered phase, and the massive/massless phases, cannot be distinguished by the breaking of the local gauge symmetry. But is there some other global symmetry which would serve this purpose? We will turn to this issue in the next chapter.

References

 1. Elitzur, S.: Impossibility of spontaneously breaking local symmetries. Phys. Rev. D **12**, 3978–3982 (1975)
 2. Wegner, F.J.: Duality in generalized Ising models and phase transitions without local order parameter. J. Math. Phys. **12**, 2259–2272 (1971)
 3. DeGrand, T., DeTar, C.E.: Lattice Methods for Quantum Chromodynamics. World Scientific, Singapore (2006)
 4. Seiberg, N.: Electric–magnetic duality in supersymmetric non-Abelian gauge theories. Nucl. Phys. B **435**, 129–146 (1995) (arXiv:hep-th/9411149)
 5. Intriligator, K.A., Seiberg, N.: Phases of N = 1 supersymmetric gauge theories and electric–magnetic triality. In: Strings 95: Future Perspectives in String Theory (arXiv:hep-th/9506084)
 6. Diakonov, D., Petrov, V.Y.: A formula for the Wilson loop. Phys. Lett. B **224**, 131–135 (1989)
 7. Karp, R.L., Mansouri, F., Rno, J.S.: Product integral formalism and non-Abelian stokes theorem. J. Math. Phys. **40**, 6033–6043 (1999) (arXiv:hep-th/9910173)
 8. Arefeva, I.: Non-Abelian Stokes formula. Theor. Math. Phys. **43**, 353–356 (1980) (Teor. Mat. Fiz. **43**, 111 (1980))
 9. Fishbane, P.M., Gasiorowicz, S., Kaus, P.: Stokes's theorems for non-Abelian fields. Phys. Rev. D **24**, 2324–2329 (1981)
10. DiGiacomo, A., Dosch, H.G., Shevchenko, V.I., Simonov, Yu. A.: Field correlators in QCD: Theory and applications. Phys. Rep. **372**, 319–368 (2002) (arXiv:hep-ph/0007223)
11. Halpern, M.B.: Field strength and dual variable formulations of gauge theory. Phys. Rev. D **19**, 517–530 (1979)
12. Batrouni, G.G., Halpern, M.B.: String, corner and plaquette formulation of finite lattice gauge theory. Phys. Rev. D **30**, 1782–1790 (1984)

Chapter 3
What is Confinement?

What is meant by the word "confinement"?

Most efforts aimed at proving (or at least understanding) confinement attempt to show that the potential energy between a static quark–antiquark pair (the *static quark potential*), grows linearly with the separation between the quarks. In other words, since a linear potential is associated with an area-law falloff for Wilson loops, the aim is to show that $SU(3)$ gauge theory in $D = 4$ dimensions is in the phase of magnetic disorder. But if, by the word "confinement," we mean that there is a static quark potential which rises indefinitely with quark separation, then by this definition QCD is not confining, because the static quark potential must eventually go flat due to a process, described below, known as "string breaking."

For this reason, since all of the known hadrons are singlets under the $SU(3)$ color group, many people prefer to define confinement as the absence of color charged asymptotic particle states. This defines confinement as "color confinement." Although this definition ties in very well with the historical understanding of confinement as the absence of free quarks, it has its own problems if taken seriously. The difficulty is that the asymptotic particle states of a gauge-Higgs theory, where the forces are all Yukawa-like and the dynamics something like the weak interactions, are *also* color singlet states. So if by "confinement" we mean color confinement, then Higgs theories are also confining.

Before settling on a choice of terminology, at least for the purposes of this book, it is useful to understand *why* a linear static quark potential is believed to be relevant to QCD, why that potential must ultimately level off, and how we know that the spectrum of a Higgs theory consists only of color singlets.

3.1 Regge Trajectories, and the Spinning Stick Model

What is special about the dynamics of QCD, as opposed to, e.g., the electroweak theory? A very strong clue comes from the QCD spectrum, when the spin of

mesons (and baryons) is plotted against their squared mass, as shown in Fig. 3.1. In such plots the mesons and baryons of given flavor quantum numbers seem to lie on nearly parallel straight lines, known as linear Regge trajectories. This is a very striking feature of the QCD spectrum, nothing similar is found in the electroweak theory, and the question is why it occurs.

Suppose that we picture a meson as a straight line of length $L = 2R$, with mass per unit length σ. The line rotates about a perpendicular axis through its midpoint, such that the endpoints of the line are moving at the speed of light. Then for the energy (mass) of the spinning stick we have

$$
\begin{aligned}
m = E &= 2 \int_0^R \frac{\sigma dr}{\sqrt{1 - v^2(r)}} \\
&= 2 \int_0^R \frac{\sigma dr}{\sqrt{1 - r^2/R^2}} \\
&= \pi \sigma R,
\end{aligned}
\tag{3.1}
$$

and for the angular momentum

$$
\begin{aligned}
J &= 2 \int_0^R \frac{\sigma r v(r) dr}{\sqrt{1 - v^2(r)}} \\
&= \frac{2}{R} \int_0^R \frac{\sigma r^2 dr}{\sqrt{1 - r^2/R^2}} \\
&= \frac{1}{2} \pi \sigma R^2.
\end{aligned}
\tag{3.2}
$$

Fig. 3.1 Regge trajectories for the low-lying mesons (figure from Bali [1])
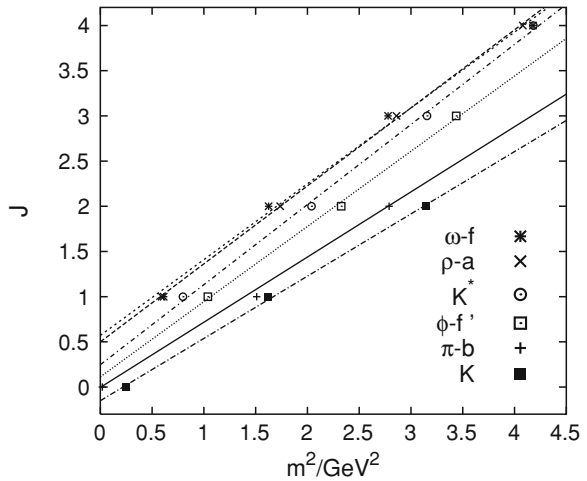
Comparing the two expressions, we see that

$$J = \frac{1}{2\pi\sigma}m^2 = \alpha' m^2. \tag{3.3}$$

The constant α' is known as the *Regge slope*. From the data, $\alpha' = 1/(2\pi\sigma) = 0.9\,\text{GeV}^{-2}$, which gives a mass/unit length of the string, or *string tension*, of

$$\sigma \approx 0.18\,\text{GeV}^2 \approx 0.9\,\text{GeV/fm}. \tag{3.4}$$

The spinning stick model is not perfect; in fact the various Regge trajectories shown in Fig. 3.1 do not pass through the origin, and have slightly different slopes. To make the model more realistic, one might want to relax the requirement of rigidity, and allow the "stick" to fluctuate in transverse directions. This line of thought leads to the formidable subject of string theory. But since QCD is the theory of quarks and gluons, the question is how a stick-like or string-like object actually emerges from that theory.

One possible answer is via the formation of a color electric flux tube. Imagine that the color electric field running between a static quark and antiquark is, for some reason, squeezed into a cylindrical region, whose cross-sectional area is nearly constant as quark–antiquark separation L increases. In that case the energy stored in the color electric field E_k will grow linearly with quark separation, i.e.

$$\text{Energy} = \sigma L \text{ with } \sigma = \int d^2 x_\perp \frac{1}{2} E_k^a(x) E_k^a(x), \tag{3.5}$$

where $E_k^a(x) = F_{0k}^a(x)$, and the integration is over a cross-section of the flux tube. This means that there will be a linearly rising potential energy associated with static sources (the "static quark potential,") and an infinite energy is required to separate these charges an infinite distance. However, this is not quite what happens in real QCD.

In real QCD with light fermions, the linear potential does not extend indefinitely. For sufficiently large quark separations it is energetically favorable to

Fig. 3.2 String breaking by quark–antiquark pair production
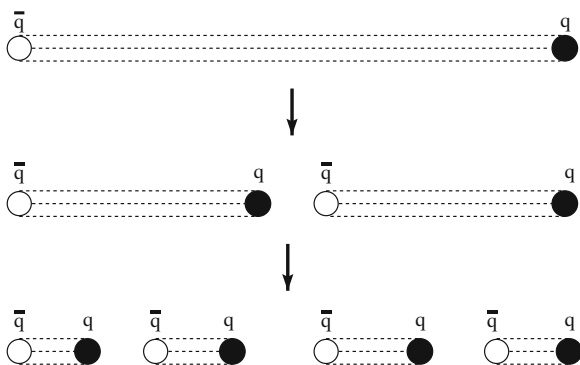
Fig. 3.3 Qualitative features
of the static quark potential in
QCD, including the effect of
quark fields
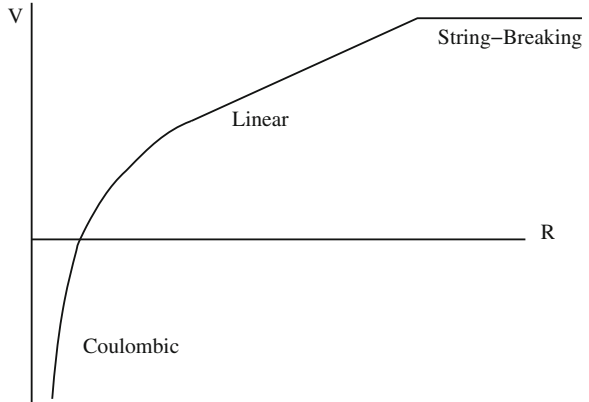
pair-produce a quark and antiquark of masses m_q, and thereby break the flux tube
(or QCD 'string') as indicated in Fig. 3.2. The static quark potential then looks
something like the sketch in Fig. 3.3. At very small distances asymptotic freedom
prevails, and the static potential is Coulomb-like. At intermediate distances a flux
tube forms, and the potential is linear. At large distances, the color field of the
static quarks is screened by the dynamical matter fields. This screening behavior
can be demonstrated in strong-coupling lattice gauge theory calculations and, more
to the point, it has been seen in lattice Monte Carlo simulations when matter fields
are added to the action [2–4].

3.2 The Fradkin–Shenker–Osterwalder–Seiler Theorem

I have already stated, without giving any reasons, that the spectrum of a gauge-
Higgs theory, like that of QCD, consists only of color-singlet states. The reasons
for thinking this is so were given by Fradkin and Shenker [5], who based their
argument on an earlier theorem due to Osterwalder and Seiler [6].

Consider a lattice gauge theory with the Higgs field in the fundamental rep-
resentation of the $SU(2)$ gauge group, and let the modulus of the Higgs field be
fixed. This theory can be realized on the lattice by the action [7]

$$S = \beta \sum_{\text{plaq}} \frac{1}{2}\text{Tr}[UUU^{\dagger}U^{\dagger}] + \gamma \sum_{x,\mu} \frac{1}{2}\text{Tr}[\phi^{\dagger}(x)U_{\mu}(x)\phi(x+\hat{\mu})], \qquad (3.6)$$

where $\phi(x)$, like the link variables $U_{\mu}(x)$, is $SU(2)$ group-valued, and transforms
under a local gauge transformation as $\phi(x) \rightarrow \phi'(x) = G(x)\,\phi(x)$. For small values
of the Higgs coupling constant γ, the dynamics of the theory resembles QCD, in
the sense that there is electric flux tube formation followed by string breaking, and
a potential (extracted from Wilson loops) which rises linearly up to the point of
string breaking, and then goes flat. At large values of γ, the situation is much more

like the weak interactions. There is no electric flux-tube formation, there are only short-range Yukawa-type interactions, and there exist massive vector bosons that might as well, in analogy to the electroweak theory, be identified as W-bosons. Many textbooks on quantum field theory, in defiance of Elitzur's theorem, would call this region a phase of spontaneously broken gauge symmetry. I will refer to it as a "Higgs-like" region, in contrast to the "confinement-like" region at small γ.

What Osterwalder and Seiler proved was that there is always a path from a point deep in the confinement-like regime, at $\beta, \gamma \ll 1$, to a point deep in the Higgs-like region, at $\beta, \gamma \gg 1$, such that the expectation value of all local, gauge-invariant observables $\langle A(x_1)\rangle, \langle B(x_2)\rangle, \langle C(x_3)\rangle, \ldots,$ and products $\langle A(x_1)B(x_2) C(x_3), \ldots, \rangle$ of such observables, vary analytically along the path. What this implies, as emphasized by Fradkin and Shenker, is that there can be no phase transition along the path, and no discontinuous change in the asymptotic particle spectrum. There is no phase boundary which separates the Higgs and confinement-like regions, and by varying the gauge and Higgs couplings the theory evolves smoothly from one region to another.

The Osterwalder–Seiler theorem does not cover the entire coupling plane, and it is in principle possible that there are phase transitions somewhere in the coupling constant space, at moderate values of β, γ. It could be, for example, that there is a massless phase somewhere in the phase diagram. The theory has been studied by lattice Monte Carlo techniques, and the situation is indicated schematically in Fig. 3.4. All indications are that the gauge theory is in a single phase, namely the massive phase, at all values of β, γ except $\gamma = 0$, where the theory is in a phase of magnetic disorder. The solid line, from roughly $\beta = 2.0$ to $\beta = 2.7$ indicates a region where certain observables (such as plaquette variables) vary very rapidly with coupling, but still analytically in the infinite volume limit. This is known as a "crossover." The most recent simulations, by Bonati et al. [8], find that this turns into a line of genuine first-order phase transitions at $\beta > 2.7$. But whether the line is a crossover, or a line of genuine first order phase transitions, is not so important. The point is that that there exists a path from any point in the phase diagram, to

Fig. 3.4 Schematic phase diagram of the gauge-Higgs model
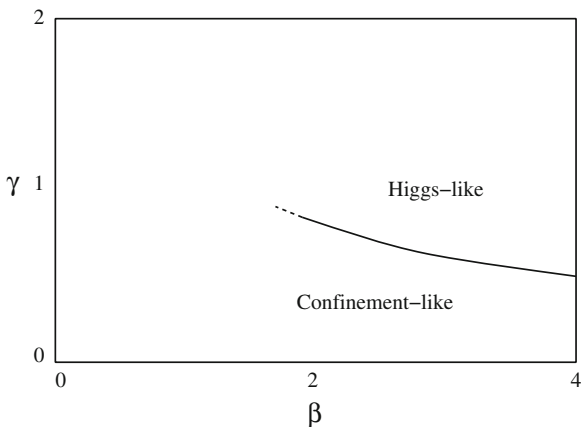
any other point in the phase diagram, such that all local gauge-invariant observables, and products of such observables, evolve analytically. This means, in particular, that a discontinuous change from a color-singlet spectrum in the confinement-like region, to a color non-singlet spectrum in the Higgs region, is ruled out.

One way of seeing that there are only singlets in the spectrum is to consider the color field emanating from a static color-charged source, in both the confinement-like and Higgs-like regions of the phase diagram. In neither case is the color field detectable far from the source. In the confinement-like region, a scalar particle will bind to the charged source, producing a color singlet which has no associated long-range color field. In the Higgs-like region, gauge forces are short range, and again the color field is undetectable a short distance away from the source. What this means, in terms of the Gauss Law, is that the Higgs field in the vacuum has adjusted itself in the neighborhood of the source to cancel out the static charge. The Fradkin–Shenker work implies that one charge cancellation mechanism evolves continuously into the other, as we move along some path from the confinement-like to the Higgs-like region in the coupling-constant plane.

One might object that gauge particles appear in the spectrum of Higgs theories as massive W-bosons. How can these gauge particles evolve continuously, along some path in the coupling-constant space, from color-singlet states in the confinement-like region? The answer is that the same gauge-invariant operators, which create composite, color singlet vector mesons in the confinement-like region, also create particle states which can be interpreted as massive gauge bosons in the Higgs-like region. An example of such operators in the $SU(2)$ gauge-Higgs theory (3.6) would be

$$\text{Tr}\left[\frac{\sigma_a}{2}\phi^\dagger(x)U_\mu(x)\phi(x+\widehat{\mu})\right] \tag{3.7}$$

where the σ_a denote the Pauli matrices. These color-singlet vector meson operators also create the massive W-bosons in the Higgs-like region, which is made clear by going to a unitary gauge $\phi = \mathbb{1}$, and expanding $U_\mu(x) \approx \mathbb{1} + igA_\mu^a\sigma_a/2$. From this fact it is evident that W-bosons in the Higgs-like region *can* evolve continuously, from states which we would interpret as bound states of scalar particles in the confinement-like region.

3.3 Remnant Gauge Symmetries

Elitzur's theorem, and the Fradkin–Shenker–Osterwalder–Seiler theorem, do some violence to one's preconceptions about gauge-Higgs theories. Consider a Higgs field transforming in the fundamental representation of the gauge group, so in an $SU(N)$ theory the Higgs field $\boldsymbol{\varphi}$ has N components φ^i. Standard introductions to the Higgs mechanism often teach that when the Higgs potential has a certain form, usually a "Mexican-hat," i.e.

$$V(\varphi) = \lambda(\boldsymbol{\varphi} \cdot \boldsymbol{\varphi} - \mu)^2, \tag{3.8}$$

such that the potential is minimized away from $\boldsymbol{\varphi} = 0$, then the Higgs potential acquires a vacuum expectation value $\langle\boldsymbol{\varphi}\rangle \neq 0$, and the gauge symmetry is spontaneously broken. The fact is, however, that in the absence of gauge fixing we have $\langle\boldsymbol{\varphi}\rangle = 0$ regardless of the form of the Higgs potential. If, on the other hand, we fix to a unitary gauge, where one uses the gauge freedom to rotate the Higgs field into a definite direction in color space, then $\langle\boldsymbol{\varphi}\rangle \neq 0$, also regardless of the Higgs potential. This is just due to quantum fluctuations away from the classical minimum.

Certain gauge choices, however, such as Landau and Coulomb gauge, do not fix all of the gauge symmetry, but leave some global subgroup unfixed. Elitzur's theorem does not apply to these remnant gauge symmetries, and in principle such remnant symmetries *can* break spontaneously. There are a number of suggestions in the literature which argue that the distinction between confining and non-confining gauge theories is related to whether or not some remnant gauge symmetry is, or is not, spontaneously broken. Three examples are the Kugo–Ojima confinement criterion [9, 10], the Coulomb confinement criterion [11], and the Pisa criterion [12, 13] associated with dual superconductivity (which will be discussed in Chap. 8).

3.3.1 Landau Gauge and the Kugo–Ojima Criterion

If a gauge field A_μ satisfies the Landau gauge condition $\partial_\mu A_\mu = 0$, then so does the gauge-transformed field $GA_\mu G^\dagger$, where $G(\mathbf{x}, t) = G$ is a spacetime-independent gauge transformation; i.e. a global transformation. There is also, as pointed out by Hata [14, 15], a remnant space-time dependent transformation which preserves the Landau gauge condition:

$$G(x) = \exp[i \wedge^a (\varepsilon; x) \frac{1}{2} \sigma_a]$$
$$\wedge^a(\varepsilon; x) = \varepsilon_\mu^a x^\mu - g \frac{1}{\partial^2} (A_\mu \times \varepsilon_\mu)^a + O(g^2), \tag{3.9}$$

where g is the gauge coupling of the $SU(2)$ gauge theory. This transformation also counts as global, since it depends on a finite number of parameters ε_μ^a, and is therefore not subject to the Elitzur theorem. These remnant symmetries exist in any covariant gauge.

Kugo and Ojima introduce the function $u^{ab}(p^2)$, defined by the expression

$$u^{ab}(p^2)\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right)$$
$$= \int d^4x e^{ip(x-y)} \langle 0|T[D_\mu c^a(x) g(A_\nu \times \overline{c})^b(y)|0\rangle, \tag{3.10}$$

where $c, \bar{c}$ are ghost and antighost fields required, in the Faddeev-Popov approach, for quantization in any covariant gauge. They then show that the expectation value of color charge in any physical state vanishes

$$\langle \text{phys}|Q^a|\text{phys}\rangle = 0, \tag{3.11}$$

providing that (a) remnant symmetry with respect to spacetime-independent gauge transformations $G(\mathbf{x}, t) = G$ is unbroken; and (b) the following condition is satisfied:

$$u^{ab}(0) = -\delta^{ab}. \tag{3.12}$$

This latter condition is known as the Kugo–Ojima confinement criterion, and it implies that the ghost propagator is more singular, and the gluon propagator less singular, than a simple pole at $p^2 = 0$ [10]. A number of efforts have focussed on verifying this condition (or its corollaries) both analytically [16] and numerically [17].

It turns out that the Kugo–Ojima condition (3.12) is itself tied to the unbroken realization of remnant gauge symmetry in covariant gauges. It was shown by Hata [14, 15] (see also Kugo [10]) that the condition (3.12) is a necessary (and probably sufficient) condition for the unbroken realization of the residual spacetime-dependent symmetry (3.9), while an unbroken, spacetime independent symmetry is required, in *addition* to (3.12), for the vanishing of $\langle \psi|Q^a|\psi\rangle$ in physical states.

Thus the Kugo–Ojima scenario requires that the full remnant gauge symmetry in Landau gauge, i.e. both the spacetime dependent and the spacetime independent residual gauge symmetries, must be unbroken. Both of these symmetries are necessarily broken if a Higgs field acquires an expectation value in a covariant gauge, so in the Kugo–Ojima scheme the confinement phase is a phase of unbroken remnant gauge symmetry.

3.3.2 Coulomb Confinement

The Coulomb gauge condition $\nabla \cdot A = 0$ is preserved by time-dependent, but space-independent gauge transformations $G(\mathbf{x}, t) = G(t)$. This is a remnant symmetry which is local in time, but global in space, and if we restrict ourselves to a single time-slice, or a finite set of time-slices, then in principle this symmetry can break spontaneously.[1] On the lattice, Coulomb gauge is equivalent to maximizing, on each time slice, the quantity

---

[1] What this means is that the order parameter for the transition can be non-zero in the infinite 3-volume limit at any fixed time, but have different values at different times, averaging to zero if we average over all times. Thus the symmetry can be spontaneously broken on any single timeslice, or finite set of timeslices, without violating the Elitzur theorem.

$$R(t) = \sum_{\mathbf{x}} \sum_{k=1}^{3} \text{Tr}[U_k(\mathbf{x}, t)], \tag{3.13}$$

and this quantity is obviously undisturbed by $U_k(\mathbf{x}, t) \rightarrow G(t) U_k(\mathbf{x}, t) G^\dagger(t)$. On the other hand, for timelike links $U_0(\mathbf{x}, t) \rightarrow G(t) U_0(\mathbf{x}, t) G^\dagger(t + 1)$, so if the remnant symmetry is unbroken at fixed times t and $t + 1$, then necessarily

$$\langle \text{Tr}[U_0(\mathbf{x}, t)] \rangle = 0 \tag{3.14}$$

because the left hand side is not invariant with respect to the remnant symmetry. It will be shown in Chap. 9 that (3.14) implies that the energy of the Coulomb field associated with a static, isolated color charge is infinite, even when the usual ultraviolet divergence is regulated with a lattice cutoff. For this reason, Marinari, Paciello, Parisi and Taglienti [11] (see also [18]) suggested that a possible criterion for confinement is that the remnant gauge symmetry in Coulomb gauge is unbroken. In principle the criterion can work even in the presence of string-breaking matter fields such as quarks, because the Coulomb field is not necessarily the actual gauge field surrounding a color charge, and the color Coulomb potential is only an upper bound on the static quark potential.v

3.3.3 Remnant Symmetry Breaking

The problem with confinement criteria which rely on the broken or unbroken status of a remnant gauge symmetry is that, first of all, different remnant symmetries break at different couplings (so which is the "right" remnant symmetry to consider?), and secondly, they break where in fact there is no change whatever in the physical state of the theory. The situation for the gauge-Higgs model of (3.6) is shown in Fig. 3.5. We may use, as the order parameter for spontaneous breaking of remnant Landau gauge symmetry, the spatial average of the Higgs field

Fig. 3.5 The location of remnant global gauge symmetry breaking in Landau and Coulomb gauges, in the $\beta - \gamma$ coupling plane. From [19]
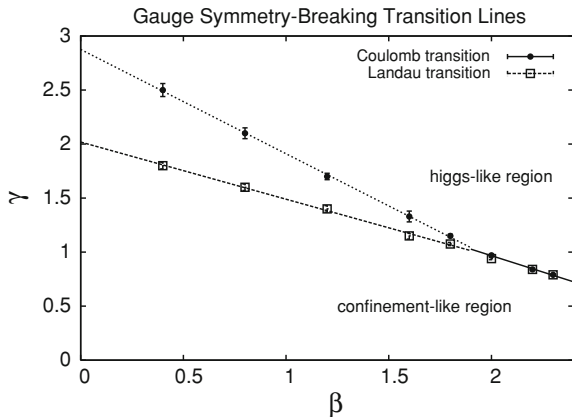
$$\widetilde{\phi} = \frac{1}{V} \sum_x \phi(x), \tag{3.15}$$

and for spontaneous breaking of remnant symmetry in Coulomb gauge, the spatial average of timelike links on any given timeslice

$$\widetilde{U}(t) = \frac{1}{V_3} \sum_{\mathbf{x}} U_0(\mathbf{x}, t). \tag{3.16}$$

The test is whether or not the expectation values of the magnitudes of $\tilde{\phi}$ and $\tilde{U}(t)$ extrapolate to zero as the 4-volume V, and 3-volume V_3, respectively, are taken to infinity.

The result of a lattice Monte Carlo simulation [19] is shown in Fig. 3.5. What we see is that the Coulomb and Landau gauge remnant symmetries break along lines which coincide along the crossover, but diverge from one another at lower β, and occur in a theory where there is, in fact, no transition between different physical states.[2] This calls into question both the Kugo–Ojima criterion, and the Coulomb confinement criterion. A third example (the Pisa criterion [12, 13]), which identifies confinement as the broken phase of a remnant "dual" $U(1)$ gauge symmetry, was studied in ref. [20]. It turns out that this criterion also predicts transitions between confining and non-confining phases, where in fact no such transitions exist.

3.4 Center Symmetry

In the Ising model example, the distinction between the ordered and disordered phases corresponded to whether or not a certain global symmetry of the theory is spontaneously broken. But in the gauge-Higgs theory we have discussed, it seems impossible to distinguish between the confinement-like and Higgs-like regions of the theory in terms of the spontaneous symmetry breaking of a gauge symmetry. If we classify the theory based on the behavior of very large Wilson loops, we find that the theory is *almost* everywhere in a massive phase, characterized by a perimeter-law falloff. The exception is along the line at $\gamma = 0$. As this line is approached, string-breaking occurs at ever larger charge separations. Precisely at $\gamma = 0$ the Higgs field decouples from the gauge field, string breaking disappears entirely and (it is believed) Wilson loops have an area-law falloff at arbitrarily

---

[2] Order parameters such as (3.15) and (3.16) are non-local quantities, in the sense that they are defined via a non-local functional of the fields, i.e. the gauge transformation to Landau or Coulomb gauge respectively. One could argue that a non-analyticity in these order parameters (breaking of a remnant symmetry), must, by definition, indicate a phase transition. The point, however, is that if the non-analyticity in a non-local parameter is unaccompanied by a non-analyticity in either the spectrum, the scattering amplitudes, or any thermodynamic quantity, then there is really no physical distinction between phases across the supposed transition.

large scales. In other words, at $\gamma = 0$ the system is in a phase of magnetic disorder. The same thing happens in a gauge theory like QCD, if the quark masses are taken to infinity. It is then worth asking: is there some global symmetry which distinguishes the massive phase from the magnetically disordered phase?

There is. It is known as *center symmetry*, and, although a global symmetry, it is associated with the center of the gauge group.

We will need a few basic facts from group theory. First of all, the center of a group consists of those group elements which commute with all elements of the group. For an $SU(N)$ gauge theory, this is the set of all group elements proportional to the identity:

$$z_n \mathbb{1}_N = \exp\left(\frac{2\pi i n}{N}\right) \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & 1 \end{bmatrix} \qquad (n = 0, 1, 2, ..., N-1). \qquad (3.17)$$

The $\{z_n = \exp[2\pi i n/N]\}$ are the elements of the discrete abelian subgroup Z_N, and the set $\{z_n \mathbb{1}_N\}$ is the Z_N subgroup of $SU(N)$. While there are an infinite number of representations of $SU(N)$, there are only N representations of Z_N. This means that every representation of $SU(N)$ falls into one of N subsets (known as *class* or *N-ality*), depending on the representation of the Z_N subgroup. The N-ality k of a given representation is given by the number of boxes in the corresponding Young Tableau, mod N. If $M[g]$ is the matrix representation of group element g in a representation of N-ality k, and $z \in Z_N$ is an element in the center, then

$$M[zg] = z^k M[g]. \qquad (3.18)$$

Now consider the lattice action (2.24) of a pure gauge theory, i.e. a gauge theory with no matter fields. It is easy to see that this action is invariant under the following transformation: Pick any fixed time $t = t_0$ and multiply the timelike links on that timeslice by an element of the center subgroup, i.e.

$$U_0(\boldsymbol{x}, t_0) \rightarrow z U_0(\boldsymbol{x}, t_0) \text{ all } \boldsymbol{x}, \text{ fixed } t_0, \qquad (3.19)$$

with $z \in Z_N$ if the gauge group is $SU(N)$; all other link variables are unchanged. The transformation is indicated in Fig. 3.6; transformations defined at other fixed times are equivalent up to a gauge transformation. The only elements of the lattice action which might be affected are the timelike plaquettes containing the links $U_0(\boldsymbol{x}, t_0)$ but in fact these are unchanged:

$$\begin{aligned} &U_i(\boldsymbol{x}, t_0) z U_0(\boldsymbol{x} + \hat{i}, t_0) U_i^\dagger(\boldsymbol{x}, t_0 + 1) U_0^\dagger(\boldsymbol{x}, t_0) z^{-1} \\ &= U_i(\boldsymbol{x}, t_0) U_0(\boldsymbol{x} + \hat{i}, t_0) U_i^\dagger(\boldsymbol{x}, t_0 + 1) U_0^\dagger(\boldsymbol{x}, t_0) \end{aligned} \qquad (3.20)$$
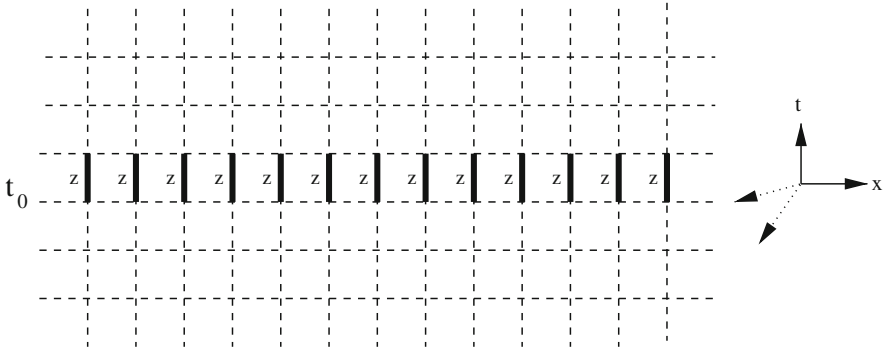
Fig. 3.6 The global center transformation. Each of the indicated links in the time direction, on the timeslice $t = t_0$, is multiplied by the same element z of the center subgroup. The lattice action of a pure gauge theory is left unchanged by this operation

because of the fact that the center elements commute with all elements of the group, and with any link variable in particular.

The reason that center symmetry is so important is that if this symmetry

- is broken explicitly, e.g. by matter fields in the fundamental representation of the gauge group, or
- is broken spontaneously, which happens at high temperature, and *may* happen if there are matter fields in the adjoint representation of the gauge group, or
- is trivial, i.e. the center subgroup consists only of the unit element, as in the exceptional group G_2.

Then the static quark potential cannot be linear asymptotically; it must become flat.

Let's see why matter fields in the fundamental representation, or in any representation of the gauge group of N-ality $k \neq 0$, must break the center symmetry explicitly. A matter field in color representation r couples to the gauge field via a term in the action

$$\gamma \sum_{x,\mu} \phi^\dagger(x) U_\mu^{(r)}(x)\phi(x+\hat{\mu}) + \text{c.c.,} \tag{3.21}$$

where the superscript r means that the link variables are in the r-representation. Under the global center symmetry transformation

$$\phi^\dagger(\boldsymbol{x}, t_0) U_0^{(r)}(\boldsymbol{x}, t_0)\phi(\boldsymbol{x}, t_0+1) \rightarrow z^k \phi^\dagger(\boldsymbol{x}, t_0) U_0^{(r)}(\boldsymbol{x}, t_0)\phi(\boldsymbol{x}, t_0+1). \tag{3.22}$$

So if the N-ality k of the representation of the matter field is non-zero, the matter action breaks the global Z_N center symmetry. But matter fields of N-ality $k \neq 0$ are exactly the fields associated with string-breaking. Suppose we have two static color charges, in the fundamental representation of the gauge group. In order to break the electric flux tube which forms between these sources, by the process pictured in Fig. 3.2, there must be dynamical particles which can bind to the

sources and produce a color singlet. This is not possible if the particles are in a color representation with N-ality $k = 0$ (such as gluons, or any matter fields in an adjoint representation). Gluons can screen color in the sense that they can bind to a source of N-ality $k \neq 0$, and produce a particle with color in a lower dimensional representation of the gauge group with the *same* N-ality k, but it is not possible for the bound state to change the N-ality of the source. Particles with color charge in the fundamental representation, however, with $k = 1$, can bind and form a color singlet with any color-charged source. So if Z_N center symmetry is completely broken by $k = 1$ matter fields, then the static potential between any two sources must ultimately become flat.

In the case of a group such as G_2, in which the center subgroup is only the unit element, gluons can bind with any color charge to form a singlet. So in this theory also, the static potential between any two sources is flat at large distances.

We will see in the next chapter that, if center symmetry is unbroken, then the energy of an isolated color charge in the fundamental representation is infinite, and the potential energy of a particle-antiparticle pair rises without limit as the charges separate. Conversely, if the symmetry is spontaneously broken, the energy of an isolated color charge is finite, and a particle-antiparticle pair *can* be separated.

In short, center symmetry is associated with an infinitely rising static quark potential, and is therefore the global symmetry associated with the magnetically disordered phase of a gauge theory. When the symmetry is broken in some way, then the gauge theory exists in some other phase.

Now it is time to return to the semantic question that opened this chapter. What is "confinement"? We have seen that "color confinement" exists in a Higgs theory; it is not necessarily associated with a linear potential, flux tube formation, and Regge trajectories. It is perfectly all right to define confinement as color confinement, providing it is understood that such a term does not necessarily refer to the phenomena we are interested in. I prefer to reserve the term "confinement" to refer to a particular phase of gauge theories, namely, the magnetically disordered phase. Confinement in this sense is not the same thing as color confinement (although that is implied), nor is it the same thing as a mass gap.[3] This terminology has a price; it means that QCD is a "confinement-like" theory, i.e. it resembles a magnetically disordered theory only below some string-breaking length scale, which goes off to infinity as the quark masses are taken to infinity. But string-breaking is easy to understand; the hard problem is to understand the origins of the linear potential. For this reason it makes sense to concentrate on magnetic disorder. The confinement problem is solved if we can understand why non-abelian gauge theories with a non-trivial center symmetry can only exist, in $D \leq 4$ dimensions, in the magnetically disordered phase.

---

[3] The Clay Mathematics Institute offers a large cash prize [21] for proving that Yang–Mills theory has a mass gap, which would be true in any gauge theory which is in either a magnetically disordered phase, as in pure Yang–Mills theory, or a massive phase, as in Higgs theories and in real QCD.

References

1. Bali, G.: QCD forces and heavy quark bound states. Phys. Rep. **343**, 1–136 (2001) arXiv: hep-ph/0001312
2. Philipsen, O., Wittig, H.: String breaking in non-Abelian gauge theories with fundamental matter fields. Phys. Rev. Lett. **81**, 4056–4059 [Erratum-ibid. 83:2684 (1999)] (1998) arXiv:hep-lat/9807020
3. Knechtli, F.: String breaking and lines of constant physics in the SU(2) Higgs model. Nucl. Phys. Proc. Suppl. **83**, 673–675 (2000) arXiv:hep-lat/9909164
4. DeTar, C.E., Heller, U.M., Lacock, P.: First signs for string breaking in two flavor QCD. Nucl. Phys. Proc. Suppl. **83**, 310–312 (2000) arXiv:hep-lat/9909078
5. Fradkin, E. Shenker, S.: Phase diagrams of lattice gauge theories with Higgs fields. Phys. Rev. D **19**, 3682–3697 (1979)
6. Osterwalder, K. Seiler, E.: Gauge field theories on the lattice. Ann. Phys. **110**, 440–471 (1978)
7. Lang, C., Rebbi, C., Virasoro, M.: The phase structure of a non-Abelian Gauge Higgs field system. Phys. Lett. **104B**, 294–300 (1981)
8. Bonati, C., Cossu, G., D'Elia, M., Di Giacomo, A.: Phase diagram of the lattice SU(2) Higgs model. Nucl. Phys. B **828**, 390–403 (2010) arXiv:0911.1721 [hep-lat]
9. Kugo, T., Ojima, I.: Local covariant operator formalism of non-Abelian gauge theories and quark confinement problem. Prog. Theor. Phys. Suppl. **66**, 1–232 (1979)
10. Kugo, T.: The universal renormalization factors Z(1)/Z(3) and color confinement condition in non-Abelian gauge theory. arXiv:hep-th/9511033
11. Marinari, E., Paciello, M.L., Parisi, G., Taglienti, B.: The string tension in gauge theories: a suggestion for a new measurement method. Phys. Lett. B **298**, 400–404 (1993) arXiv:hep-lat/9210021
12. Del Debbio, L., Di Giacomo, A., Paffuti, G.: Detecting dual superconductivity in the ground state of gauge theory. Phys. Lett. B **349**, 513–518 (1995) arXiv:hep-lat/9403013
13. DiGiacomo, A., Lucini, B., Montesi, L., Paffuti, G.: Color confinement and dual superconductivity of the vacuum. 1. Phys. Rev. D **61**, 034503-1–034503-10 (2000) arXiv: hep-lat/9906024
14. Hata, H.: Restoration of the local gauge symmetry and color confinement in non-Abelian gauge theories. II. Prog. Theor. Phys. **69**, 1524–1536 (1983)
15. Hata, H.: Restoration of the local gauge symmetry and color confinement in non-Abelian gauge theories. Prog. Theor. Phys. **67**, 1607–1618 (1982)
16. Alkofer, R., Fischer, C.S., Llanes-Estrada, F.J., Schwenzer, K.: The Quark–gluon vertex in Landau gauge QCD: Its role in dynamical chiral symmetry breaking and quark confinement. Ann. Phys. **324**, 106–172 (2009) arXiv:0804.3042 [hep-ph]
17. Nakajima, H., Furui, S., Yamaguchi, A.: Numerical study of the Kugo–Ojima criterion and the Gribov problem in the Landau gauge. arXiv:hep-lat/0007001
18. Greensite, J., Olejnik, S., Zwanziger, D.: Coulomb energy, remnant symmetry, and the phases of non-Abelian gauge theories. Phys. Rev. D **69**, 074506-1–074506-18 (2004) arXiv:hep-lat/0401003
19. Caudy, W., Greensite, J.: On the ambiguity of spontaneously broken gauge symmetry. (2008) Phys. Rev. D **78**, 025018-1–025018-11 [arXiv:0712.0999 [hep-lat]]
20. Greensite, J., Lucini, B.: Is confinement a phase of broken dual gauge symmetry? (2008) Phys. Rev. D **78**, 085004-1–085004-7 [arXiv:0806.2117 [hep-lat]]
21. http://www.claymath.org/millennium/Yang-Mills_Theory/

Chapter 4
Order Parameters for Confinement

There are certain observables whose behavior indicates that a gauge system is, or is not, in a phase of magnetic disorder. These are the Wilson loop, the Polyakov loop, the 't Hooft loop, and the vortex free energy. The Wilson loop has been discussed already, to some extent. The Polyakov loop is a Wilson loop which winds around a finite, periodic lattice in the time direction; it is a crucial probe of center symmetry and the high temperature deconfinement phase transition, in which center symmetry is spontaneously broken. The 't Hooft loop is an operator which creates an object known as a *center vortex*. The center vortex scenario for confinement will be discussed in some detail in Chaps. 6 and 7.

4.1 The Wilson Loop

The Wilson loop operator [1] can be thought of in two ways:

1. rectangular timelike loops represent the creation, propagation, and destruction of two static sources of quark and antiquark color charge, respectively, located at fixed points in space (cf. Chap. 2);
2. spacelike loops are probes of gauge-field fluctuations in the vacuum state, independent of any external source.

Of course, spacelike and timelike loops are not intrinsically different, but are related to one another by Lorentz (or, in the Euclidean formulation, rotation) invariance. This means that the interaction energy between static sources is related to vacuum fluctuations in the absence of external sources.

Let us again consider the relation of timelike loops to the static quark potential, this time for static charges in any group representation r. The theory consists of an $SU(N)$ lattice gauge field coupled to a single, very massive, scalar field in color group representation r. The action is

$$S = -\frac{\beta}{N} \sum_p \mathrm{ReTr}[U(p)] + S_{\mathrm{matter}}, \tag{4.1}$$

where

$$S_{\mathrm{matter}} = -\sum_{x,\mu} \left(\phi^\dagger(x) U_\mu^{(r)}(x)\phi(x+\hat{\mu}) + \mathrm{c.c.} \right) + \sum_x (m^2 + 2D)\phi^\dagger(x)\phi(x), \tag{4.2}$$

and $U_\mu^{(r)}(x)$ denotes the link variable in group representation r. Again let Q be an operator which creates a two-particle color-singlet state, with separation R,

$$Q(t) = \phi^\dagger(0,t) \left\{ \prod_{n=0}^{R-1} U_i^{(r)}(n\hat{\imath},t) \right\} \phi(R\hat{\imath},t). \tag{4.3}$$

The Q operator not only creates the two charged massive particles, but also creates a thin line of color electric flux running between those particles. In general this is not the minimal energy configuration for the color electric field.

We then consider the vacuum expectation value $\langle Q^\dagger(T)Q(0) \rangle$. This expression can be interpreted as creating a massive color charged pair joined by a line of color electric flux, which propagates for some Euclidean time T, and is then annihilated. As $m^2 \to \infty$, string-breaking effects can be neglected, and we can integrate out the scalar field

$$\langle Q^\dagger(T)Q(0) \rangle \sim m^{-4(T+1)} \frac{1}{Z} \int DU \chi_r[U(R,T)]e^{-S_U}$$
$$\sim m^{-4(T+1)} W_r(R,T). \tag{4.4}$$

Note that $W_r(R, T)$ is computed in the pure gauge theory, with no matter fields. Here $\chi_r[g]$ is the *group character* of group element g in representation r, S_U is the action without the S_{matter} term, $U(R, T)$ is the *holonomy*, i.e. the ordered product of link variables around the loop, and $W_r(R, T)$ is the expectation value

$$W_r(R,T) = \langle \chi_r[U(R,T)] \rangle. \tag{4.5}$$

A holonomy is just a Wilson loop before taking the trace; a "group character" is the trace of a group element in the given representation. In the continuum, the holonomy $U(C)$ is the path-ordered Wilson loop

$$U(C) = P \exp\left[ig \oint_C dx^\mu A_\mu(x) \right], \tag{4.6}$$

where the path-ordering symbol P means that, in the power-series expansion of the exponential, a product of $A_\mu(x)$ operators at different points on the loop are ordered according to where they appear on the loop relative to some beginning point on the loop.

Then we have, by the same logic as (2.33)

$$W_r(R, T) \propto \sum_n |c_n|^2 e^{-\Delta E_n T}. \tag{4.7}$$

where the sum runs over states containing two charged particles separated by a distance R, and ΔE_n is the energy of the n-th state above the vacuum energy. The minimal energy state of the color field joining the charged particles is singled out in the $T \to \infty$ limit, and

$$V_r(R) = -\lim_{T \to \infty} \log\left[\frac{W_r(R, T+1)}{W_r(R, T)}\right] \tag{4.8}$$

is the static quark potential for color sources in group representation r. Again, this is the potential for massive color charged sources. We assume the absence of light matter fields which could, by pair production, screen the charge of the massive sources. The confinement problem is to show that

$$V_r(R) \sim \sigma_r R \tag{4.9}$$

at large R, for non-zero N-ality representations, or more generally an area-law falloff for Wilson loops

$$W_r(C) \sim \exp[-\sigma_r \, \text{Area}(C)] \tag{4.10}$$

for contours C enclosing a large minimal area.

The relationship of the Wilson loop to the potential between massive, color-charged particles, and the fact that the timelike lines of a loop can be interpreted as the worldlines of static particles, sometimes obscures the fact that the Wilson loop is an observable which measures the vacuum fluctuations of gauge fields, and that this observable makes sense even if there are no matter particles whatever in the theory. This is particularly clear for spacelike loops C, contained entirely in some spatial volume at constant time. In that case

$$W_r(C) = \langle \Psi_0 | \chi_r[U(C)] | \Psi_0 \rangle, \tag{4.11}$$

where Ψ_0 is the ground state of the gauge-field Hamiltonian, and there are no charged sources whatever. In fact, in relating the Wilson loop to the static potential, we assumed that the mass of the matter field was large but finite. On the other hand, if the mass is finite there must be string-breaking at some distance scale. To avoid string-breaking it is necessary to take the $m \to \infty$ limit, in which case the correlator in Eq. (4.4) is actually zero. In extracting the Wilson loop we have factored out the mass term, retaining the holonomy that depends only on the gauge field, and whose expectation value makes sense in the $m \to \infty$ limit.

It should also be noted that the expectation value of a Wilson loop is always defined with an implicit (usually lattice) ultraviolet regularization. In the continuum limit, the gluon propagator is singular at short distances, and this leads to a perimeter contribution to the expectation value with a singular coefficient.

For timelike loops, this singularity can be viewed as a divergent self-energy of the charged particles and antiparticles running along the loop. To really make sense of Wilson loops in the continuum, some "smearing" of the loop, i.e. a superposition of many nearby loops, is required (see, e.g. [2]). The same is true for the Polyakov lines discussed in the next section.

4.2 The Polyakov Loop

Consider an $SU(N)$ gauge theory with either no matter fields, or only N-ality = 0 matter fields, formulated on a lattice which is finite and periodic in the time direction. As we have seen, the action is invariant under the global Z_N group of center symmetry transformations

$$U_0(\mathbf{x}, t_0) \rightarrow z U_0(\mathbf{x}, t_0), \;\; z \in Z_N, \quad \text{all } \mathbf{x}, \tag{4.12}$$

implemented on all timelike link variables at any fixed time t_0. This transformation, which we have already encountered in Sect. 3.4 of Chap. 3, does not change the action or plaquette variables or Wilson loops. But there are certain gauge-invariant observables which *are* affected. Consider the trace of a loop holonomy that winds once around the lattice in the periodic time direction, i.e.

$$P(\mathbf{x}) = \text{Tr}[U_0(\mathbf{x}, 1)U_0(\mathbf{x}, 2)\ldots U_0(\mathbf{x}, L_t)], \tag{4.13}$$

where L_t is the extension of the lattice in the time direction. This observable is known as a *Polyakov loop* [3]. Under a center transformation $U_0(x, t_0) \rightarrow z U_0(x, t_0)$, we find

$$P(\mathbf{x}) \rightarrow z P(\mathbf{x}). \tag{4.14}$$

Since the Polyakov loop is not invariant under the global center transformation, its expectation value serves as an order parameter for the spontaneous breaking of the symmetry. There are two possibilities:

$$\langle P(\mathbf{x}) \rangle = \begin{cases} 0 & \text{unbroken } Z_N \text{ symmetry phase} \\ \text{non-zero} & \text{broken } Z_N \text{ symmetry phase} \end{cases}. \tag{4.15}$$

Which of these possibilities is realized by the theory has a lot to do with confinement, because Polyakov loops are, in fact, order parameters for the confinement property. In order to understand this important point, a brief digression into the thermodynamics of quantum field theory is called for.

On the lattice, finite temperature is represented by time-asymmetric lattices, with temperature proportional to $1/L_t$, where L_t is the lattice length in the time direction. The argument for this is straightforward. In quantum statistical mechanics, the partition function is defined to be

$$\begin{aligned} Z &= \text{Tr} e^{-\beta H} \\ &= \sum_n \langle n | e^{-\beta H} | n \rangle, \end{aligned} \tag{4.16}$$

where, in the standard notation, β is not the lattice coupling but rather $\beta = 1/T$, where T is the temperature.[1] The sum is over any complete orthonormal set of states. Now in non-relativistic quantum mechanics, the relationship between the Hamiltonian and path-integral formulations is given by

$$\langle x_f | e^{-iH(t_f - t_0)} | x_i \rangle = \int Dx(t_0 < t < t_f) e^{iS}, \tag{4.17}$$

with the integral over paths $x(t)$ subject to boundary conditions $x(t_0) = x_i$, $x(t_f) = x_f$. Likewise in field theory, after a Wick rotation to Euclidean time

$$\langle \phi_f | e^{-H\beta} | \phi_i \rangle = \int D\phi(\mathbf{x}, 0 < t < \beta) e^{-S}, \tag{4.18}$$

with boundary conditions $\phi(\mathbf{x}, 0) = \phi_i(\mathbf{x})$, $\phi(\mathbf{x}, \beta) = \phi_f(\mathbf{y})$. Here $|\phi(\mathbf{x})\rangle$ represents an eigenstate of the field operator; i.e. it is a state in which the quantum field has a definite value at each point. The set of all such states is a complete set of states in the field theory. This means that the partition function at finite temperature can be expressed as a path integral

$$\begin{aligned} Z &= \int D\phi(\mathbf{x}, 0) D\phi(\mathbf{x}, \beta) \delta[\phi(\mathbf{x}, 0) - \phi(\mathbf{x}, \beta)] \int D\phi(x, 0 < t < \beta) e^{-S} \\ &= \int D\phi(x, 0 \le t \le \beta) \delta[\phi(\mathbf{x}, 0) - \phi(\mathbf{x}, \beta)] e^{-S}, \end{aligned} \tag{4.19}$$

where the delta function simply imposes periodic boundary conditions in the time direction. We can also just write

$$Z = \int D\phi(x, 0 \le t < \beta) e^{-S} \tag{4.20}$$

with the understanding that periodic boundary conditions in time are imposed.

Thermal averages, in quantum statistical mechanics, are defined as

$$\langle Q \rangle = \frac{1}{Z} \mathrm{Tr} e^{-\beta H} Q, \tag{4.21}$$

and by the same reasoning as above, this quantity is given by

$$\langle Q \rangle = \frac{1}{Z} \int D\phi(x, 0 \le t < \beta) Q[\phi(\mathbf{x}, 0)] e^{-S} \tag{4.22}$$

with periodic boundary conditions in the time direction. After lattice regularization, $\beta = L_t a$, where a is the lattice spacing. Then the temperature is related to the extension of the periodic lattice in the time direction via $T = 1/(L_t a)$.

---

[1] The notation, while standard, is potentially confusing, because β is also used for lattice coupling, and T for time; so I emphasize again that in this section T stands for temperature, not time. In fact, as we will see, it is β which represents a time extension.

A quantity of particular interest in non-abelian gauge theory is the difference F_q between the free energy of a gauge field theory containing an isolated, static quark, and the free energy of a pure gauge system. Let Tr_q denote the trace over states with a single static color charge, due to a massive quark, at some point $\mathbf{x}$. Then

$$
\begin{aligned}
e^{-\beta F_q} &\equiv \frac{Z_q}{Z} \\
&= \frac{\text{Tr}_q e^{-\beta H}}{\text{Tr} e^{-\beta H}} \\
&= \frac{\sum_n \langle \Psi_n(1 \text{ quark})|e^{-\beta H}|\Psi_n(1 \text{ quark})\rangle}{\sum_n \langle \Psi_n(0 \text{ quark})|e^{-\beta H}|\Psi_n(0 \text{ quark})\rangle}.
\end{aligned}
\tag{4.23}
$$

Passing to the path-integral representation for thermal averages, the 1-quark state is obtained by creating a massive quark at point $\mathbf{x}$ and time $t = 0$, propagation of the static quark to point $\mathbf{x}$, $t = \beta$, and annihilation at $t = \beta$. Integrating over the massive quark field at times $0 \leq t < \beta$ produces a Wilson line running between $(\mathbf{x}, 0)$ and $(\mathbf{x}, \beta)$, and the imposition of periodic boundary conditions results in a closed Wilson loop running through the periodic lattice. Discarding an overall constant proportional to $m^{-\beta}$, the final result is that

$$
\begin{aligned}
e^{-\beta F_q} &\propto \frac{1}{Z} \int DU_\mu \text{Tr}[U_0(\mathbf{x}, 1)U_0(\mathbf{x}, 2)\ldots U_0(\mathbf{x}, L_t)]e^{-S} \\
&= \langle P(\mathbf{x})\rangle,
\end{aligned}
\tag{4.24}
$$

with periodic boundary conditions in the time direction understood.

This is the crucial result: the free energy F_q of a quark on the lattice, at temperature $T = 1/(L_t a)$, is infinite if (and only if) the Polyakov loop expectation value is zero, and the center symmetry is unbroken. We can therefore identify the unbroken Z_N center symmetry phase (the "disordered phase") with the confinement phase, and the Polyakov line is an order parameter for confinement. In general, center symmetry is broken spontaneously at high temperatures, so that $\langle P(\mathbf{x})\rangle \neq 0$ and isolated quarks are finite energy states. This means that quarks are unconfined, in the sense that there exist isolated, finite energy states with the flavor quantum numbers of quarks. The broken center symmetry phase is known as the *deconfined phase*, separated from the confined phase (in a center-symmetric theory) by a deconfinement phase transition which occurs at some critical temperature.[2]

---

[2] We note again that an order parameter for the spontaneous breaking of a global symmetry, such as $\langle P(\mathbf{x})\rangle$, is only non-zero, strictly speaking, in the infinite volume limit. In practice what is done is to compute, on each lattice, the magnitude of the sum of Polyakov loops of the lattice, divided by the number of loops (cf. 6.62) in (Chap. 6). The expectation value of this quantity, which is guaranteed to be positive, is then extrapolated to the infinite volume limit, and if the limit is vanishing, the symmetry is unbroken.

4.3 The 't Hooft Loop

The center symmetry transformation (4.12) can be generalized; it is a special case of a so-called "singular gauge transformation." Consider a transformation

$$U_0(x,t) \rightarrow G(x,t)U_0(x,t)G^\dagger(x,t+1) \tag{4.25}$$

in which $G(x, t)$ is non-periodic in the periodic time direction, or more precisely, is periodic only up to a center transformation

$$G(x, L_t + 1) = z^* G(x, 1). \tag{4.26}$$

This again transforms Polyakov lines $P(x) \rightarrow z\, P(x)$, but plaquettes and ordinary Wilson loops are not affected. In the continuum, the singular gauge transformation amounts to transforming the gauge field by the usual formula

$$A_\mu(x) \rightarrow G(x)A_\mu(x)G^\dagger(x) - \frac{i}{g}G(x)\partial_\mu G^\dagger(x), \tag{4.27}$$

except that at $t = L_t$ and $\mu = 0$ we drop the second term (which would be a delta-function). Because of this, a singular gauge transformation is not a true gauge transformation.

Instead of gauge transformations which are discontinuous on loops which wind around the periodic lattice, we could also consider transformations which are discontinuous on other sets of loops. Unlike the center symmetry transformations we have already discussed, these singular gauge transformations do not leave the action invariant; rather, they create a singular loop of magnetic flux known as a "thin center vortex." There is, in fact, a familiar example of such a vortex creation operator in classical electrodynamics: the exterior field of a solenoid is the result of a singular gauge transformation applied to $A_\mu = 0$.

So let's start with electrodynamics. The Wilson loop holonomy (which is an element in the $U(1)$ gauge group) is

$$U(C) = \exp\left[ie \oint_C dx^\mu A_\mu(x)\right] = e^{ie\Phi_B}, \tag{4.28}$$

where we take C to be a closed spacelike loop, and Φ_B is the magnetic flux through the loop. If the loop C winds around the exterior of a solenoid, it is certainly possible to have $B = 0$ everywhere along the loop, yet $\Phi_B \neq 0$. In the case of a solenoid of radius R, oriented along the z-axis, we may have

$$A_\theta = \frac{\Phi_B}{2\pi r}\,\widehat{\theta} \quad (r > R), \tag{4.29}$$

which is obtained from a singular gauge transformation of $A_\mu = 0$, with

$$G(r, \theta, z, t) = \exp\left[-ie\Phi_B \frac{\theta}{2\pi}\right]. \tag{4.30}$$

Note the discontinuity

$$G(r, \theta = 2\pi, z, t) = e^{-i\Phi_B} G(r, \theta = 0, z, t). \tag{4.31}$$

The gauge transformation is discontinuous on a surface (defined by $y = 0$, $x > 0$, all z) with a boundary at the z-axis. In obtaining A from the gauge transformation g, we drop the delta-function that would arise in A_θ due to the discontinuity of G on the $y = 0$, $x > 0$ surface, and which would have to be included if this were a true gauge transformation.

Suppose the gauge transformation $G(r, \theta, z, t)$ has the discontinuity (4.31) for any $r > 0$. Then the effect of the singular gauge transformation on any loop winding once around the z-axis is

$$U(C) \rightarrow e^{\pm ie\Phi_B} U(C), \tag{4.32}$$

where the sign in the exponent depends on whether the loop runs clockwise or counterclockwise around the z-axis. From (4.32) and (2.44) it is clear that the singular gauge transformation has created a line of magnetic flux along the z-axis.

We can easily generalize this construction, to obtain an operator which creates a line of magnetic flux running along a closed curve C'. Let S be some surface bounded by the loop C', and let C be a loop which is topologically linked to C', parametrized by $\mathbf{x}(\tau)$, where τ runs from $\tau = 0$ to 1, and $\mathbf{x}(0) = \mathbf{x}(1)$ lie on S. Then we consider a singular gauge transformation $G(\mathbf{x})$ which is discontinuous on S; i.e. for any loop C linked to C'

$$G(\mathbf{x}(1)) = e^{-ie\Phi_B} G(\mathbf{x}(0)). \tag{4.33}$$

Once again, for such loops the singular gauge transformation changes the loop holonomy according to (4.32), so we must conclude that a singular magnetic field has been created along the loop C'.

"Linking" is a topological concept: two loops are linked if it is impossible to translate one loop an arbitrary distance from the other loop without the two objects actually crossing one another. Loops can be linked in three dimensions, but not in four. Suppose two loops C_1 and C_2 appear to be linked at a given time t. The loops can be separated an arbitrary distance, without crossing, by simply translating one of them in time. So the sorts of objects that can be linked depends on spacetime dimension, as indicated in Fig. 4.1, and the classification is as follows:

1. In $D = 2$ dimensions, loops can link to points. A singular gauge tranformation is discontinuous along a line ending at a point, and any loop linked to the point is changed by the transformation according to (4.32). The transformation creates a singular magnetic field (a thin vortex) at a point.
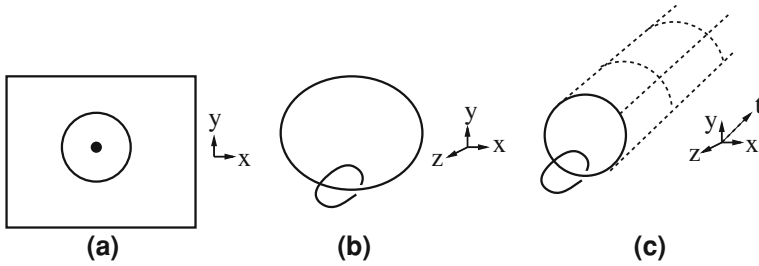
Fig. 4.1 A loop can be topologically linked to: **a** a point, in $D = 2$ dimensions; **b** another loop, in $D = 3$ dimensions; **c** a surface, in $D = 4$ dimensions

2. In $D = 3$ dimensions, loops link to loops. A singular gauge transformation, discontinuous on a surface S bounded by loop C', creates a singular magnetic field (thin vortex) along loop C'. Again, a holonomy along loop C, linked to C', is changed according to (4.32).
3. In $D = 4$ dimensions loops link to surfaces. A singular gauge transformation, discontinuous on a 3-volume V_3 which is bounded by a surface S, creates a surface of magnetic flux on S. A holonomy along loop C, linked to the surface S, is changed according to (4.32).

The general case, of course, follows by induction: In D dimensions, loops link to $D - 2$ dimensional hypersurfaces. Now in $D = 2$ dimensions a loop can wind more than once around a point; the number of times that the loop goes around the point is known as the *winding number*. In $D = 3$ dimensions two loops can obviously wind more than once around each other, and the topological invariant which describes this intertwining is known as the *linking number*. The concept of linking number generalizes to describe the topological connectivity of a surface and a loop in $D = 4$ dimensions, but for now it will be sufficient to consider only connections with linking number = 1.

Returning to $U(1)$ gauge theory, we see that a singular gauge transformation which is discontinuous on a region of dimension $D - 1$, bounded by a hypersurface of dimension $D - 2$, creates a singular magnetic field (a thin vortex) on the $D - 2$ dimensional hypersurface. The $D - 1$ dimensional region of discontinuity is known as a Dirac line, surface, 3-volume, etc., depending on the dimension D. The same procedure can be applied to define a magnetic vortex creation operator in $SU(N)$ theories, but with one importance difference as compared to abelian gauge theories. In an abelian gauge theory, the discontinuity (4.33) consists of multiplication by *any* element $\exp[ie\Phi_B]$ of the abelian gauge group. In a non-abelian theory, if the singular gauge transformation is applied to a non-zero field configuration, this transformation would not only change the action at the location of the $D - 2$ dimensional thin vortex, but would also affect the action all along the $D - 1$ dimensional volume of discontinuity. To avoid changing the action on the discontinuity volume, in the same way that this is avoided in (3.20), and create *only* a magnetic vortex on the $D - 2$ dimensional surface, the discontinuity must be restricted to multiplication by a center element; i.e.

$$G(\mathbf{x}(1)) = zG(\mathbf{x}(0)), \quad z \in Z_N \tag{4.34}$$

for $SU(N)$ gauge theories. Then any loop holonomy $U(C)$ which is topologically linked to the vortex, with linking number = 1, will transform as

$$U(C) \rightarrow z^* U(C). \tag{4.35}$$

For linking number l, $U(C) \rightarrow (z^*)^l \, U(C)$. Since z is a center element, the non-abelian magnetic "flux" created by such singular gauge transformations is quantized, and the magnetic objects themselves are known as "thin center vortices." As in QED, the singular magnetic field on the $D - 2$ dimensional hypersurface can be smeared out in the directions transverse to the hypersurface into a region of finite thickness, much like the field of a solenoid. This is a *thick* center vortex.

The center vortex creation operator is closely related to the center symmetry transformation (4.12). In $D = 2$ dimensions we can create a thin vortex at the plaquette at site x_0, y_0 by the following operation (Fig. 4.2):

$$U_y(x, y_0) \rightarrow z U_y(x, y_0) \quad \text{for } x > x_0. \tag{4.36}$$

In higher dimensions, the operation

$$U_y(x, y_0, x_\perp) \rightarrow z U_y(x, y_0, x_\perp) \quad \text{for } x > x_0 \tag{4.37}$$

creates a line of magnetic flux parallel to the z-axis ($D = 3$), or a surface of magnetic flux parallel to the $z - t$ plane ($D = 4$).

In the Hamiltonian formulation of $3 + 1$ dimensional gauge theory, we may consider an operator $B(C)$ which creates a thin center vortex along loop C at a particular time t. Then if C and C' are loops at a given time t, and if the two loops are linked (with unit linking number) in the $D = 3$ dimensional timeslice, we have

$$B(C)U(C') = zU(C')B(C), \quad z \in Z_N. \tag{4.38}$$

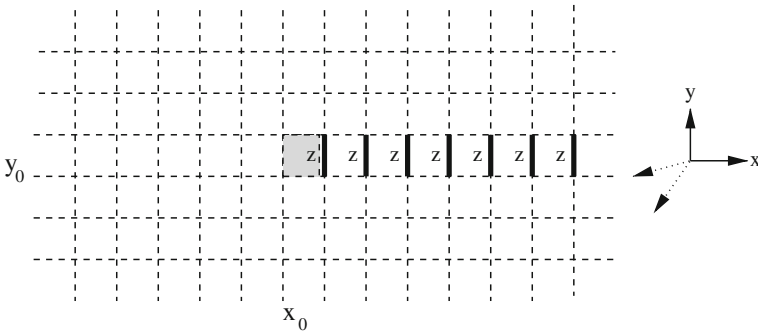Using only this relation, 't Hooft [4] argued that

Fig. 4.2 Creation of a thin center vortex. The shaded plaquette, and all other x–y plaquettes at sites $(x_0, y_0, \mathbf{x}_\perp)$ form the center vortex. The stack of vortex plaquettes lie along a line in $D = 3$ dimensions, or a surface in $D = 4$ dimensions

- only an area-law or a perimeter-law falloff for $W[C]$ and $\langle B[C] \rangle$ is possible; and
- in the absence of massless excitations, it is impossible to have both

$$W(C) \sim e^{-aP(C)} \quad \text{and} \quad \langle B(C) \rangle \sim e^{-bP(C)}. \tag{4.39}$$

So a perimeter law falloff for $B(C)$ implies an area-law falloff for $W(C)$, i.e. confinement. $B(C)$ is known as the *'t Hooft loop operator*. It is a creation operator for thin center vortices, which (at fixed time t) are loops of color magnetic flux, and can be thought of as being in some sense "dual" to the Wilson loop operator, which (at fixed time) creates a closed loop of color electric flux.

4.4 The Vortex Free Energy

't Hooft devised another procedure for inserting a center vortex into a finite lattice [5]. Consider $SU(N)$ gauge theory on a finite two-dimensional lattice. Pick one plaquette p', with

$$U(p') = U_1(x_0, y_0)U_2(x_0 + 1, y_0)U_1^\dagger(x_0, y_0 + 1)U_2^\dagger(x_0, y_0), \tag{4.40}$$

and modify the action by replacing $U(p')$ by $zU(p')$, where $z \in Z_N$, i.e.

$$S = -\frac{\beta}{2N}\left\{ \sum_{p \neq p'}(\text{Tr}[U(p)] + c.c.) + (\text{Tr}[zU(p')] + c.c.) \right\}. \tag{4.41}$$

On a finite lattice this replacement can be formulated as a modification of the ordinary periodic boundary conditions, and is known as "twisted boundary conditions." Twisted boundary conditions can also be regarded as a shift $\beta \rightarrow z\beta$ in the coupling constant at the plaquette p'. It is clear that setting all links equal to the identity matrix (or a gauge transformation of the identity matrix) no longer minimizes the action. On an infinite lattice, the action would be minimized by any configuration gauge-equivalent to

$$U_2(x, y_0) = z\mathbb{1}_N \quad \text{for } x \leq x_0, \tag{4.42}$$

with all other links set equal to the $N \times N$ identity matrix $\mathbb{1}_N$. This is a Dirac line, running parallel to the x-axis from $x = x_0$ to $x = -\infty$. On a finite lattice the line has to end somewhere; it must end on a center vortex. There is no need, however, for the line to end on a thin center vortex, with all of the center flux contained in a single plaquette. That arrangement costs a great deal of action at large β. The action is lowered if the center flux is spread over many plaquettes, in a region of some finite extension.

In $D = 4$ dimensions, the twisted boundary conditions involve replacing $U(p')$ by $zU(p')$ in the action, for all plaquettes p' oriented in the $x - y$ plane, at fixed x_0, y_0 and all z, t, i.e.

$$U(p') = U_1(x_0, y_0, z, t)\ U_2(x_0 + 1, y_0, z, t)\ U_1^\dagger(x_0, y_0 + 1, z, t)\ U_2^\dagger(x_0, y_0, z, t).$$

(4.43)

This replacement then creates a thick center vortex sheet somewhere on the lattice, oriented parallel to the $z - t$ plane and closed by ordinary periodic boundary conditions in the z and t directions. The actual thickness of the vortex is not determined by the $U(p') \to zU(p')$ replacement; it is instead a dynamical issue. Of course, there is an upper limit to the thickness of a center vortex on a finite lattice, the limit is the cross-sectional area of the lattice in the $x - y$ plane.

For simplicity, consider the SU(2) gauge group, with $z = -1$ for twisted boundary conditions. We define Z_+ as the lattice partition function with ordinary boundary conditions, and Z_- as the partition function with twisted boundary conditions. The magnetic free energy F_{mg} of a Z_2 vortex is defined by

$$e^{-F_{mg}} = \frac{Z_-}{Z_+}.$$

(4.44)

The electric free energy F_{el} is defined by a Z_2 Fourier transform

$$\begin{aligned} e^{-F_{el}} &= \sum_{z=\pm} z \frac{Z_z}{Z_+} \\ &= 1 - e^{-F_{mg}}. \end{aligned}$$

(4.45)

Let C be a rectangular loop of area $A[C]$. The following inequality was proven by Tomboulis and Yaffe in [6]

$$W(C) \le \{\exp[-F_{el}]\}^{A(C)/(L_x L_y)}.$$

(4.46)

So if the vortex magnetic free energy falls off with cross-sectional area $L_x L_y$ as

$$F_{mg} \sim L_z L_t e^{-\rho L_x L_y},$$

(4.47)

then for large L_x, L_y we have $F_{mg} \ll 1$ and $\exp[-F_{el}] \approx F_{mg}$. The behavior (4.47) is therefore a sufficient condition for confinement, because it implies, from (4.46), an area law bound for Wilson loops.

Numerical investigations of these quantities were begun by Kovács and Tomboulis [7]; their computation of Z_-/Z_+ is shown in Fig. 4.3, and is consistent with the vortex magnetic free energy falling very rapidly to zero on large lattice volumes, in accordance with the condition (4.47). Much additional work along these lines was carried out by von Smekal and de Forcrand [8–10]. Note that, in Fig. 4.3, the ratio Z_-/Z_+ attains its limiting value of unity for lattice extensions of roughly

Fig. 4.3 The behavior of the vortex free energy versus lattice extension (same in all directions). From Kovács and Tomboulis [7]
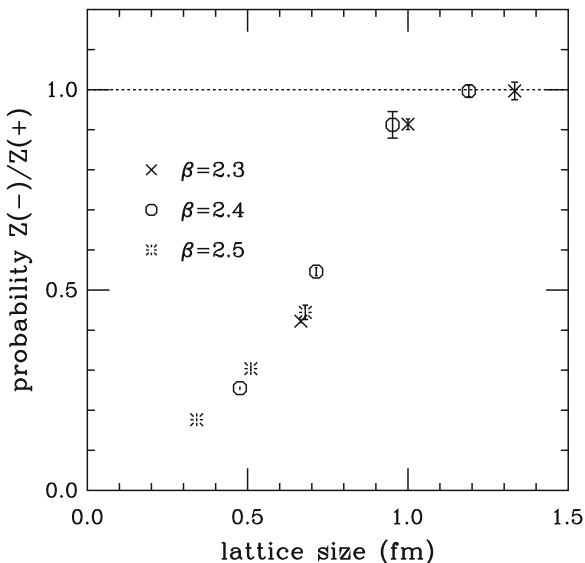
1.25 fm. It appears that the free energy of a vortex is lowered by allowing the vortex to spread out to this extent, but there is no advantage in spreading to a thickness much beyond that. This suggests that the thickness of a center vortex on a large lattice is about 1.25 fm.

4.5 Summary

To summarize the lessons of this chapter: We have seen that the existence of a non-vanishing asymptotic string tension requires that the gauge group has a non-trivial center, and that any matter fields must transform trivially with respect to the center subgroup. When these two conditions are satisfied, the action of the gauge theory is invariant with respect to a global center symmetry transformation, which is not a gauge transformation. All of the signals we have seen for confinement

- an area law for Wilson loops;
- vanishing Polyakov lines;
- perimeter-law 't Hooft loops;
- area-law falloff for the vortex magnetic free energy

can only be satisfied if a non-trivial global center symmetry exists, and is not spontaneously broken.

   This motivates the idea that the vacuum fluctuations responsible for the asymptotic string tension must be associated, in some way, with center symmetry.

References

1. Wilson, K.G.: Confinement of quarks. Phys. Rev. D **10**, 2445–2459 (1974)
2. Narayanan, R., Neuberger, H.: Infinite N phase transitions in continuum Wilson loop operators. JHEP **0603**, 064-1–064-31 (2006) [arXiv:hep-th/0601210]
3. Polyakov, A.M.: Compact gauge fields and the Infrared Catastrophe. Phys. Lett. B **59**, 82–84 (1975)
4. 't Hooft, G.: Compact gauge fields and the Infrared Catastrophe. Nucl. Phys. B **138**, 1–25 (1978)
5. 't Hooft, G.: A property of electric and magnetic flux in non-Abelian gauge theories. Nucl. Phys. B **153**, 141–160 (1979)
6. Tomboulis, E., Yaffe, L.: Finite temperature SU(2) lattice gauge theory. Commun. Math. Phys. **100**, 313–341 (1985)
7. Kovács, T., Tomboulis, E.: Computation of the vortex free energy in SU(2) gauge theory. Phys. Rev. Lett. **85**, 704–707 (2000) [arXiv:hep-lat/0002004]
8. von Smekal, L., de Forcrand, Ph.: 't Hooft loops, electric flux sectors and confinement in SU(2) Yang–Mills theory. Phys. Rev. D **66**, 011504-1–011504-5 (2002) [arXiv: hep-lat/0107018]
9. von Smekal, L., de Forcrand, Ph.: Electric and magnetic fluxes in SU(2) Yang–Mills theory. Nucl. Phys. Proc. Suppl. **119**, 655–657 (2003) [arXiv: hep-lat/0209149]
10. von Smekal, L., de Forcrand, Ph., Jahn, O.: More on electric and magnetic fluxes in SU(2) [arXiv: hep-lat/0212019]

Chapter 5
Properties of the Confining Force

Any fully satisfactory theory of confinement ought to explain the following features of the confining force, for which there is strong theoretical and/or numerical support:

1. asymptotic linearity of the static potential;
2. Casimir scaling of string tensions at intermediate distance scales;
3. N-ality dependence of asymptotic string tensions;
4. evidence of quantum string-like behavior: roughening and the Lüscher term.

Let's begin with asymptotic linearity. It was proven by Bachas in [1] that the static quark potential must be everywhere convex, i.e.

$$\frac{dV}{dR} > 0 \quad \text{and} \quad \frac{d^2V}{dR^2} \leq 0. \tag{5.1}$$

The inequality is saturated by a linear potential; $V(R)$ can be bounded from above by a straight line. Monte Carlo simulations provide very convincing evidence that the static quark potential is indeed linear asymptotically. A sample numerical calculation of $V(R)$ in $SU(3)$ gauge theory, in units of the "Sommer scale" $r_0 \approx$ 0.5 fm, is shown in Fig. 5.1.

Casimir scaling is the property that at some intermediate range of distances, the slope of the static quark potential for a quark in group representation r, and its antiquark, is approximately proportional to the quadratic Casimir C_r of the representation; i.e.

$$\sigma_r = \frac{C_r}{C_F}\sigma_F, \tag{5.2}$$

where F denotes the fundamental, defining representation. This relationship is certainly true at weak couplings in $D = 2$ dimensional gauge theory, where Wilson loops can be calculated from one-gluon exchange. The proposal that Casimir scaling might also hold in higher dimensions [3] is based on the notion of "dimensional reduction," which suggests that the calculation of a large planar

Fig. 5.1 The static quark
potential in SU(3) lattice
gauge theory, normalized to
$V(r_0) = 0$. The reference
distance r_0 is the "Sommer
scale," which roughly 0.5 fm.
From Bali [2]
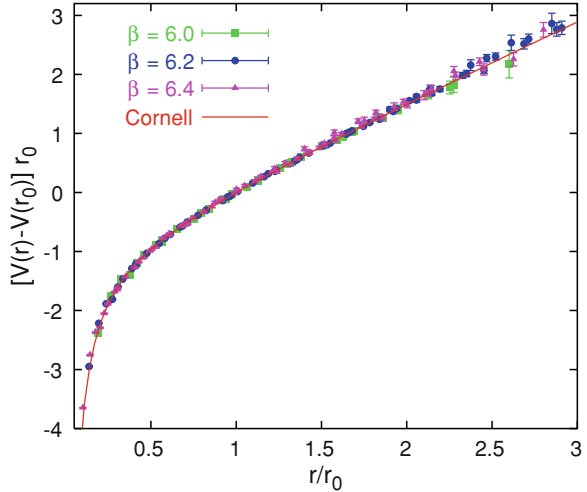
Wilson loop in $D = 4$ dimensions reduces somehow to the corresponding calcu-
lation in $D = 2$ dimensions. At strong couplings, dimensional reduction is
demonstrably correct; the leading term for the string tension derived from the
strong-coupling expansion is the two-dimensional result. On the other hand, the
non-leading terms *do* depend on dimension. Arguments for why dimensional
reduction might hold nonetheless, at weak couplings, were advanced indepen-
dently by Olesen [4] and myself [5, 6], and the interested reader can consult those
references. In any case, Casimir scaling in $SU(2)$ and $SU(3)$ gauge theory seems to
hold with remarkable precision. Figure 5.2 shows the data for string tensions of the
static quark potential, in $SU(3)$ lattice gauge theory, in various representations of
the color group. The solid lines shown for each representation are obtained by
multiplying a fit to the static potential of the fundamental (3) representation by the
ratio of quadratic Casimirs C_r/C_3.[1] It can be argued that Casimir scaling must hold
exactly for the $SU(\infty)$ gauge group; we will reserve that argument for Chapter 11,
which is concerned with the large-N limit.

Casimir scaling cannot continue out to arbitrary quark-antiquark separations; apart
from $D = 2$ dimensions or $N = \infty$ colors it is an intermediate distance phenomenon.
Consider a quark-antiquark pair in representation r, with N-ality k_r. Gluons can bind to
the quark and antiquark, and reduce the color charge to the lowest-dimensional rep-
resentation with the same N-ality. It follows that after screening by gluons (i.e.
asymptotically), string tensions depend only on the N-ality of the representation

$$\sigma_r = \sigma(k_r). \tag{5.3}$$

---

[1] It should be noted that the calculation on which this figure is based uses a method which creates
metastable flux tubes, which are then allowed to propagate for a relatively short Euclidean time
interval. This procedure is insensitive to the string-breaking process, and hence one can only
calculate the string tension of the metastable states.

Fig. 5.2 Numerical evidence for Casimir scaling in $SU(3)$ lattice gauge theory. The solid lines are obtained from a fit of the potential in fundamental representation, multiplied by a ratio of quadratic Casimirs C_r/C_F. The reference distance r_0 is again about 0.5 fm. From Bali [7]
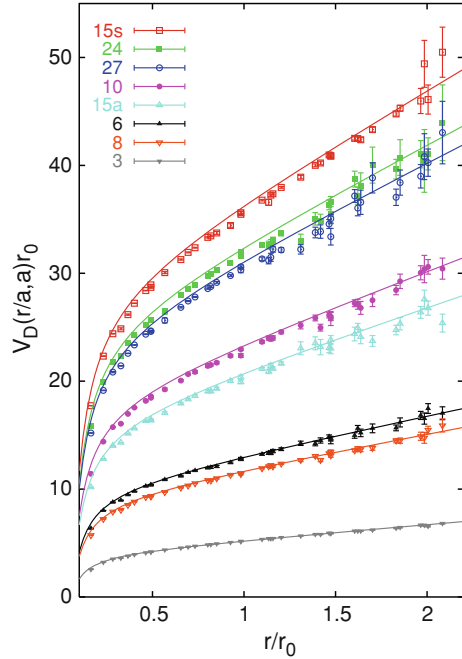
String tensions of the lowest-dimensional representation of N-ality k are often called "k-string tensions."

It is not known precisely how k-string tensions depend on k; two proposals are so-called "Casimir scaling," and Sine-law scaling. "Casimir scaling," in this context, is something of a misnomer. The term as originally introduced by Del Debbio, Faber, Olejnik and myself [8, 9] refers to intermediate string tensions (prior to screening by gluons), and is defined in (5.2) above. Somewhat confusingly, Casimir scaling has also been used in the literature in connection with k-string tensions, which refer to string tensions at asymptotic, rather than intermediate, distances. In that usage, the Casimir C_r is replaced by the Casimir of the lowest dimensional representation with the same N-ality as r. Then, for an $SU(N)$ gauge theory, the "Casimir scaling" prediction is

$$\sigma_r = \sigma(k_r) = \frac{k(N-k)}{N-1}\sigma_F, \tag{5.4}$$

where σ_F is the string tension for quarks in the fundamental, defining representation.

An alternative proposal, motivated by supersymmetric gauge theories, is the Sine-law dependence

$$\sigma_r = \sigma(k_r) = \frac{\sin(\pi k/N)}{\sin(\pi/N)}\sigma_F. \tag{5.5}$$

In practice, the Casimir scaling and Sine-law values for the k-string tensions are not very different, and there has been some debate in the literature over whether the Sine law is favored. For our purposes, the crucial point is not Casimir scaling vs. the Sine Law, but rather the simple fact that asymptotic string tension depends only on the N-ality of the representation.

An important example is the asymptotic string tension of the adjoint representation, which has $k = 0$. Gluons can screen adjoint color charges and the confining string should break, leaving two bound states known as "gluelumps", which can be thought of as a gluon bound to a static quark in the color adjoint representation. Suppose the gluonic energy of a gluelump state is m_{GL}. Then the adjoint string should break for quark separations R such that

$$2m_{GL} < \sigma_A R, \tag{5.6}$$

where σ_A is the adjoint string tension, which is approximately equal to $C_A \, \sigma_F/C_F$ in the intermediate region, according to Casimir scaling. Beyond the string-breaking distance the static quark potential is flat, and equal to $2m_{GL}$.

To observe string-breaking (i.e. color screening by gluons) for adjoint representation heavy quark states, using only rectangular Wilson $R \times T$ Wilson loops, is computationally demanding, but was achieved in numerical simulations by de Forcrand and Kratochvila [10], using a noise reduction technique due to Lüscher and Weisz [11]. The result of their calculation, for $SU(2)$ lattice gauge theory in $D = 3$ dimensions, is shown in Fig. 5.3. The adjoint string breaks at about ten lattice spacings, corresponding (at the given value of the lattice coupling) to 1.25 fm in physical units. We can see that Casimir scaling of the adjoint potential is fairly good up to 10 lattice spacings, at $\beta = 6.0$, after which the potential abruptly flattens out.

Although I have used the terms "QCD string" and "string breaking," there is the question of whether, or to what extent, the QCD flux tube really behaves like a quantum string as described by string theory. If the QCD flux tube resembles a Nambu string (see Chapter 13), then transverse fluctuations of the string induce a

Fig. 5.3 The adjoint and $\frac{8}{3}\times$ fundamental static potentials $V(R)$ vs R, in $D = 3$ dimensional $SU(2)$ lattice gauge at $\beta = 6.0$. The horizontal line at 2.06(1) represents twice the energy of a gluelump. From de Forcrand and Kratochvila [10]
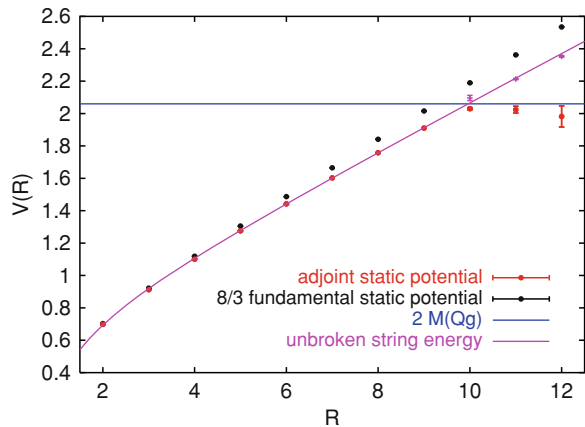
universal (coupling and scale independent) $1/R$ modification to the linear potential [12, 13] in D-dimensions

$$V(R) = \sigma R - \frac{\pi(D-2)}{24}\frac{1}{R} + c. \tag{5.7}$$

This modification is known to lattice gauge theorists as the "Lüscher term."[2] Another prediction of the Nambu model is that the cross-section of the QCD flux tube is not actually constant, but should grow logarithmically with quark separation, a phenomenon known as "roughening" [14, 15]. There is some evidence in favor of both the Lüscher term [11, 16] and logarithmic growth of the flux tube cross-section [17], as well as a spectrum of flux tube excited states which also appears to correspond to that of a string theory [18]. Perhaps the most accurate calculations to date [19] have been carried out in $2 + 1$ dimensions, for the spectrum of closed flux tubes of length R which wind once around the periodic lattice in a spatial direction. Both the R-dependence of the energy of the ground state (which includes the Lüscher term), and the pattern of excited flux tube states, seem to agree rather well with predictions of the Nambu string action.

References

1. Bachas, C.: Convexity of the Quarkonium Potential. Phys. Rev. D **33**, 2723–2725 (1986)
2. Bali. G.: QCD forces and heavy quark bound states. Phys. Rept. **343**, 1–136 (2001) [arXiv: hep-ph/0001312]
3. Ambjorn, J., Olesen, P., Peterson, C.: Stochastic confinement and dimensional reduction (I): four-dimensional SU(2) lattice gauge theory. Nucl. Phys. B **240**, 189–212; 533–542 (1984)
4. Olesen, P.: Confinement and random fluxes. Nucl. Phys. B **200**(FS4), 381–390 (1982)
5. Greensite, J.: Calculation of the Yang–Mills vacuum wave functional. Nucl. Phys. B **158**, 469–496 (1979)
6. Greensite, J.: Large scale vacuum structure and new calculational techniques in lattice SU(N) gauge theory. Nucl. Phys. B **166**, 113–124 (1980)
7. Bali, G.: Casimir scaling of SU(3) static potentials. Phys. Rev. D **62**, 114503-1–114503-11 (2000) [arXiv: hep-lat/0006022]
8. Del Debbio, L., Faber, M., Greensite, J., Olejník, Š.: Casimir scaling versus Abelian dominance in QCD string formation. Phys. Rev. D **53**, 5891–5897 (1996) [arXiv: hep-lat/9510028]
9. Del Debbio, L., Faber, M., Greensite, J., Olejník, Š.: Some cautionary remarks on Abelian projection and Abelian dominance. Nucl. Phys. Proc. Suppl. **53**, 141–147 (1997) [arXiv: hep-lat/9607053]
10. de Forcrand, P., Kratochvila, S.: Observing string breaking with Wilson loops. Nucl. Phys. B **671**, 103–132 (2003) [arXiv: hep-lat/0209094]
11. Lüscher, M., Weisz, P.: Quark confinement and the bosonic string. JHEP **07**, 049-1–049-18 (2002) [arXiv: hep-lat/0207003]

---

[2] To add to the "Casimir" confusion, the Lüscher term is known to string theorists as the "Casimir energy."

12. Lüscher, M.: Symmetry breaking aspects of the roughening transition in gauge theories. Nucl. Phys. B **180**(FS2), 317–329 (1981)
13. Alvarez, O.: The static potential in string models. Phys. Rev. D **24**, 440–449 (1981)
14. Lüscher, M., Münster, G., Weisz, P.: How thick are chromoelectric flux tubes? Nucl. Phys. B **180**(FS2), 1–12 (1981)
15. Hasenfratz, A., Hasenfratz, E., Hasenfratz, P.: Generalized roughening transition and its effect on the string tension. Nucl. Phys. B **180**(FS2), 353–367 (1981)
16. Juge, K.J., Kuti, J., Morningstar, C.: QCD string formation and the Casimir energy [arXiv: hep-lat/0401032]
17. Bornyakov, V.G., Kovalenko, A.V., Polikarpov, M.I., Sigaev, D.A.: Confining string and P vortices in the indirect Z(2) projection of SU(2) lattice gauge theory. Nucl. Phys. Proc. Suppl. **119**, 739–741 (2003) [arXiv:hep-lat/0209029]
18. Kuti, J.: Lattice QCD and string theory. PoS LAT **2005**, 1–22 (2006) [arXiv:hep-lat/0511023]
19. Athenodorou, A., Bringoltz, B., Teper, M.: The closed string spectrum of SU(N) gauge theories in 2 + 1 dimensions. Phys. Lett. B **656**, 132–140 (2007) [arXiv:0709.0693 [hep-lat]]

Chapter 6
Confinement from Center Vortices I

We have already seen a number of indications that the center of the gauge group has something to do with confinement. For one thing, the existence of a non-vanishing asymptotic string tension is tied to the existence, and unbroken realization, of a global center symmetry. Moreover, two of the order parameters for confinement, namely the 't Hooft loop $B(C)$ and vortex free energy, are explicitly associated with center vortex creation. But perhaps the strongest motivation for the center vortex mechanism of confinement, to be discussed below, comes from the fact that the asymptotic string tension depends only on the N-ality of the quark charge. Whatever vacuum fluctuations of the gauge field are responsible for an asymptotic string tension for quarks in the fundamental representation, those same fluctuations should not also give rise to an asymptotic string tension for quarks in the adjoint representation, or in any other zero N-ality representation of the gauge group.

6.1 The Mechanism

There is, of course, no great mystery in the fact that the adjoint string tension vanishes asymptotically. We have already seen why that must be true. As the flux tube gets longer, at some stage it is energetically favorable to pair-create some gluons which bind to each adjoint quark to form a singlet, and the flux tube breaks. This explanation is certainly correct, but when we talk of pair-creating gluons out the vacuum, we are using the language of Feynman diagrams, and talking about particle states. However, the path-integral itself is a sum over field configurations. When a Wilson loop is evaluated at a given location by Monte Carlo methods, the configurations which are generated have no knowledge whatever of the location of the Wilson loop being measured, or where the pair-created gluons should appear.

The Wilson loop is just an observable, not a part of the action. Nevertheless, the Monte Carlo evaluation must somehow give us the same answer that we deduce, on the basis of energetics, from the particle picture of string breaking. So this raises the question: What feature of a typical vacuum field fluctuation can give us both a finite string tension for quarks in the fundamental representation, and a vanishing string tension for zero N-ality quark charges, at large distances? What we are looking for is a "field" explanation of screening, to complement the particle picture.

Suppose that there is a certain type of solitonic object, identifiable in typical field configurations, whose fluctuations in position (or some other degree of freedom) give rise to an area law for Wilson loops. In order that the string tension depend only on N-ality, it seems necessary that Wilson loops which are in different representations, but have the same N-ality, should be affected by these solitonic objects in exactly the same way. Loops of zero N-ality should not be affected at all. These requirements are almost a definition of center vortices; I do not know of any other object with the same properties. Suppose we create a center vortex linked to loop C. This affects the loop holonomy as follows: $U(C) \rightarrow zU(C)$ where $z \in Z_N$ for an $SU(N)$ gauge group. Then, for a Wilson loop in representation r with N-ality k

$$W_r(C) \rightarrow z^k W_r(C), \tag{6.1}$$

so that Wilson loops of the same N-ality are, indeed, affected in the same way, while zero N-ality loops are unchanged. Creation of a set of vortices, with center elements z_a, z_b, z_c..., linked to loop C, with linking numbers l_a, l_b, l_c ... respectively, will transform the loop as

$$W_r(C) \rightarrow Z^k(C)W_r(C)$$
$$Z(C) = z_a^{l_a} z_b^{l_b} z_c^{l_c}\dots. \tag{6.2}$$

The center vortex scenario for confinement [1–6] is very simple: Any vacuum (thermalized lattice) configuration can be viewed as a set of center vortices superimposed on a non-confining configuration, and the area law for Wilson loops is obtained from random fluctuations in the number of vortices (and their linking numbers) topologically linked to the loop. More explicitly, we can write for any holonomy

$$U(C) = Z(C)u(C), \tag{6.3}$$

where $u(C)$ is the contribution to the holonomy from the non-confining background. Then the idea is that

$$\langle \chi_r[U(C)] \rangle \approx \langle Z^k(C) \rangle \langle \chi_r[u(C)] \rangle$$
$$\approx e^{-\sigma_k A(C)} e^{-\mu_r P(C)}. \tag{6.4}$$

To get this result, we only need that

(A) For a single large loop C, $Z(C)$ and $u(C)$ are weakly correlated; and
(B) for two large loops C_1 and C_2, $Z(C_1)$ and $Z(C_2)$ are weakly correlated; i.e.

$$\langle Z(C_1)Z(C_2)\rangle \approx \langle Z(C_1)\rangle\langle Z(C_2)\rangle \tag{6.5}$$

and in fact this weak correlation applies to the product of any number of large loops lying in the same plane.

   To see how we arrive at an area law from these assumptions, consider a large rectangular Wilson loop C of area A, in group representation r of N-ality k. We have, by assumption A,

$$W_r(C) = \langle \chi_r[U(C)]\rangle = \langle Z^k(C)\rangle\langle \chi_r[u(C)]\rangle. \tag{6.6}$$

Now subdivide the area A into square $L \times L$ subareas bounded by loops $\{C_i\}$, so that

$$\begin{aligned} \langle Z^k(C)\rangle &= \left\langle \prod_{i=1}^{A/L^2} Z^k(C_i) \right\rangle \\ &\approx \prod_{i=1}^{A/L^2} \langle Z^k(C_i)\rangle \\ &= \exp[-\sigma(k)A(C)], \end{aligned} \tag{6.7}$$

where assumption B is used in going from the first to second line, and

$$\sigma(k) = -\frac{\log\big[\langle Z^k(C_i)\rangle\big]}{L^2} \tag{6.8}$$

for any of the $L \times L$ subloops C_i. Solving for $\langle Z^k(C_i)\rangle$, we have

$$\langle Z^k(C_i)\rangle = \exp[-\sigma(k)L^2]. \tag{6.9}$$

which means that the string tension extracted from the smaller loops is identical to that of the larger loop, providing the smaller loops are large enough so that assumption B holds, i.e. providing the larger loop can be subdivided into smaller loops such that the $Z(C_i)$ fluctuate independently. Note that we have simply applied the reasoning given in Sect. 2.4 to the center-valued loops $Z(C)$.

   It is interesting to consider the limiting case, where $Z(p)$ fluctuates independently from one plaquette p on the lattice to another, for plaquettes lying in the same plane. Start from

$$Z(C) = \prod_{p\in A(C)} Z(p), \tag{6.10}$$

where $A(C)$ is the minimal area of the loop. For simplicity consider the group $SU(2)$, where

$$Z(p) = \begin{cases} -1 & \text{if the vortex pierces the plane at } p \\ +1 & \text{otherwise} \end{cases}, \qquad (6.11)$$

and define f to be the probability that $Z(p) = -1$ on any given plaquette p. Then, assuming the $Z(p)$ fluctuate independently,

$$\langle Z(C) \rangle = \left\langle \prod_{p \in A(C)} Z(p) \right\rangle = \prod_{p \in A(C)} \langle Z(p) \rangle$$

$$= \prod_{p \in A(C)} [(1-f)(+1) + f(-1)]$$

$$= \exp[A(C) \ln(1 - 2f)], \qquad (6.12)$$

and the string tension is

$$\sigma = -\ln(1 - 2f). \qquad (6.13)$$

Now, what kind of link configurations produce holonomies, for any C, which are Z_N elements $U(C) = Z(C) \in Z_N$? The answer is that these are configurations which can be transformed, by an $SU(N)$ gauge transformation, into link configurations of Z_N lattice gauge theory, i.e.

$$\mathcal{U}_\mu(x) = z_\mu(x) g(x) g^{-1}(x + \hat{\mu}), \quad z_\mu(x) \in Z_N. \qquad (6.14)$$

The excitations of Z_N lattice gauge theory are precisely thin Z_N vortices. Then we may subdivide link variables into degrees of freedom responsible for confinement $z_\mu(x)$, and those which are simply a non-confining background $V_\mu(x)$, as follows[1]

$$U_\mu(x) = g(x) z_\mu(x) V_\mu(x) g^{-1}(x + \hat{\mu}). \qquad (6.15)$$

This form motivates us to look for the gauge transformation $g(x)$ which would bring a lattice configuration $\mathcal{U}_\mu(x)$ into the form of a simple product $U = zV$. If V, the non-confining background, is (mainly) a small fluctuation around the unit element then we can easily determine the large fluctuations $z_\mu(x)$, and, from those, the center vortex locations.

6.2 Center Gauges and Center Projection

The vortex mechanism is probably the simplest route to confinement, and is well motivated by the local, gauge-invariant order parameters for confinement ('t Hooft

---

[1] $V_\mu(x)$ would also be responsible for "thickening" the flux of the thin vortices, so there must be some short range correlation between $V_\mu(x)$ and vortex position.

loop, Polyakov loop, vortex free energy), and by the known facts about N-ality dependence. But is it right?

   To find out, we turn to lattice Monte Carlo simulations. The first problem is to figure out how to spot thick center vortices in a list of what looks like random numbers, i.e. the lattice link variables.

6.2.1 Direct Maximal Center Gauge

We would like to transform the link variables to a gauge in which link variables can be factored into center elements, whose fluctuations bring about the asymptotic string tension, and non-confining fluctuations $V_\mu(x)$, which have no influence on the far-infrared physics. But what gauge is that? The most intuitive choice [7] is to transform the configuration so that the link variables $U_\mu(x)$ are as close as possible to some Z_N lattice configuration or, to put it another way, the link variables in the adjoint representation are as close as possible to unity. The center elements of $SU(N)$ are all mapped to the identity in the adjoint representation, so the deviation of adjoint representation links from the identity matrix is a measure of the deviation of the lattice configuration from a pure Z_N configuration. The gauge in which this deviation is minimized is known as *"direct maximal center gauge."* It is defined as the gauge which maximizes

$$R = \sum_x \sum_\mu \text{Tr}[U_{A\mu}(x)]$$
$$= \sum_x \sum_\mu \text{Tr}[U_\mu(x)]\text{Tr}[U_\mu^\dagger(x)] - 1, \qquad (6.16)$$

where $U_{A\mu}$ is the link variable in the adjoint representation. This is simply Landau gauge in the adjoint representation. In this gauge we decompose

$$U_\mu(x) = Z_\mu(x)V_\mu(x), \qquad (6.17)$$

where $Z_\mu(x)\mathbb{1}_N$ is the center element in the $SU(N)$ group manifold which is closest to $U_\mu(x)$. For example, in $SU(2)$,

$$Z_\mu(x) = \text{signTr}[U_\mu(x)]. \qquad (6.18)$$

In general, in an $SU(N)$ gauge theory, we take $Z_\mu(x) = z_n$, where $z_n \in Z_n$ is the element for which $\text{ReTr}[z_n^* U_\mu(x)]$ is largest.

   The mapping of the gauge-fixed $SU(N)$ lattice configuration to a Z_N lattice configuration

$$U_\mu(x) \to Z_\mu(x) \qquad (6.19)$$

is known as *"center projection."* Plaquettes with $Z(p) \neq 1$ on the projected lattice, where $Z(p)$ is the product of $Z_\mu(x)$ link variables around plaquette p, are known as

"P-plaquettes," and together they identify the position of thin center vortices known as *"P-vortices."* The claim is that this procedure locates center vortices on the unprojected lattice; P-vortices lie somewhere in the middle of the thick vortices of the original lattice configuration.

P-vortices, strictly speaking, are not located on the original lattice, but rather on the corresponding "dual" lattice, whose sites are shifted away from the sites of the original lattice by half a lattice spacing in the $\mu = 1, \ldots, D$ directions. In $D = 2$ dimensions, a plaquette is said to "dual to" (intersected in the middle by) a site of the dual lattice, located at the center of the plaquette. In three dimensions, a plaquette is dual to a link on the dual lattice, orthogonal to the plaquette, which runs through the center of the plaquette. In four dimensions a plaquette is dual to a plaquette on the dual lattice, which is oriented in a plane orthogonal to original plaquette; the areas of the two plaquettes intersect at a common midpoint. Suppose, for example, that at a fixed time we have a set of P-plaquettes oriented orthogonal to the x-axis, as shown in Fig. 6.1. A P-vortex in this timeslice is a line which pierces the center of each P-plaquette, and carries magnetic flux in the center of the gauge group. As the vortex line propagates in time, it traces out an area on the dual lattice, formed by plaquettes which are dual to the P-plaquettes.

The center-projected lattice is a configuration of Z_N lattice gauge theory, which of course is an abelian gauge theory, and has a simple Stokes Law

$$Z(C) = \prod_{p \in S(C)} Z(p), \tag{6.20}$$

where $S(C)$ is any surface bounded by C, and the product is over the plaquettes p which make up that surface. Let C be a planar loop, and $S(C)$ be the minimal surface. If there are no P-plaquettes in the minimal surface, then $Z(C) = 1$. If there
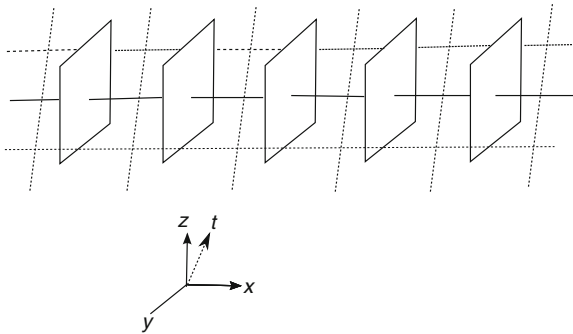

Fig. 6.1 A set of P-plaquettes, oriented parallel to the *y–z* plane. A thin vortex in $D = 3$ dimensions is a line running through the middle of the P-plaquettes, carrying a unit of center flux. The vortex line is dual to the P-plaquettes and runs along links of the dual lattice. In $D = 4$ dimensions the thin vortex is a surface on the dual lattice, indicated the dashed lines, formed by plaquettes which are dual to the P-plaquettes. Note that the sites of the dual lattice are displaced by half a lattice spacing in time, as well as in the space directions

is one or more, then the product can be restricted to the P-plaquettes ($Z(p) \neq 1$) only,

$$Z(C) = \prod_{\substack{p \in S(C) \\ Z(p) \neq 1}} Z(p). \tag{6.21}$$

Since each P-plaquette is dual to a plaquette (four dimensions) or a link (three dimensions) on a P-vortex which intersects the P-plaquette at a point in the middle, we say that the minimal surface is "pierced" by a P-vortex at the location of the P-plaquette.

Let us refer to the maximum distance between any two points on a vortex as the *extension* of the vortex, and suppose that there were an upper limit L_{max} to vortex extension. For very large Wilson loops, this means that the only vortices which can link to the loop are those which are in the neighborhood (i.e. within a distance L_{max}) of the loop perimeter, and the minimal area is irrelevant. It is not too hard to see that a perimeter, rather than an area-law falloff of center-projected Wilson loops $\langle Z(C) \rangle$ would be the result (cf. [8]). This means that if the center-projected loops do, in fact, have an area-law falloff, then the only possible upper bound on vortex extension is the finite lattice volume. The picture is that P-vortices are random surfaces which "percolate" throughout the spacetime volume. The piercings, by such vortices, of widely separated plaquettes on any given planar surface are uncorrelated, leading to a lack of correlation among large center-projected loops.

6.2.2 Finding Thin Vortices

If the gauge-fixing and center projection procedure really locates center vortices on the unprojected lattice, whose position is initially unknown, then it should also locate vortices whose position is known beforehand. In other words, suppose we have some lattice configuration U, and insert by hand a center vortex, via a singular gauge transformation, somewhere in the lattice. Assuming it is possible to gauge transform to a unique global maximum of R, will this center vortex be among the set of vortices identified by the center projection procedure? The answer is yes, and the argument is as follows: Let U denote some thermalized lattice configuration. A thin center vortex is created, on the background U, by a singular gauge transformation. Denote the resulting configuration as U'. Since the discontinuity of the singular gauge transformation is a center element, this discontinuity is invisible in the adjoint representation. It follows that the two lattice configurations in the adjoint representation, U_A and U'_A, are gauge equivalent. Then maximal center gauge will take U_A and U'_A into the *same* gauge-fixed configuration $\widetilde{U}_A$. In the passage to maximal center gauge, the link variables U and U' in the

fundamental representation are transformed to configurations $\widetilde{U}$ and $\widetilde{U}'$ which both equal $\widetilde{U}_A$ in the adjoint representation. Since they correspond to the same configuration in the adjoint representation, the $\widetilde{U}$ and $\widetilde{U}'$ lattice configurations can differ only by continuous and/or singular Z_N transformations. This means that the singular $SU(N)$ gauge transformation which inserted the thin vortex in U' must appear as a singular Z_N transformation relating $\widetilde{U}'$ to $\widetilde{U}$.

What has happened here is that the original singular $SU(N)$ gauge transformation, which may be quite smooth (up to the discontinuity) and extended, has been squeezed by the gauge-fixing condition to the identity everywhere except on a Dirac volume (bounded by the thin vortex sheet), where it has the effect of simply multiplying a certain set of links by -1. Upon center projection, $\widetilde{U}_\mu(x) \to Z_\mu(x)$ and $\widetilde{U}' \to Z'_\mu(x)$, and the projected configurations differ by the same discontinuous Z_2 gauge transformation. This discontinuity then shows up as an additional P-vortex in Z', not present in Z, at the location of the vortex inserted by hand.

While encouraging, this argument has two weaknesses: First, the inserted vortex is thin. The vortices which we are looking for in typical lattice configurations are thick. Secondly, the argument that U_A and U'_A transform into a unique configuration $\widetilde{U}_A$ in the adjoint representation ignores the fact that there exist a vast number of local maxima of the gauge-fixing functional R in (6.16), known as "Gribov copies." Existing numerical techniques are only able to fix the lattice to one of these local maxima, not to the global maximum. Despite this fact, the method has been shown, in numerical tests, to successfully locate thin vortices that were inserted by hand into the lattice [9].

The ability to find thin vortices inserted in the lattice is not unique to maximal center gauge, according to the above argument, but is shared by any choice of gauge which is defined as a condition on link variables in the adjoint representation. It is therefore worthwhile to explore some other possible gauge choices.

6.2.3 Indirect Maximal Center Gauge

This gauge [10] is useful in exploring connections between abelian monopoles and vortices.

The starting point is *maximal abelian gauge* [11], which minimizes the off-diagonal elements of the link variables, leaving a residual $U(1)^{N-1}$ gauge invariance (we will have more to say about this gauge, and its motivation, in Sect. 8.4) For $SU(2)$, maximize

$$R = \sum_{x,\mu} \text{Tr}[U_\mu(x)\sigma_3 U_\mu^\dagger(x)\sigma_3]. \tag{6.22}$$

This is equivalent to maximizing the (33) matrix element of the $SU(2)$ link variables in the adjoint representation. The link variables are decomposed as

$$U_\mu(x) = C_\mu(x)D_\mu(x), \tag{6.23}$$

where D is the diagonal part of the link variable, rescaled to restore unitarity

$$
\begin{aligned}
D_\mu &= \frac{1}{\sqrt{\left|[U_\mu]_{11}\right|^2 + \left|[U_\mu]_{22}\right|^2}} \begin{bmatrix} [U_\mu]_{11} & 0 \\ 0 & [U_\mu]_{22} \end{bmatrix} \\
&= \begin{bmatrix} e^{i\theta_\mu} & 0 \\ 0 & e^{-i\theta_\mu} \end{bmatrix}.
\end{aligned} \tag{6.24}
$$

Then we use the residual $U(1)$ symmetry to maximize

$$\widetilde{R} = \sum_{x,\mu} |\text{Tr}[D_\mu(x)]|^2 = \sum_{x,\mu} 4\cos^2\left(\theta_\mu(x)\right), \tag{6.25}$$

leaving a residual center (in this case Z_2) gauge symmetry. This brings the configuration into the indirect maximal center gauge.

Both the direct and indirect maximal center gauges have Gribov copies: for any lattice configuration there are a huge number of local maxima of R (much like spin glasses [12]), and there is no known technique for finding the global maximum. There are two strategies for dealing with this problem. One can just make an effort to find the "best" copy one can, e.g. by a simulated annealing technique, or else give up on trying to find a best copy, and average over all Gribov copies. In Monte Carlo simulations this last strategy is easy; it is equivalent to simply picking a gauge copy at random.

6.2.4 Laplacian Center Gauge

Laplacian gauges, originally introduced by Vink and Wiese [13, 14], are gauges which avoid the Gribov problem. Consider a Yang-Mills theory with two scalar fields ϕ_1^c, ϕ_2^c in the adjoint representation, where c is the color index. The unitary gauge

$$
\begin{aligned}
\phi_1^c(x) &= \rho(x)\delta^{c3} \\
\phi_2^2(x) &= 0, \quad \phi_2^1(x) > 0
\end{aligned} \tag{6.26}
$$

can be fixed uniquely, leaving a residual Z_2 symmetry. In the Laplacian center gauge [15, 16], the two scalar fields are taken to be the two lowest eigenmodes

$$\sum_y -\Delta^{ab}(x,y)f_n^b(y) = \lambda_n f_n^a(x) \tag{6.27}$$

of the (negative) covariant Laplacian $-\Delta$ in adjoint representation

$$-\Delta^{ab}(x,y) = -\sum_{\mu}\Big([U_{A\mu}(x)]^{ab}\delta_{y,x+\hat{\mu}} + [U_{A\mu}(x-\hat{\mu})]_{ba}\delta_{y,x-\hat{\mu}} - 2\delta_{xy}\delta^{ab}\Big). \quad (6.28)$$

The Laplacian center gauge is the unitary gauge defined by (6.26), with the two lowest eigenmodes serving as the two scalar fields $\phi_{1,2}$.

6.2.5 Direct Laplacian Center Gauge

This is a hybrid [17], which for $SU(2)$ uses the two lowest eigenmodes of the adjoint covariant Laplacian to select a particular Gribov copy of direct maximal center gauge. In practice, the results obtained in this gauge are indistinguishable from picking gauge copies at random in direct maximal center gauge.

In the end, the only real justifications for using one or another gauge are empirical. Most of the results I will show were obtained in the direct Laplacian center gauge, but the corresponding results obtained in direct and indirect maximal center gauges are very similar.

6.3 The Numerical Evidence

The numerical results which bear on the center vortex mechanism fall into several categories:

1. Center Dominance: What string tension is obtained from P-vortices?
2. Vortex-Limited Wilson loops: What is the correlation between center-projected loops and Wilson loops on the unprojected lattice? Do P-vortices really locate thick vortices on the original lattice?
3. Vortex Removal: What is the effect of removing vortices, identified by center projection, from the lattice configuration?
4. Scaling: Does the density of vortices scale according to the asymptotic freedom prediction?
5. Finite Temperature
6. Chiral Condensate/Topological Charge.

6.3.1 Center Dominance

The very first question is whether, under the factorization in one of the center gauges

$$U_{\mu}(x) = Z_{\mu}(x)V_{\mu}(x), \quad (6.29)$$

the variables $Z_\mu(x)$ of the center-projected configuration carry the disorder responsible for confinement.

The expectation value of a rectangular $R \times T$ Wilson loop typically has the following dependence on R and T:

$$W[R, T] = \exp[-(\sigma RT + \mu(R + T) - a(T/R + R/T) + b]. \tag{6.30}$$

The term $\mu(R + T)$ is a self-energy term, and is divergent in the continuum limit. Creutz noticed that one could form a ratio of rectangular loops[2]

$$\chi[R, T] \equiv -\log\left\{\frac{W[R, T]W[R - 1, T - 1]}{W[R - 1, T]W[R, T - 1]}\right\}, \tag{6.31}$$

such that the self-energy terms cancel out, and in the limit of large loop areas the Creutz ratio converges to the asymptotic string tension

$$\chi[R, T] \to \sigma. \tag{6.32}$$

Let $W_{cp}[R, T]$ be the Wilson loops of the center-projected lattice, i.e. loops constructed from the link variables $Z_\mu(x)$ identified in maximal center gauge, and let $\chi_{cp}[R, T]$ be the corresponding Creutz ratios. Then the question is whether, at large R, T, these Creutz ratios converge to the asymptotic string tension of the unprojected lattice.

Figure 6.2 is a first look at $\chi_{cp}[R, R]$ versus coupling β for $R = 1$–6. There are three points, in particular, worth noting:

- At each β, the $\chi_{cp}[R, R]$ are almost identical for all $R > 1$. In other words, the asymptotic string tension is obtained already at $R = 2$.
- There is excellent agreement with asymptotic freedom, indicated by the straight line passing through the data points at $R = 2$. The slope of this line is a prediction of the Yang-Mills two-loop beta function.
- Even the data for $\chi_{cp}[1, 1]$ seems to fit with asymptotic freedom.

Creutz ratios in the full theory, by contrast, certainly do depend on R, and only converge to the asymptotic string tension, at a given β, at large R.

Figure 6.3 displays the (nearly constant) values for the Creutz ratios $\chi_{cp}[R, R]$ vs. R at several β values. At each β the central straight line represents the accepted value of the asymptotic string tension of lattice Yang-Mills theory; the upper and lower straight lines indicate the associated statistical errors. The agreement of the string tension on the center-projected lattice with the full string tension is known as *"center dominance."* The fact that $\chi_{cp}[R, R]$ is nearly R-independent means that the center-projected potential is linear starting from $R = 2$. This feature is known as *"precocious linearity."*

---

[2] The Creutz ratios $\chi[R, T]$ should not be confused with the group characters $\chi_r[g]$ discussed previously.

Fig. 6.2 Creutz ratios of center-projected Wilson loops. From Faber et al. [17]
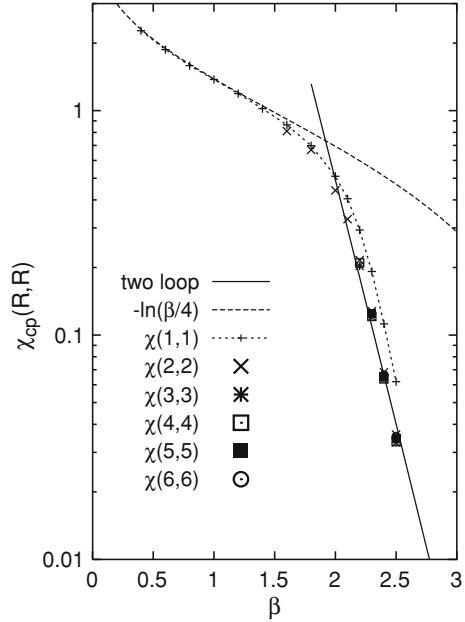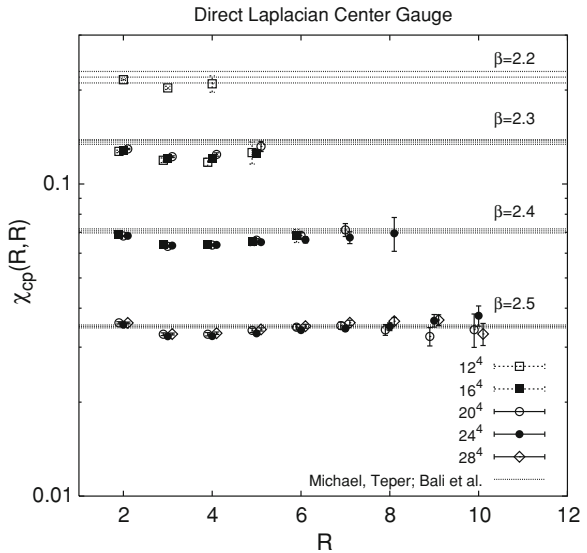

Fig. 6.3 Combined data, at $\beta = 2.2$–2.5, for center-projected Creutz ratios obtained after direct Laplacian center gauge fixing. Horizontal bands indicate the accepted values of asymptotic string tensions on the unprojected lattice, with the corresponding errorbars [18, 19]. From Faber et al. [17]

In order to understand precocious linearity, recall that center vortices on the unprojected lattice are thick objects, and the full effect on a Wilson loop, i.e. multiplication by a center element, is only obtained for thick vortices linked to large loops. Center projection essentially shrinks the thickness of the vortex to one lattice spacing, so the full effect of linking to a vortex appears for even the smallest

center-projected Wilson loops. Thus, if P-vortex plaquettes are completely uncorrelated in a plane, then we must see a linear potential from the smallest distances. If this is not seen, then either the vortex surface is very rough, bending in and out of the plane, or else there are very small vortices. In either case there are correlations between nearby P-plaquettes, and a delay in the onset of the linear projected potential.

6.3.2 Vortex-Limited Wilson Loops

The fact that P-vortices in the projected lattice reproduce the correct, or nearly correct, asymptotic string tension does not, by itself, prove that these objects locate something physical on the unprojected lattice. P-vortices, after all, are identified with the help of gauge fixing; perhaps they are artifacts of the gauge choice. To address that issue, the crucial question to ask is whether P-vortices are correlated with gauge-invariant observables, such as ordinary Wilson loops.

A vortex-limited Wilson loop $W_n(C)$ is the expectation value of a Wilson loop on the unprojected lattice, evaluated in the subensemble of configurations in which the minimal area of the loop is pierced by precisely n P-vortices (i.e. there are n P-plaquettes in the minimal area). Here the center projection is used only to select the data set; the loop itself is evaluated using unprojected link variables.

If P-vortices on the projected lattice locate center vortices on the unprojected lattice, then for $SU(2)$ we would expect, asymptotically, that

$$\frac{W_n(C)}{W_0(C)} \rightarrow (-1)^n. \tag{6.33}$$

The reason is as follows: In maximal center gauge we have

$$\begin{aligned}
W_n(C) &= \langle Z(C)\text{Tr}[V(C)]\rangle_{n_p(C)=n} \\
&= (-1)^n \langle \text{Tr}[V(C)]\rangle_{n_p(C)=n}, \tag{6.34}
\end{aligned}$$

where the $\langle\ldots\rangle_{n_p(C)}$ notation indicates that the planar loop is only evaluated in those configurations for which $n_p(C)$ P-plaquettes lie in the interior of the minimal area of the loop. If we assume that $V_\mu(x)$ has only short range correlations, then on a large loop this variable is insensitive to the presence or absence of vortices deep in the interior of the loop, i.e.

$$\langle \text{Tr}[V(C)]\rangle_{n_p(C)=n} \approx \langle \text{Tr}[V(C)]\rangle_{n_p(C)=0} \tag{6.35}$$

for large loops. From this, (6.33) follows immediately.

Fig. 6.4 W_n/W_0 ratios at $\beta = 2.3$ vs. loop area. Note that $W_1/W_0 \approx -1$, and $W_2/W_0 \approx 1$ for the largest loops. From Faber et al. [17]
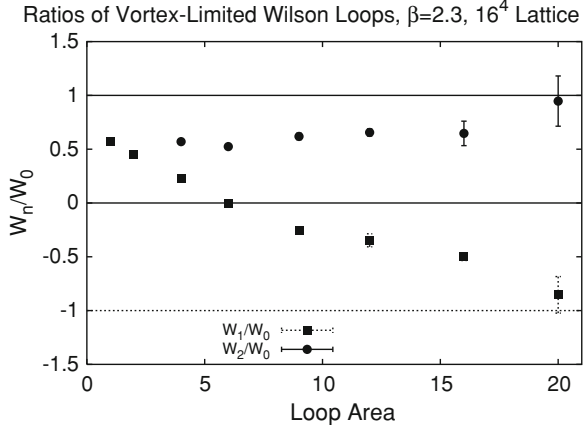
Figure 6.4 shows results for W_1/W_0, and W_2/W_0 for square $R \times R$ and rectangular $R \times (R + 1)$ loops. The data is quite consistent with (6.33). One can also look at loops with even or odd numbers of P-vortices piercing the loop. We find, for $SU(2)$ [17]

$$\frac{W_{odd}(C)}{W_{even}(C)} \rightarrow -1. \tag{6.36}$$

From the fact that $W_n/W_0 \rightarrow (-1)^n$, we conclude that P-vortices are correlated with the sign of the Wilson loop, in just the way expected if these P-vortices are correlated with the location of thick center vortices. If we restrict ourselves to loops with $n = 0$, or $n =$ even, then there are no fluctuations in the sign of the loop due to vortices, and in the vortex scenario the asymptotic string tension should vanish. In fact, Creutz ratios constructed from loops of this kind show exactly the expected behavior, i.e.

$$\chi_0(R, R) \rightarrow 0, \quad \chi_{even}(R, R) \rightarrow 0. \tag{6.37}$$

6.3.3 Vortex Removal

A powerful consistency test was devised by de Forcrand and D'Elia in [20]. Suppose we remove center vortices from the unprojected configuration, by multiplying a lattice configuration, in a center gauge, by its own center projection, i.e.

$$U_\mu(x) \rightarrow U'_\mu(x) = Z^*_\mu(x)U_\mu(x) = V_\mu(x). \tag{6.38}$$

The center-projection of $U'_\mu(x)$ has no center vortices at all. What the procedure actually does is to insert a thin vortex in the middle of a thick vortex.

Fig. 6.5 Creutz ratios at $\beta = 2.3$ in the unprojected configuration (*open circle*), in the center projected configurations derived from direct Laplacian gauge (*filled circle*), and in the unprojected configurations with vortices removed (*open square*). The horizontal band indicates the accepted asymptotic string tension of the unprojected theory. From Faber et al. [17]
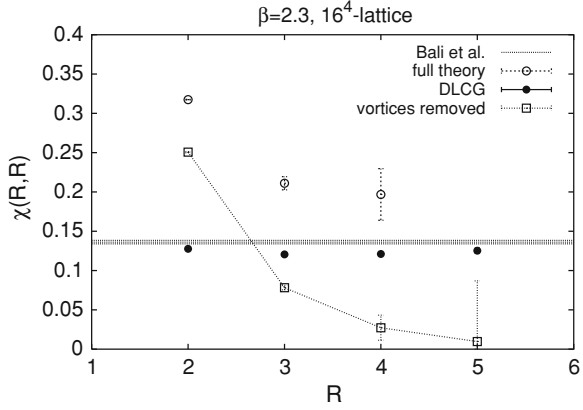
The asymptotic fields of the thin and thick vortices would cancel out, removing the vortex disordering effect on large loops. Thus, if (i) P-vortices locate thick vortices (the evidence for that comes from the vortex-limited Wilson loops); and (ii) vortex disorder is confining disorder (the evidence is center dominance); then removing vortices in this way should also remove the asymptotic string tension. This is, in fact, exactly what happens, as can be seen in Fig. 6.5.

6.3.4 Scaling of the P-vortex Density

If center vortices are physical objects, it makes sense that their density (vortex area per unit volume) in the vacuum is lattice-spacing independent in the continuum limit. If P-vortices lie in the middle of center vortices, it would likewise follow that P-vortex density is lattice-spacing independent, a property which was first checked in [21]. Let

$$N_{vor} = \text{total no. of P-plaquettes} = \text{total P-vortex area in lattice units}$$
$$N_T = \text{total no. of plaquettes} = \text{total lattice volume} \times 6.$$
(6.39)

Then the density p of P-plaquettes is related to the vortex density ρ in physical units via

$$p = \frac{N_{vor}}{N_T} = \frac{N_{vor}a^2}{N_T a^4}a^2$$
$$= \frac{\text{Total Vortex Area}}{6 \times \text{Total Lattice Volume}}a^2$$
$$= \frac{1}{6}\rho a^2.$$
(6.40)

If ρ is a physical quantity (i.e. β-independent), then we can substitute the asymptotic freedom expression for lattice spacing $a(\beta, \Lambda)$ to obtain the prediction

$$p = \frac{\rho}{6\Lambda^2}\left(\frac{6\pi^2}{11}\beta\right)^{102/121} \exp\left[-\frac{6\pi^2}{11}\beta\right]. \tag{6.41}$$

The average value of p is obtained from $W_{cp}[1, 1]$, because

$$\frac{1}{2}W_{cp}[1, 1] = (1 - p) + p \times (-1) = 1 - 2p \tag{6.42}$$

Figure 6.6 displays p vs. β; the straight line has the slope predicted from asymptotic freedom. In Fig. 6.7 we show closely related results by Gubarev et al. [22]

Fig. 6.6 Evidence of asymptotic scaling of the P-vortex surface density. The data points are the number of P-plaquettes per unit volume, and the solid line is the asymptotic freedom prediction with $\sqrt{\rho/6\Lambda^2} = 50$. From Faber et al. [17]
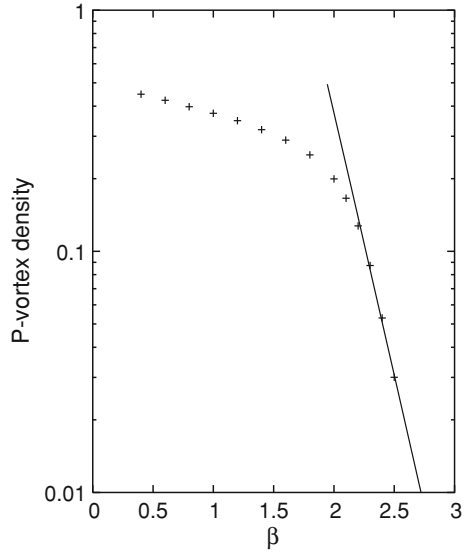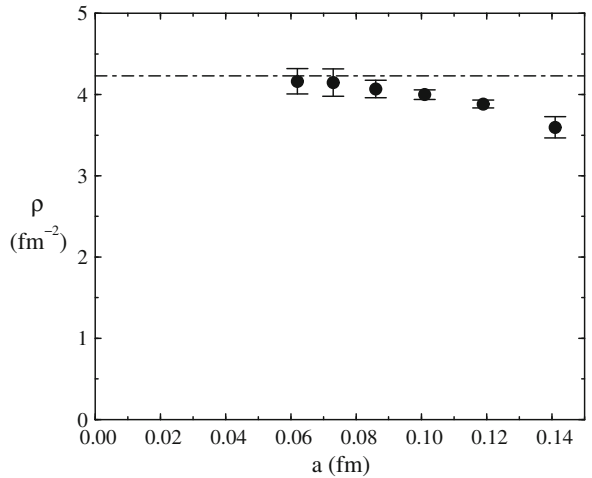

Fig. 6.7 P-vortex density vs. lattice spacing. From Gubarev et al. [22]

using the indirect maximal center gauge. This is a plot of ρ in physical units vs. lattice spacing, and the insensitivity to lattice spacing at small lattice spacings is quite evident.

Gubarev et al. [22] also investigated the (gauge-invariant) action density, above the vacuum average, at the location of P-plaquettes, i.e.

$$\Delta S = -\frac{1}{2}\beta(W_1[1,1] - W[1,1]). \tag{6.43}$$

It was found that this action density is singular in the continuum limit. The result means that in terms of action density, the middle of a center vortex stands out from the background, in the continuum limit, as a surface (or "brane") of infinite action density. The fact that the density of vortices is constant when taking the continuum limit, despite the infinite action density limit at the location of P-vortices, suggests a delicate cancellation between surface action and surface entropy.[3]

6.3.5 Vortices at High Temperatures

We have already encountered Polyakov loops, which measure the quark free energy in a periodic lattice at finite temperature $T = 1/(L_t a)$, where L_t is the extension of the lattice in the time direction. As the temperature is raised to $T_c = 220$ MeV, quantized $SU(2)$ gauge fields theory go through a "deconfinement" transition, where hadrons dissolve into their constituents. One can show numerically that at $T > T_c$ the quark free energy, as measured by the Polyakov loop, becomes finite. On the other hand, the vacuum of the deconfined phase still retains some features associated with the confined phase, since it has also been shown that spacelike Wilson loops (which are a measure of vacuum fluctuations) retain an area law and asymptotic string tension beyond the phase transition, even though the static quark potential measured by Polyakov loop correlators goes flat. A theory of confinement must be consistent with both of these features, i.e. an area law for spacelike loops but non-zero Polyakov loop expectation values; at $T > T_c$.

Center vortices fit in nicely with these requirements, and provide an intuitive picture of the transition which can be checked via numerical simulation. It is easiest to visualize the situation in 2 + 1 dimensions. Let us consider the expectation value of the product of two Polyakov loops, $P(\mathbf{x})P^*(\mathbf{x} + \mathbf{R})$. By the same argument as for a single Polyakov line, this quantity is related to the free energy of a quark–antiquark pair

$$e^{-\beta F_{q\bar{q}}} = \langle P(\mathbf{x})P^*(\mathbf{x} + \mathbf{R})\rangle, \tag{6.44}$$

---

[3] Zakharov [23] has argued that surfaces with singular action density fit in well with the need for certain power corrections in QCD sum rules.

where $\beta = 1/T$. In the confinement phase a flux tube, with some temperature-dependent string tension, will form between the quark–antiquark pair, and at large R the free energy of a separated quark-antiquark pair will approximate $\sigma(T)R + V_0$. This means that, in lattice units,

$$\langle P(\mathbf{x})P^*(\mathbf{x} + \mathbf{R})\rangle \propto e^{-\sigma(T)L_t R}. \tag{6.45}$$

In the vortex theory, this exponential falloff, proportional to the minimal area bounded by the Polyakov loops, is due to random (uncorrelated) piercings of the minimal area by vortices. However, vortices have a certain thickness, and it is known from the work of Kovacs and Tomboulis [24], and de Forcrand and von Smekal [25–27], that when vortices are "'squeezed" by the finite extension of the lattice, i.e. the thickness of the lattice is less than the normal thickness of center vortices, then the free energy of the squeezed vortices rises rapidly away from zero as the lattice extension is reduced, and at some point squeezed vortices cease to percolate. The situation in 2 + 1 dimensions is sketched in Fig. 6.8. When the squeezed vortices no longer percolate, then piercings of the minimal surface between Polyakov loops are correlated, and the piercings can no longer result in the exponential falloff shown in (6.45). An exponential falloff of the correlator guarantees that $\langle P(x)\rangle = 0$, but in the absence of this falloff there is no reason for the expectation value of the Polyakov loop to vanish, and we have

$$\begin{aligned} e^{-\beta F_{q\bar{q}}} &= \langle P(\mathbf{x})P^*(\mathbf{x} + \mathbf{R})\rangle \\ &\rightarrow \langle P(\mathbf{x})\rangle\langle P^*(\mathbf{x} + \mathbf{R})\rangle \\ &> 0. \end{aligned} \tag{6.46}$$

Since the "cost" in free energy of having an infinitely separated quark and anti-quark is finite, this is the deconfined phase.

On the other hand, let us consider spacelike loops at high temperature. The area law for these loops is due to piercing by vortices oriented in the time direction, as indicated in Fig. 6.9. However, the thickness of such vortices is not constrained by

Fig. 6.8 Vortices running in the spacelike directions disorder Polyakov loops. When the time extension L_t is smaller than the diameter of the vortex (high temperature case), then spacelike vortices are "squeezed" and cease to percolate
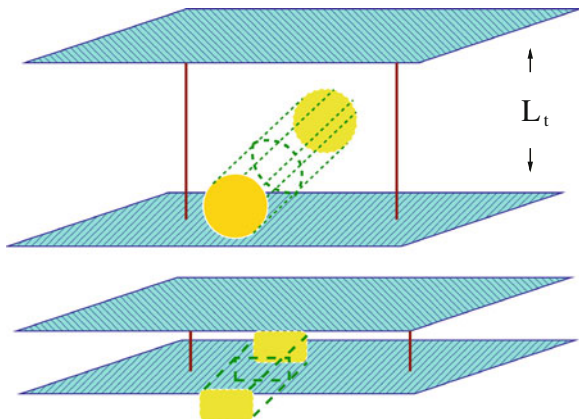
Fig. 6.9 Vortices running in
the timelike direction dis-
order spacelike Wilson loops.
The vortex cross section is
not constrained by a small
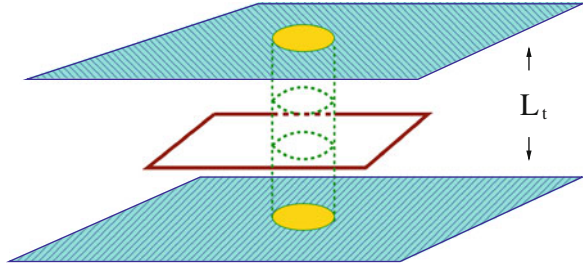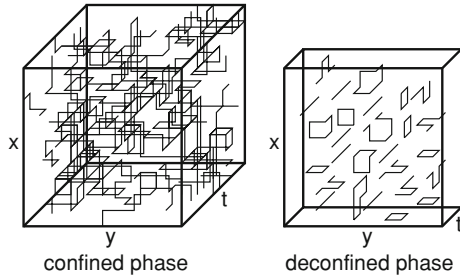extension in the time
direction

Fig. 6.10 Schematic picture
of P-vortices in a "space-
slice" of the lattice, below
and above the deconfinement
phase transition. From
Engelhardt et al. [8]

confined phase   deconfined phase

the finite lattice extension in the time direction, and their free energy on a lattice
with large extension in all space directions is essentially zero. It is necessary,
however, that these unsuppressed configurations are oriented in the time direction
all along their length, which means such vortices will extend through the lattice in
the time direction, and close via the lattice periodicity. Vortices of this type account
for the area law falloff of spatial Wilson loops, even at very high temperatures.

   In $3 + 1$ dimensions the vortex lines become vortex sheets winding through the
periodic lattice in the time direction, but to visualize the situation it is helpful to
consider a space-slice of the lattice, say at some fixed value of the z-coordinate. In
a space-slice the vortex surfaces are again line-like. The P-vortices which we use
to locate center vortices have, of course, no thickness, and are literally lines. The
distribution of P-vortices in a space-slice is shown schematically in Fig. 6.10. At
low temperatures, the P-vortices percolate throughout the lattice, while at high
temperatures the vortex loops are small, but some of them (the ones responsible for
the spacelike string tension) wind through the lattice in the periodic time direction.

   On a time-slice, in $D = 2 + 1$ dimensions, the P-vortices would simply be a lot
of uncorrelated points (P-plaquettes) in the plane. In $D = 3 + 1$ dimensions what
we should see is the cross-section of P-vortex surfaces which are closed by lattice
periodicity in the time direction, but percolate in spatial directions. Thus, in a time
slice, we would expect to see large loops percolating throughout the spatial vol-
ume, even at high temperatures.

   This picture and its numerical verification are due to the Tübingen group in [8]
(see also Refs. [28, 29]). Figure 6.11 shows some data for a space-slice at finite
temperature. The x-axis is in units of the maximal extension of the vortex loop in
the $L^2 \times L_t$ space-slice 3-volume, i.e. $\sqrt{2(L/2)^2 + (L_t/2)^2}$. The y-axis is the
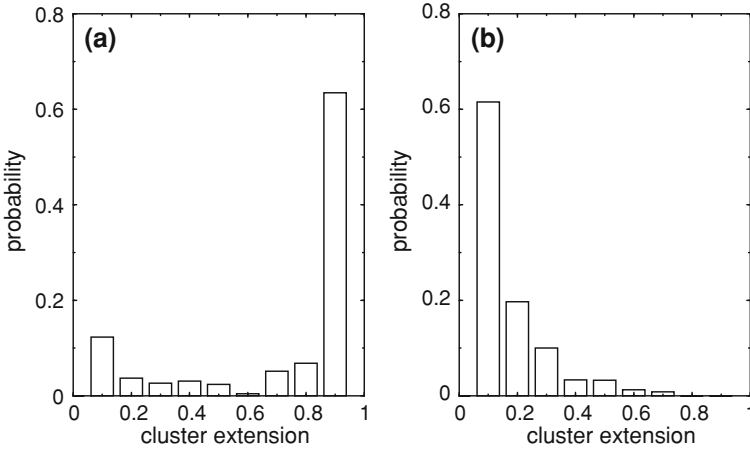
Fig. 6.11 Histograms of vortex extension in a space-slice at finite temperature, both below (left figure, $T = 0.7\,T_c$) and above (right figure, $T = 1.85\,T_c$) the deconfinement phase transition. The data is for $\beta = 2.4$ on a $12^3 \times 7$ (confined) and $12^3 \times 3$ (deconfined) lattices; center projection in direct maximal center gauge. From Engelhardt et al. [8]

Fig. 6.12 Histograms of vortex extension in a time slice at finite temperature. Vortices are identified by center projection at $\beta = 2.4$ in direct maximal center gauge. Data at three temperatures are shown on the same figure. From Engelhardt et al. [8]
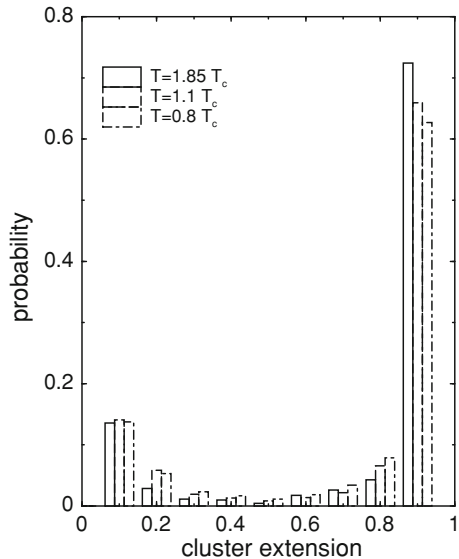
fraction of P-vortex plaquettes in the space-slice which are dual to a vortex loop of a given extension. At low temperature, most P-plaquettes belong to vortex loops whose extensions are comparable to the size of the lattice. At high temperatures, most P-plaquettes belong to short loops in the space-slice. The same sort of data for a time slice is shown in Fig. 6.12, at temperatures just below, just above, and

well above the deconfinement transition. On a time slice there is little difference with temperature. At any temperature, most P-plaquettes belong to P-vortices whose extension is comparable to the size of the lattice. Thus we have percolation of vortices in a time-slice, and a corresponding area-law falloff for spacelike Wilson loops.

6.3.6 Chiral Condensates and Topological Charge

Consider a gauge theory with two light quarks, u and d, which together form a flavor doublet. As long as the masses of these quarks are zero, the Lagrangian is invariant under the transformations

$$\begin{bmatrix} u \\ d \end{bmatrix} \rightarrow U_B \times U_A \times U_L \times U_R \begin{bmatrix} u \\ d \end{bmatrix}, \tag{6.47}$$

where

$$\begin{aligned}
U_B &= \exp[i\theta_B] \\
U_A &= \exp[i\theta_A \gamma_5] \\
U_L &= \exp\left[i\theta_L^i (1 - \gamma_5)\frac{\sigma_i}{2}\right] \\
U_R &= \exp\left[i\theta_R^i (1 + \gamma_5)\frac{\sigma_i}{2}\right].
\end{aligned} \tag{6.48}$$

The symmetry with respect to $U(1)$ transformations U_B results, via Noether's Theorem, in conservation of baryon number. The axial $U(1)$ symmetry with respect to transformations U_A, while a symmetry of the Lagrangian, is not a symmetry of the full quantum theory. This symmetry is said to be *anomalous*. One way of understanding this anomaly in the continuum, in the context of path-integral quantization, is that the path-integration measure for the fermions is not invariant under U_A [30].[4] With most lattice regularizations the lattice measure is invariant under the axial $U(1)$ symmetry, and the symmetry is instead broken by the regularized lattice action [32]. Finally, while the chiral $SU(2)_L \times SU(2)_R$ symmetry transformations $U_L \times U_R$ are not broken by any anomaly, neither is this symmetry realized in the spectrum of the theory. If it were, then particles would appear in equal mass doublets of opposite parity, and this is something which is not found in nature. Instead, chiral symmetry is spontaneously broken to ordinary $SU(2)$ isospin symmetry, which consists of the subgroup of transformations for

---

[4] Another picture, in the Hamiltonian formulation, involves the notion of the Dirac sea. Under the influence of an external gauge field, the energy levels of the Dirac operator shift up or down, depending on the chirality of the state. This level-shifting can *appear* as the creation of quark-antiquark pairs by a gauge field, with quantum numbers violating conservation of the axial $U(1)$ symmetry [31].

which $\theta_L^i = \theta_R^i$. When a continuous global symmetry is broken spontaneously, the Goldstone theorem assures us that there must be some associated massless particles; in this case these are the pions, which would be massless if the quarks were exactly massless.[5]

Whenever there is a possibility of spontaneously broken symmetry, it is useful to introduce a local operator which is not invariant under the symmetry, and whose expectation value would average to zero if the symmetry were unbroken. The expectation value of such an operator is the order parameter for the symmetry breaking. For the Ising model the operator was the average spin, whose expectation value is the magnetization, and this is non-zero if the Z_2 global symmetry of the Ising model is spontaneously broken. In the case of global center symmetry, the Polyakov loop is the natural order parameter. For chiral symmetry, an appropriate order parameter is $\langle \overline{\psi}\psi \rangle$, and the symmetry is broken when this quantity is non-zero. This order parameter, known at the *chiral condensate*, is also of phenomenological importance. If the quark masses (known as *current* quark masses) are not exactly zero, then the pions, now described as "pseudo-Goldstone" particles, also acquire a mass, and this mass can be expressed in terms of the pion decay constant f_π, the quark masses, and the order parameter for chiral symmetry breaking:

$$m_\pi^2 = -2\frac{m_u + m_d}{f_\pi^2}\langle \overline{\psi}\psi \rangle. \qquad (6.49)$$

A derivation can be found in standard texts, e.g. [33].

As we saw in the case of the Ising model, a non-zero value for the order parameter can only be obtained via a careful order of limits: First we introduce a term which breaks the symmetry explicitly (an external magnetic field, for the Ising model, or a quark mass term, in the present case), take the infinite volume limit, and only then remove the explicit breaking term. Proceeding in this way, Banks and Casher [34] found a beautiful relationship between the chiral condensate and the density of near-zero eigenvalues of the Dirac operator. The derivation is as follows: Introduce a small quark mass m, which is to be taken to zero at the end of the calculation, and let

$$i \not{D}\phi_n = \lambda_n \phi_n \qquad (6.50)$$

be the eigenvalue equation for the Euclidean-space Dirac operator, with $\not{D} = \gamma^\mu D_\mu$. Then the massive quark propagator can be expressed as

$$\langle \psi(x)\overline{\psi}(y) \rangle = \sum_n \left\langle \frac{\phi_n(x)\phi_n^\dagger(y)}{-i\lambda_n + m} \right\rangle. \qquad (6.51)$$

---

[5] For a theory with N_f massless flavors of quarks, the chiral symmetry is $SU(N_f)_R \times SU(N_f)_L$, which is spontaneously broken to an $SU(N_f)$ flavor symmetry, and the massless Goldstone particles belong to a multiplet transforming in the adjoint representation of the remaining flavor symmetry.

We set $x = y$ to obtain the chiral condensate, then integrate over x (noting the orthonormality of the ϕ_n), and divide by spacetime volume, which gives

$$\langle \overline{\psi}\psi \rangle = -\frac{1}{V}\left\langle \sum_n \frac{i\lambda_n + m}{\lambda_n^2 + m^2} \right\rangle$$

$$= -\frac{2m}{V}\left\langle \sum_{\lambda_n > 0} \frac{1}{\lambda_n^2 + m^2} \right\rangle, \tag{6.52}$$

where we have used the fact that, apart from zero modes, the $\{i\lambda_n\}$ come in complex–conjugate pairs, so the term with λ_n in the numerator averages to zero. Next take the infinite volume limit, where the Dirac operator has a continuous spectrum, and

$$\frac{1}{V}\left\langle \sum_{\lambda_n > 0} \cdots \right\rangle \to \int d\lambda \rho(\lambda)\langle \cdots \rangle, \tag{6.53}$$

where $\rho(\lambda)$ is the density of eigenvalues per unit volume. This gives

$$\langle \overline{\psi}\psi \rangle = -2m \int d\lambda \frac{\rho(\lambda)}{\lambda^2 + m^2}. \tag{6.54}$$

The infinite volume limit is followed by the zero mass limit, using $m/(\lambda^2 + m^2) \to \pi\delta(\lambda)$, from which we finally obtain the Banks-Casher formula

$$\langle \overline{\psi}\psi \rangle = -\pi\rho(0). \tag{6.55}$$

From the point of view of the Banks-Casher formula, the mechanism of chiral symmetry breaking is tied to whatever property of vacuum field configurations gives rise to $\rho(0) > 0$. Note that $\rho(0) = 0$ at the level of perturbation theory. For example, the eigenvalues of the ordinary Laplacian operator $-\nabla^2$ are $\lambda = p^2$. Then, from

$$\int d^D p = \int d\Omega \int_0^\infty dp\, p^{D-1} \propto \int_0^\infty d\lambda \lambda^{(D-2)/2}, \tag{6.56}$$

we read off that in this case $\rho(\lambda) \propto \lambda^{(D-2)/2}$ in D Euclidean spacetime dimensions, so that $\rho(0) = 0$ for any $D > 2$. A similar argument holds for the density of states of the free Dirac operator.

Chiral symmetry breaking clearly does not require a non-vanishing asymptotic string tension, since the asymptotic string tension in real QCD, with light quarks, vanishes due to the string breaking mechanism. Chiral symmetry breaking does not even require gauge fields. The breaking can occur in theories with other types of forces between fermions, as in the Nambu Jona-Lasinio model [35], where the effect is due to four-fermion interactions. Nor is chiral symmetry breaking a necessary feature of asymptotically free gauge field theories; some counterexamples in supersymmetric confinement-like theories are known [36]. But for present purposes it is sufficient to note that chiral symmetry breaking certainly

does occur in QCD, and $\langle \overline{\psi}\psi \rangle$ is perfectly well-defined even in a theory without dynamical fermions; it is simply the trace of the quark propagator evaluated in a pure gauge theory.[6] It is entirely possible that the mechanism for chiral symmetry breaking in non-abelian gauge theories is not so closely related to the confinement mechanism. But it would be more economical, and more satisfying, if the objects which we suspect are responsible for confinement, i.e. the center vortices, are also somehow essential for a non-zero value for the chiral condensate.

On the lattice the chiral condensate can be computed directly, and it is also possible to numerically compute the low-lying spectrum of the Dirac operator. However, the lattice regularization of this operator introduces some thorny technical issues, which are discussed in most modern texts on lattice gauge theory (cf. in particular, Smit [37] and DeGrand and DeTar [38]). Briefly, the problem is that a naive lattice discretization, simply replacing covariant derivatives in the continuum by the one-link lattice covariant derivatives, leads to the notorious doubling problem: instead of one light fermion, one actually ends up with 2^D light fermions on a hypercubic lattice. One method of handling this problem, by introducing a one link mass term with a carefully tuned coefficient, breaks the chiral symmetry explicitly, at least at finite lattice spacing. Another formulation, known as *staggered fermions*, which places different fermionic components on neighboring lattice sites, reduces the multiplicity to four (known as "tastes") in $D = 4$ dimensions, while preserving a subgroup of the chiral symmetry. A variation of staggered fermions, known as *asqtad* fermions, introduces some next-to-nearest neighbor operators in order to reduce lattice artifacts, and improve convergence to the continuum limit. It has been used extensively in numerical simulations. Finally, the *overlap* Dirac operator has no doubling problem whatever, and is endowed with an exact invariance which, while not exactly the chiral symmetry shown in (6.48), approximates the standard $SU(2)_L \times SU(2)_R$ symmetry for a sufficiently smooth lattice gauge field background. We will not be detained by these issues here, except to mention that the results described below were obtained using either the staggered or the asqtad discretization of the Dirac operator. The overlap formulation is not so suitable for our purposes, because center-projected configurations are as far from smooth as possible, and the symmetry of the overlap operator is therefore not so close to the standard $SU(2)_L \times SU(2)_R$ chiral symmetry (cf. the discussion in [39]).

Lattice simulations by de Forcrand and co-workers [16, 20] have found that when vortices are removed via the procedure of Eq. 6.38, then the chiral condensate vanishes in the vortex-removed configuration, while in center-projected configurations, the value of the condensate is greatly enhanced, compared to its value in unprojected configurations. Their results are shown in Fig. 6.13. The low-lying spectrum of the Dirac operator in center-projected and vortex-removed configurations has also been calculated, in Ref. [39], and the results for the first

---

[6] It should be noted, however, that $\langle \overline{\psi}\psi \rangle$ diverges like $\log(m)$ in theories with no dynamical fermions, as $m \to 0$.

twenty complex–conjugate pairs of smallest-magnitude Dirac eigenvalues are shown in Fig. 6.14 (numerical details can be found in the cited reference). What is striking is that when vortices are removed, a large gap opens up around the zero eigenvalue; the eigenvalue density is clearly zero at $\lambda = 0$. In the center-projected configurations, the density of eigenvalues appear to be larger than in the unmodified configurations. This is all consistent with the Banks-Casher relation, and the results shown in Fig. 6.13. The evidence is that center vortices are not only responsible for the confining force, but also, by themselves, can be responsible for the spontaneous breaking of chiral symmetry.

Fig. 6.13 Chiral symmetry breaking order parameter $\langle \bar{\psi}\psi \rangle$ vs. lattice bare quark mass m_q, for the unmodified (*plus*), center-projected (*open square*), and vortex removed (*times*) configurations in pure $SU(2)$ lattice gauge theory. The order parameter for vortex removed configurations extrapolates to zero, as $m_q \rightarrow 0$. From Alexandrou et al. [16]
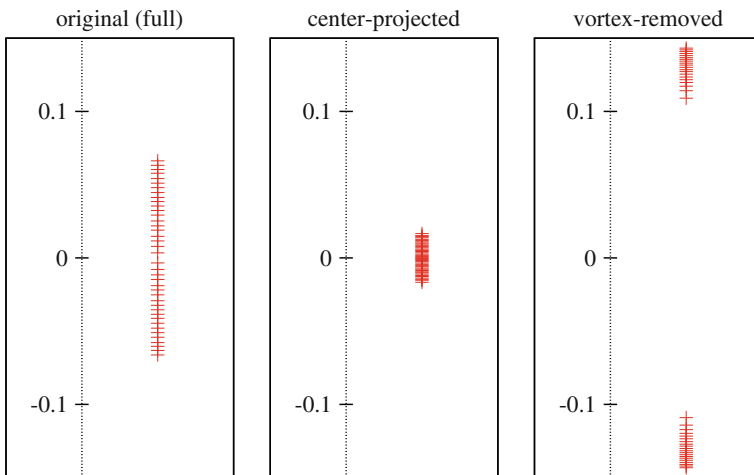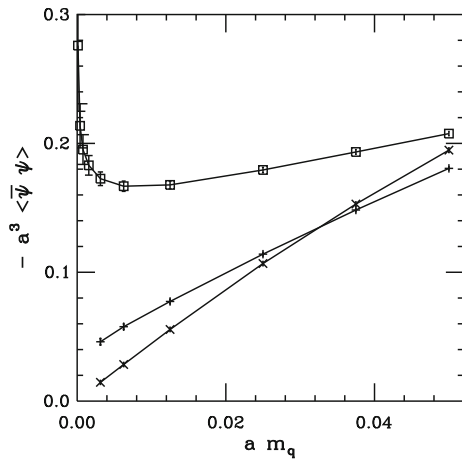

Fig. 6.14 Spectrum of the lowest-lying eigenvalues of the lattice Dirac operator (asqtad formulation), for unmodified, center projected, and vortex-removed configurations, in a pure $SU(2)$ gauge theory. See Hollwieser et al. [39] for details of the simulation

Next, let us return to the axial $U(1)$ symmetry, which is broken at the quantum level by an anomaly. In the absence of the anomaly, Noether's theorem would lead to a conserved current, $\partial^\mu j_{5\mu} = 0$, where

$$j_{5\mu}(x) = \overline{\psi}(x)\gamma_\mu\gamma_5\psi(x). \tag{6.57}$$

In fact the correct answer for massless quarks is [40]

$$\partial^\mu j_{5\mu} = -\frac{g^2 N_f}{16\pi^2}\varepsilon^{\mu\nu\alpha\beta}\mathrm{Tr}[F_{\mu\nu}F_{\alpha\beta}], \tag{6.58}$$

where N_f is the number of light flavors ($= 2$, if we consider only a doublet of u, d quarks). The integral of the right-hand side of this equation is proportional to a quantity known as the *topological charge*

$$Q = \frac{1}{32\pi^2}\int d^4x\,\varepsilon^{\mu\nu\alpha\beta}\mathrm{Tr}[F_{\mu\nu}F_{\alpha\beta}]. \tag{6.59}$$

The integrand $\varepsilon^{\mu\nu\alpha\beta}\mathrm{Tr}[F_{\mu\nu}F_{\alpha\beta}]$ can be expressed as a total divergence. There are, however, finite action configurations (such as instantons) for which the integral of this quantity is non-vanishing, and Q is an integer [41]. The restriction to integer values is of course no accident. There is a celebrated result in mathematics (known as the Atiyah-Singer Index Theorem) which tells us that the topological charge of any given gauge-field configuration is equal to the number of zero modes of the Dirac operator of positive chirality (i.e. they are also eigenstates of the γ_5 Dirac matrix with positive eigenvalue), minus the number of zero modes with negative chirality; this difference is obviously an integer. The mean-square variation of the topological charge, per unit spacetime volume, is known as the topological susceptibility

$$\chi = \frac{\langle Q^2\rangle}{V}, \tag{6.60}$$

and this quantity also plays an important role in phenomenology. If the axial $U(1)$ symmetry were conserved but spontaneously broken, then there would be an associated massless Goldstone boson (if the quarks were massless) or a light pseudo-Goldstone boson (if the quarks had a small current mass). This would be the η' boson, which at about 1 GeV is far to heavy to be considered a pseudo-Goldstone particle. An important result due to Witten [42] and Veneziano [43] relates the topological susceptibility χ, computed in a pure gauge theory, to the η' mass

$$m_{\eta'}^2 = \frac{2N_f}{f_\pi^2}\chi, \tag{6.61}$$

and this is known, naturally enough, as the Witten–Veneziano mass formula.

The topological susceptibility, like the chiral condensate, is a property of the vacuum state, and has a well defined expectation value in a pure gauge theory with no quark fields whatever. It is not obvious that topological susceptibility has any

Fig. 6.15 Intersection points (**a**) and writing points (**b**) which contribute to the topological charge of a P-vortex surface. From Reinhardt [45]

(a) **(b)**
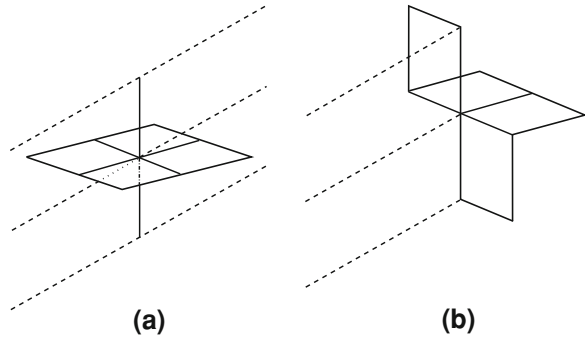
direct connection to the ultra long-range property of confinement, or to center vortices. It could be that other non-confining configurations are responsible for the density of topological charge, and instantons (which carry integer amounts of topological charge) have been the traditional candidates. An indication that center vortices might play a role was the result, due to de Forcrand and D'Elia [20], that topological charge plummets to zero when vortices are removed from lattice configurations. A more detailed analysis is due to Engelhardt [44], who has pointed out that on a thin vortex surface, topological charge density can arise at sites on the surface where the tangent vectors to the surface are in all four space-time directions. These sites are of two sorts:

- Self-intersections, where two surface segments intersect at a point in four dimensions.
- "Writhing" points, in which the vortex surface twists about in such a way as to produce four linearly independent tangent vectors to the surface.

Writhing points are best illustrated rather than described, and both types of sites generating topological charge are indicated in Fig. 6.15. One can show that zero modes of the Dirac operator tend to peak at these intersection and writhing points; the plot in Fig. 6.16 shows the modulus of Dirac zero modes in a background of four intersecting P-vortex sheets.

Calculation of the topological charge of vortex surfaces still requires the assignment of an "orientation" for the field strength, and a suitable algorithm was proposed by Engelhardt [44]. Using this proposal, Bertle et al. [47] have calculated the topological susceptibility derived from P-vortices, and have found results which are consistent with the conventional measurements of this quantity.

6.3.7 Center Symmetry Breaking by Matter Fields

We have so far only considered center vortices in theories with a center-symmetric action. But what happens in a theory where the center symmetry is explicitly broken

Fig. 6.16 Quark zero modes
in a background of intersect-
ing P-vortex sheets. From
Reinhardt et al. [46]
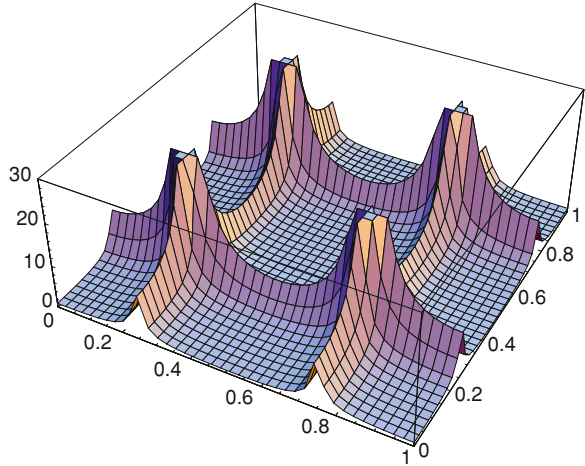
by dynamical matter fields (such as quarks) in the fundamental representation of the
gauge group? Theories of this sort are at most "confinement-like," in the sense that
the static quark potential rises linearly only up to some string-breaking scale, which
increases with the mass of the matter fields. Beyond that scale the string tension is
zero. The question is whether the assumed dominance of large-scale vacuum
fluctuations by center vortices can be consistent with this behavior.

This question is easiest to address numerically in a theory where the matter
fields are scalar, rather than fermionic, such as the $SU(2)$ gauge-Higgs theory
introduced in Sect. 3.2, Eq. 3.6. We recall that the phase diagram of this theory,
sketched in Fig. 3.4, has a line of sharp crossover behavior. The confinement-like
region (with a string tension at intermediate scales) lies below this line, and the
Higgs-like region above. In this theory we can locate vortices in the usual way, by
gauge-fixing to direct maximal center gauge and then applying center projection.
The aim is to test center dominance in the confinement-like region, but close
enough to the crossover line so that the screening effect of the scalar field is
detectable numerically. Screening is easiest to observe, in both the full and center-
projected lattices, in the expectation values of Polyakov loops, and Polyakov loop
correlators.

Center-projected Polyakov loops in the $SU(2)$ gauge-Higgs theory were com-
puted in [48]. At $\beta = 2.2$, the crossover to the Higgs-like region occurs at about
$\gamma = 0.84$, while center symmetry is unbroken at $\gamma = 0$. In the latter case the
expectation value of a Polyakov loop, like any order parameter for spontaneous
symmetry breaking, is vanishing in a finite volume. It is standard practice to
instead measure

$$\langle P \rangle \equiv \left\langle \frac{1}{L^3} \left| \sum_{\mathbf{x}} P(\mathbf{x}) \right| \right\rangle \tag{6.62}$$
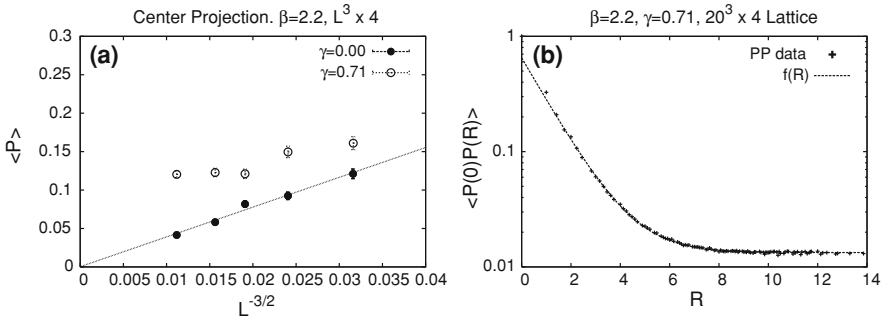
Fig. 6.17 (**a**) Center-projected Polyakov loop values in the $SU(2)$ gauge-Higgs model at $\beta = 2.2$, $\gamma = 0$ and $\gamma = 0.71$, on $L^3 \times 4$ lattice volumes with $L = 10, 12, 14, 16, 20$ The straight line is a best fit to the $\gamma = 0$ data. (**b**) Polyakov loop correlator $\langle P(0)P(R)\rangle$ on the center-projected lattice. From [48]

where $P(\mathbf{x})$ denotes the Polyakov loop passing through the point $\{\mathbf{x}, t = 0\}$. In the case of unbroken center symmetry, at $\gamma = 0$ and on an $L^3 \times L_T$ lattice, we must find

$$\langle P \rangle \propto \sqrt{\frac{1}{L^3}} \tag{6.63}$$

The reason is that if Polyakov loops have a finite correlation length l, then they must fluctuate independently in regions of extension L/l, and the number of such regions is proportional to the spatial volume L^3. This means, by the usual laws of statistics, that the magnitude of the average value of P, in any large-volume lattice configuration, must be proportional to $L^{-3/2}$. For explicitly broken center symmetry ($\gamma \neq 0$) we must instead find that $\langle P \rangle$ has a non-zero limit at $L \to \infty$.

The data for center-projected Polyakov loops in the confinement-like region at $\beta = 2.2$ and $\gamma = 0.71$, and also in the center-symmetric case of $\gamma = 0$, are shown in Fig. 6.17a, for lattice sizes $L^3 \times 4$ up to $L = 20$. The straight line is a best fit through the $\gamma = 0$ data, and errorbars for some data points are smaller than the symbol size. It is clear that the $\gamma = 0$ data is consistent with Eq. 6.63, and $\langle P \rangle$ in center projection extrapolates to zero in the infinite volume limit. At $\gamma = 0.71$ the system is still below the line of sharp crossover behavior, and in the temporary confinement region. It appears from the data that at this coupling $\langle P \rangle$ has stabilized (at $L = 14, 16, 20$) to a non-zero value of $\langle P \rangle \approx 0.120(4)$. So at $\beta = 2.2$, $\gamma = 0.71$, color screening of Polyakov loops by the matter field is detectable. A similar result was obtained in [49], for the variable Higgs-modulus version of the theory.

Figure 6.17b displays the correlator of center-projected Polyakov loops $\langle P(x)P(x + R)\rangle$ at $\beta = 2.2$, $\gamma = 0.71$, on a $20^3 \times 4$ lattice. The dashed line is a best fit to the data, for $R \geq 2$, by the function

$$f(R) = c_0 + c_1 \exp[-4\sigma R] \tag{6.64}$$

From the fit we find $c_0 = 0.0182$, $\sigma = 0.211$. Not surprisingly, c_0 is close to the square of the expectation value of the Polyakov loop in center projection. The transition from a linear static quark potential, in the region where the correlator falls exponentially, to a constant potential, occurs at around five lattice spacings. This agrees with the corresponding results on the unprojected lattice. It was also found that in the Higgs-like region, the center-projected Polyakov-loop correlator shows no evidence of a linear potential at any length scale.

What we learn from this data is that in a confinement-like theory, where the center symmetry is broken explicitly by matter fields, center vortices do not simply disappear. In the $SU(2)$ gauge-Higgs theory they appear to be responsible for the string tension at intermediate scales, while allowing for a vanishing string tension asymptotically. There is still, however, an unsettled question of exactly *how* the vortex structure in confinement-like theories differs asymptotically from the vortex structure in center-symmetric theories, so as to avoid generating an asymptotic string tension. Matter fields which break center symmetry explicitly will cause the Dirac volume associated with center vortices to acquire an action density. We therefore expect that on large scales the size of the Dirac volumes will be far smaller than in the center-symmetric theory. Since the vortices themselves do not generate an asymptotic string tension, vortex piercings of the minimal area of a Wilson loop must be correlated at large scales. It has been suggested [48] that the shrinking of the Dirac volume, and loss of an asymptotic string tension, could be achieved if the vortex surface has a branched-polymer structure at large scales, but at present this is only a speculation.

References

1. 't Hooft, G.: On the phase transition towards permanent quark confinement. Nucl. Phys. B **138**, 1–25 (1978)
2. Mack, G.: Properties of lattice gauge theory models at low temperatures. In: 't Hooft, G., et al. (eds.) Recent Developments in Gauge Theories, pp. 1–47. Plenum, New York (1980)
3. Nielsen, H.B., Olesen, P.: A quantum liquid model for the QCD vacuum: gauge and rotational invariance of domained and quantized homogeneous color fields. Nucl. Phys. B **160**, 380–396 (1979)
4. Ambjørn, J., Olesen, P.: A color magnetic vortex condensate in QCD. Nucl. Phys. B **170**, 60, 265–282 (1980)
5. Cornwall, J.: Quark confinement and vortices in massive gauge invariant QCD. Nucl. Phys. B **157**, 392–412 (1979)
6. Feynman, R.: The qualitative behavior of Yang–Mills theory in $(2 + 1)$-Dimensions. Nucl. Phys. B **188**, 479–512 (1981)
7. Del Debbio, L., Faber, M., Giedt, J., Greensite, J., Olejník, Š.: Detection of center vortices in the lattice Yang–Mills vacuum. Phys. Rev. D **58**, 094501-1–094501-15 (1998) [arXiv: hep-lat/9801027]
8. Engelhardt, M., Langfeld, K., Reinhardt, H., Tennert, O.: Deconfinement in SU(2) Yang–Mills theory as a center vortex percolation transition. Phys. Rev. D **61**, 054504-1–054504-10 (2000) [arXiv: hep-lat/9904004]
9. Faber, M., Greensite, J., Olejník, Š., Yamada, D.: The vortex finding property of maximal center (and other) gauges. JHEP **9912**, 012-1–012-22 (1999) [arXiv:hep-lat/9910033]

10. Del Debbio, L., Faber, M., Greensite, J., Olejník, Š.: Center dominance and Z(2) vortices in SU(2) lattice gauge theory. Phys. Rev. D **55**, 2298–2306 (1997) [arXiv: hep-lat/9610005]
11. Kronfeld, A., Laursen, M., Schierholz, G., Wiese, U.-J.: Monopole condensation and color confinement. Phys. Lett. B **198**, 516–520 (1987)
12. Fischer, K.H., Hertz, J.A.: Spin Glasses. Cambridge University Press, Cambridge (1991)
13. Vink, J., Wiese, U.-J.: Gauge fixing on the lattice without ambiguity. Phys. Lett. B **289**, 122–126 (1992) [arXiv:hep-lat/9206006]
14. Vink, J.: Investigation of Laplacian gauge fixing for U(1) and SU(2) gauge fields. Phys. Rev. D **51** 1292–1297 (1995) hep-lat/9407007
15. de Forcrand, Ph., Pepe, M.: Center vortices and monopoles without lattice Gribov copies. Nucl. Phys. B **598**, 557–577 (2001) [arXiv:hep-lat/0008016]
16. Alexandrou, C., de Forcrand, Ph., D'Elia, M.: The role of center vortices in QCD. Nucl. Phys. A **663**, 1031–1034 (2000) [arXiv:hep-lat/9909005]
17. Faber, M., Greensite, J., Olejník, Š.: Direct Laplacian center gauge. JHEP **11**, 53-1–53-27 (2001) [arXiv: hep-lat/0106017]
18. Bali, G., Schlichter, C., Schilling, K.: Observing long color flux tubes in SU(2) lattice gauge theory. Phys. Rev. D **51**, 5165–5198 (1995) [arXiv: hep-lat/9409005]
19. Michael, C., Teper, M.: Towards the continuum limit of SU(2) lattice gauge theory. Phys. Lett. B **199**, 95–100 (1987)
20. de Forcrand, Ph., D'Elia, M.: On the relevance of center vortices to QCD. Phys. Rev. Lett. **82**, 4582–4585 (1999) [arXiv: hep-lat/9901020]
21. Engelhardt, M., Langfeld, K., Reinhardt, H., Tennert, O.: Interaction of confining vortices in SU(2) lattice gauge theory. Phys. Lett. B **431**, 141–146 (1998) [arXiv:hep-lat/9801030]
22. Gubarev, F.V., Kovalenko, A.V., Polikarpov, M.I., Syritsyn, S.N., Zakharov, V.I.: Fine tuned vortices in lattice SU(2) gluodynamics. Phys. Lett. B **574**, 136–140 (2003) [arXiv:hep-lat/0212003]
23. Zakharov, V.I.: Nonperturbative match of ultraviolet renormalon. arXiv:hep-ph/0309178
24. Kovács, T., Tomboulis, E.: Computation of the vortex free energy in SU(2) gauge theory. Phys. Rev. Lett. **85**, 704–707 (2000) [arXiv:hep-lat/0002004]
25. von Smekal, L., de Forcrand, Ph.: 't Hooft loops, electric flux sectors and confinement in SU(2) Yang–Mills theory. Phys. Rev. D **66**, 011504-1–011504-5 (2002) [arXiv: hep-lat/0107018]
26. von Smekal, L., de Forcrand, Ph.: Electric and magnetic fluxes in SU(2) Yang–Mills theory. Nucl. Phys. Proc. Suppl. **119**, 655–657 (2003) [arXiv: hep-lat/0209149]
27. von Smekal, L., de Forcrand, Ph., Jahn, O.: More on electric and magnetic fluxes in SU(2). [arXiv: hep-lat/0212019]
28. Langfeld, K., Tennert, O., Engelhardt, M., Reinhardt, H.: Center vortices of Yang–Mills theory at finite temperatures. Phys. Lett. B **452**, 301–309 (1999) [arXiv: hep-lat/9805002]
29. Chernodub, M., Polikarpov, M., Veselov, A., Zubkov, M.: Aharonov–Bohm effect, center monopoles and center vortices in SU(2) lattice gluodynamics. Nucl. Phys. Proc. Suppl. **73**, 575–577 (1999) [arXiv: hep-lat/9809158]
30. Fujikawa, K.: Path integral measure for gauge invariant fermion theories. Phys. Rev. Lett. **42**, 1195–1198 (1979)
31. Ambjorn, J., Greensite, J., Peterson, C.: The axial anomaly and the lattice dirac sea. Nucl. Phys. B **221**, 381–408 (1983)
32. Karsten, L.H., Smit, J.: Lattice fermions: species doubling, chiral invariance, and the triangle anomaly. Nucl. Phys. B **183**, 103–140 (1981)
33. Weinberg, S.: The Quantum Theory of Fields, vol. II. Cambridge University Press, Cambridge (2005)
34. Banks, T., Casher, A.: Chiral symmetry breaking in confining theories. Nucl. Phys. B **169**, 103–125 (1980)
35. Nambu, Y., Jona-Lasinio, G.: Dynamical model of elementary particles based on an analogy with superconductivity. I. Phys. Rev. **122**, 345–358 (1961)

36. Intriligator, K.A., Seiberg N.: Lectures on supersymmetric gauge theories and electric–magnetic duality. Nucl. Phys. Proc. Suppl. **45BC**, 1–28 (1996) [arXiv:hep-th/9509066]
37. Smit, J.: Introduction to Quantum Fields on a Lattice. Cambridge University Press, Cambridge (2002)
38. DeGrand, T., DeTar, C.: Lattice Methods for Quantum Chromodynamics. World Scientific, Singapore (2006)
39. Höllwieser, R., Faber, M., Greensite, J., Heller, U.M., Olejník, Š.: Center vortices and the Dirac spectrum. Phys. Rev. D **78**, 054508-1–054508-14 (2008) [arXiv:0805.1846 [hep-lat]]
40. Adler, S., Bardeen, W.A.: Absence of higher order corrections in the anomalous axial vector divergence equation. Phys. Rev. **182**, 1517–1536 (1969)
41. Belavin, A.A., Polyakov, A.M., Shvarts, A.S., Tyupkin, Yu.S.: Pseudoparticle solutions of the Yang–Mills equations. Phys. Lett. B **59**, 85–87 (1975)
42. Witten, E.: Current algebra theorems for the U(1) Goldstone Boson. Nucl. Phys. B **156**, 269–283 (1979)
43. Veneziano, G.: U(1) without instantons. Nucl. Phys. B **159**, 213–224 (1979)
44. Engelhardt, M.: Center vortex model for the infrared sector of Yang–Mills theory: Topological susceptibility. Nucl. Phys. B **585**, 614–633 (2000) [arXiv:hep-lat/0004013]
45. Reinhardt, H.: Topology of center vortices. In: Greensite, J., Olejník, Š. (eds.) Confinement, Topology, and Other Non-Perturbative Aspects of QCD, pp. 277–285. Kluwer Academic, Dordrecht (2002) [arXiv: hep-th/0204194]
46. Reinhardt, H., Schröder, O., Tok, T., Zhukovsky, V.: Quark zero modes in intersecting center vortex gauge fields. Phys. Rev. D **66**, 085004-1–085004-12 (2002) [arXiv: hep-th/0203012]
47. Bertle, R., Engelhardt, M., Faber, M.: Topological susceptibility of Yang–Mills center projection vortices. Phys. Rev. D **64**, 074504-1–074504-10 (2001) [arXiv:hep-lat/0104004]
48. Greensite, J., Olejnik, S.: Vortices, symmetry breaking and temporary confinement in SU(2) gauge-Higgs theory. Phys. Rev. D **74**, 014502-1–014502-9 (2006) [arXiv:hep-lat/0603024]
49. Bertle, R., Faber, M., Greensite, J., Olejnik, S.: Center dominance in SU(2) gauge-Higgs theory. Phys. Rev. D **69**, 014007-6 (2004) [arXiv:hep-lat/0310057]

Chapter 7
Confinement from Center Vortices II

There are still a number of important questions relating to center vortices and confinement which must be addressed. First in importance is the issue of Casimir scaling. We have seen that there is some intermediate range of distances for which the string tension is proportional to the quadratic Casimir of the quark color representation; the dependence shifts to N-ality only at asymptotic distances. How is Casimir scaling to be understood, in the context of the center vortex mechanism? A second question is whether vortices can be understood as the classical solutions of some effective gauge theory action, which is relevant to large scales. Finally, is it possible to devise a simple theory of fluctuating vortex surfaces, which would describe the behavior of non-abelian gauge theories in the infrared regime?

7.1 Casimir Scaling and Vortex Thickness

Let's begin with Casimir scaling. Although the asymptotic string tension only depends on N-ality, so that for $SU(2)$

$$\sigma_j = \begin{cases} \sigma_{1/2} & j = \text{half-integer} \\ 0 & j = \text{integer} \end{cases}, \tag{7.1}$$

there is still an intermediate range of distances where Casmir scaling applies (at least approximately), i.e. for $SU(2)$

$$\sigma_j \propto j(j+1). \tag{7.2}$$

How do vortices fit in, since they are motivated by (and seem to only give rise to) N-ality dependence?

The essential observation is that while $j = $ integer Wilson loops are unaffected by *thin* center vortices, they can be affected by *thick* center vortices, if the vortex core overlaps the loop. But how thick are center vortices? We have already seen that, in $D = 3$ dimensions, the observed adjoint string breaking occurs at 1.25 fm

[1], when "physical units" are obtained by fixing the string tension at $\sigma = (440 \text{ MeV})^2$. It was also found, in $D = 4$ dimensions, that the center vortex free energy drops to zero at around 1.25 fm [2]. Finally, from Fig. 6.3 of Chap. 6, one may estimate that $W_1/W_0 \approx -1$ for 5×5 vortex-limited Wilson loops, which translates, at $\beta = 2.3$ into an extension of roughly 0.85 fm. Each of these observations suggests a vortex thickness of around one fermi. This is not so small, and the question is whether this finite thickness makes a difference for Wilson loops of dimensions less than one fermi.

A simple model, which shows how Casimir scaling at intermediate distances might emerge due to the finite thickness of vortices, was put forward in [3], improving on an earlier suggestion in [4].[1] Consider the projection of a vortex onto the plane of a planar Wilson loop; we will refer to this projection as a vortex "domain" in the plane. In this model, it is assumed that the effect of a domain (2D cross-section of a vortex) on a planar Wilson loop holonomy is to multiply the holonomy by a group element

$$G(\alpha^n, S) = S \exp[i\alpha^n \cdot \mathbf{H}]S^\dagger, \tag{7.3}$$

where the $\{H_i\}$ are generators of the Cartan subalgebra (the largest subset of group generators which commute with one another), S is a random group element, $\{\alpha_i^n\}$ depends on the location of the domain relative to the loop, and n indicates the domain type. If the domain lies entirely within the planar area enclosed by the loop, then

$$\exp[i\alpha^n \cdot \mathbf{H}] = z_n \mathbb{1}_N , \tag{7.4}$$

where

$$z_n = e^{2\pi i n/N} \in Z_N \tag{7.5}$$

and $\mathbb{1}_N$ is the $N \times N$ unit matrix. At the other extreme, if the domain is entirely outside the planar area enclosed by the loop, then

$$\exp[i\alpha^n \cdot \mathbf{H}] = \mathbb{1}_N . \tag{7.6}$$

For a Wilson loop in representation r, the average contribution from a domain will be

$$\begin{aligned}
\mathscr{G}_r[\alpha^n] \mathbb{1}_{d_r} &= \int dS \, S \exp[i\alpha^n \cdot \mathbf{H}]S^\dagger \\
&= \frac{1}{d_r} \chi_r \left[\exp[i\alpha^n \cdot \mathbf{H}] \right] \mathbb{1}_{d_r} ,
\end{aligned} \tag{7.7}$$

where d_r is the dimension of representation r, $\mathbb{1}_{d_r}$ is the $d_r \times d_r$ unit matrix, and χ_r is the group character in representation r.

Consider, e.g., $SU(2)$ lattice gauge theory, choosing $H = L_3$. The center subgroup is Z_2, and there are two types of domains, corresponding to $z_0 = 1$ and

---

[1] See also Deldar [5]. A still earlier idea along these lines was put forward by Cornwall [6].

$z_1 = -1$. Let f_1 represent the probability that the midpoint of a z_1 domain is located at any given plaquette in the plane of the loop, with f_0 the corresponding probability for a $z_0 = 1$ domain. Let us also assume that the probabilities of finding domains of either type centered at any two plaquettes in a plane are independent. Then, for a planar loop C,

$$W_j(C) \approx \prod_x \left\{ (1 - f_1 - f_0) + f_1 \mathscr{G}_j[\alpha_C^1(x)] + f_0 \mathscr{G}_j[\alpha_C^0(x)] \right\} W_j^{pert}(C)$$

$$= \exp\left[\sum_x \log\left\{ (1 - f_1 - f_0) + f_1 \mathscr{G}_j[\alpha_C^1(x)] \right. \right.$$

$$\left. \left. + f_0 \mathscr{G}_j[\alpha_C^0(x)] \right\} \right] W_j^{pert}(C), (7.8)$$

where the product and sum over x runs over all plaquette positions in the plane of the loop C, and $\alpha_C^n(x)$ depends on the position of the vortex midpoint x relative to the location of loop C. The expression $W_j^{pert}(C)$ contains the short-distance, perturbative contribution to $W_j(C)$; this will just have a perimeter-law falloff.

The question is what to use as an ansatz for $\alpha_C^n(x)$. The ansatz of [3] was motivated by the idea that the magnetic flux in the interior of vortex domains fluctuates almost independently, in subregions of extension l, apart from the restriction that the total flux results in a center element. For the $SU(2)$ gauge group, this leads to

$$\left(\alpha_C^1(x) \right)^2 = \frac{A_v}{2\mu} \left[\frac{A}{A_v} - \frac{A^2}{A_v^2} \right] + \left(2\pi \frac{A}{A_v} \right)^2$$

$$\left(\alpha_C^0(x) \right)^2 = \frac{A_v'}{2\mu} \left[\frac{A}{A_v'} - \frac{A^2}{A_v'^2} \right], \tag{7.9}$$

where A_v, A_v' are the cross-sectional areas of the $n = 1$ and $n = 0$ domains, respectively, and A is the area of the domain which is contained within the interior of the minimal area of the loop. The parameter μ controls the magnitude of the magnetic fluctuations in the subregions. The result is a vortex-induced potential which is proportional to the quadratic Casimir of the representation at small R (compared to the vortex thickness), and depends only on the N-ality at large R. A sample calculation for the $j = \frac{1}{2}, 1, \frac{3}{2}$ representations is shown in Fig. 7.1. Details of the calculation are found in [3].

Casimir scaling for the smaller loops is not an accident, and not specific to the $SU(2)$ gauge group. In generalizing to $SU(N)$ we must take into account the fact that there are N types of center domains, so that

$$\langle W_r(C) \rangle \approx \prod_x \left\{ 1 - \sum_{n=0}^{N-1} f_n \left(1 - \mathrm{Re}\,\mathscr{G}_r[\alpha_C^n(x)] \right) \right\} W_r^{pert}(C)$$

$$= \exp\left[\sum_x \log\left\{ 1 - \sum_{n=0}^{N-1} f_n \left(1 - \mathrm{Re}_r[\alpha_C^n(x)] \right) \right\} \right] W_r^{pert}(C), \tag{7.10}$$
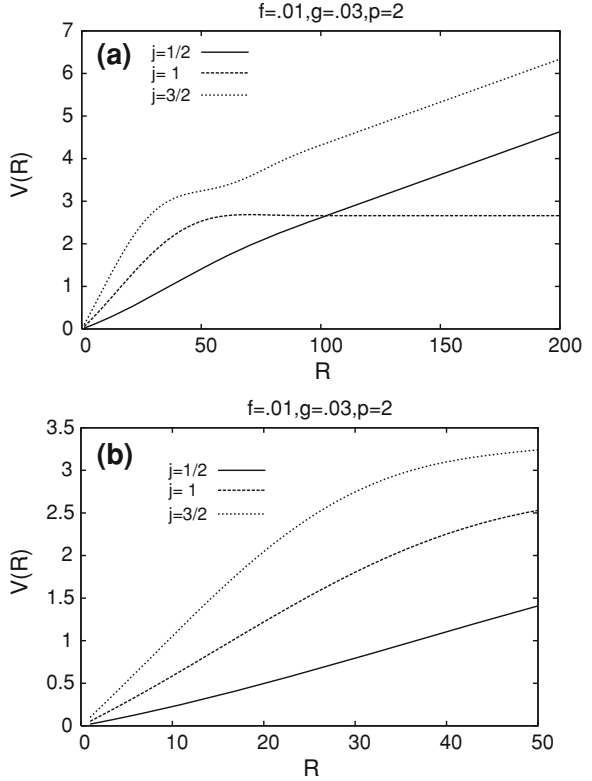
where

$$f_n = f_{N-n}, \quad \mathscr{G}_r[\alpha^n] = \mathscr{G}_r^*[\alpha^{N-n}]. \tag{7.11}$$

For very small loops, which are associated with small α, we have that

$$1 - \mathscr{G}_r[\alpha^n] \approx \frac{1}{2}\alpha_i^n \alpha_j^n \frac{1}{d_r}\mathrm{Tr}[H^i H^j]$$

$$= \frac{1}{2(N^2 - 1)}\alpha^n \cdot \alpha^n C_r, \tag{7.12}$$

where C_r is the quadratic Casimir for representation r. Once again, if $\alpha^n \cdot \alpha^n$ is proportional to the area of the vortex in the interior of the loop, for small loops, then the result for small loops is a linear potential whose string tension is proportional to the quadratic Casimir. For large loops, if the vortex domain lies entirely in the loop interior, we must have

$$\frac{1}{d_r}\chi_r[\exp[i\alpha^n \cdot \mathbf{H}]] = e^{ikn\pi/N}, \tag{7.13}$$

where k is the N-ality of representation r. Again we obtain a linear potential, whose string tension depends only on the N-ality. It is possible to choose the f_n so that for very large loops the k-string tensions follow either the Sine Law, or are

proportional to the Casimir $C_{r_{min}}$ of the lowest dimensional representation of N-ality k; cf. [7] for a full discussion of this point.

G_2 gauge theory is a special case of particular interest, because the center of the G_2 group consists only of the unit element. In this theory there would be only one type of vortex domain, corresponding to the single element of the trivial center group. One would then expect all color representations to give rise to Casimir scaling in the intermediate distance regime, and flatten out asymptotically. Casimir scaling in this theory has in fact been observed, cf. [8].

7.2   What About "Gluon Confinement?"

Center vortices account for the asymptotic string tension of the static quark potential. This asymptotic string tension depends only on the N-ality, not the dimension, of the color representation, and the string tension of N-ality $= 0$ representations is zero, as it should be. But then, one may ask, what about gluons? These are in the adjoint representation, so their N-ality is zero, yet gluons are not members of the asymptotic particle spectrum. So what accounts for the "confinement" of gluons?

To understand this point, it is worth reflecting on the fact that one would never encounter superheavy nuclei, of some very large but fixed baryon number A, with charge greater than or equal to some critical amount $Q_c \ll A$, even if such nuclei were stable to nuclear decay [9]. The reason is simple: the vacuum of quantum electrodynamics at any instant is a superposition of states containing virtual e^+e^- pairs, as seen in Fig. 7.2. If a superheavy nucleus–antinucleus pair were somehow created with electric charges exceeding the critical value (Fig. 7.3), then the

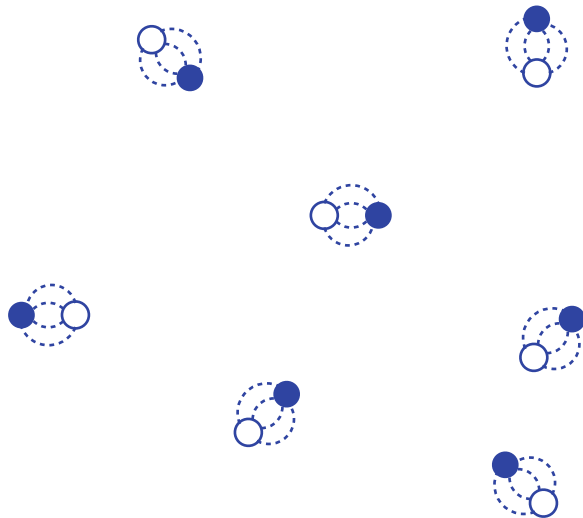**Fig. 7.2** Virtual electron–positron pairs in the QED vacuum

Fig. 7.3 Introduction of
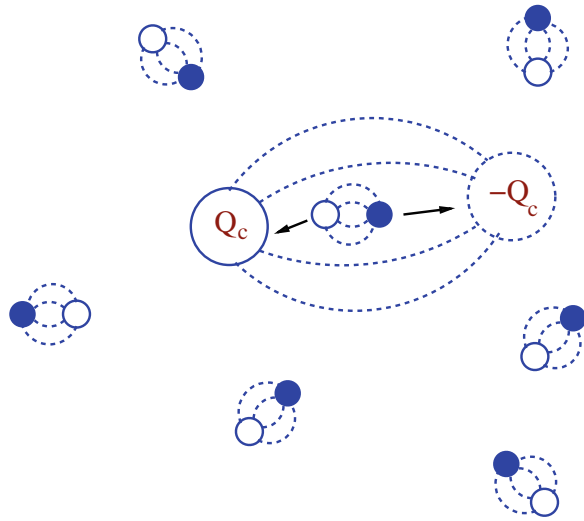nuclei with charges $\pm Q_c$
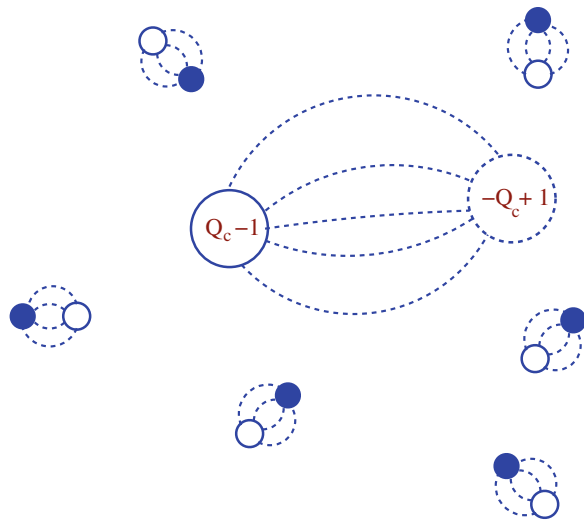beyond the threshold value

Fig. 7.4 Response of the
vacuum is to screen the
nuclear charge below the
critical value

associated electric field energy would be so strong that it is energetically favorable
for some of the virtual electrons and positrons to bind to these objects, screening
their charges down below the critical value (Fig. 7.4). Nuclear charge at Q_c and
above is therefore "confined" in nuclei of fixed baryon number A, in the sense that
nuclei with such high electric charges will never be observed.

   Similar reasoning (illustrated in Figs. 7.5, 7.6, 7.7) explains why isolated par-
ticles such as gluons, with N-ality = 0 color charge, are absent in the asymptotic
spectrum. The QCD vacuum contains, in addition to virtual quarks and antiquarks,
also virtual gluons. As two particles in the adjoint representation become widely
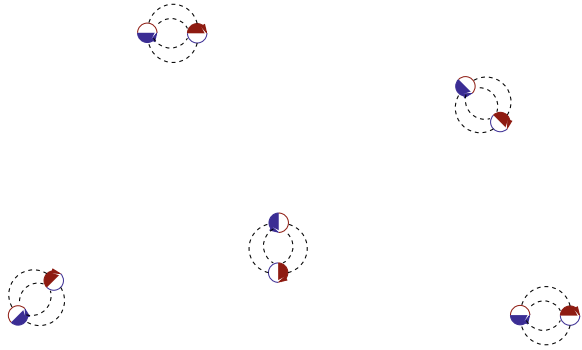
Fig. 7.5 Virtual gluons in
the QCD vacuum

Fig. 7.6 Two gluons (or
other particles of adjoint
color charge) become widely
separated, momentarily. The
effective coupling at this
separation is large, and the
associated color electric field
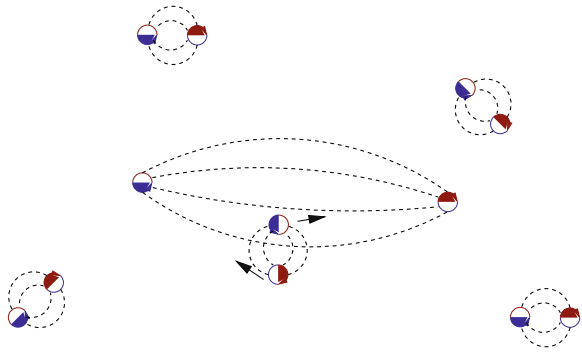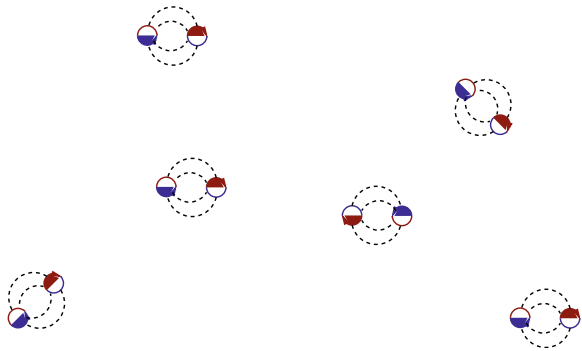energy is large

Fig. 7.7 Response of the
nearby gluons is to reduce the
color charge separation via
the mechanism of string
breaking

separated and enter the Casimir scaling regime, the energy of the color electric
field becomes large. At some point it becomes energetically favorable for virtual
gluons to bind to the sources, resulting in two color-singlet states. This is essen-
tially the string-breaking mechanism discussed previously. While it is only a
matter of semantics whether one calls this process confinement, I prefer, for
reasons discussed in Chap. 3, to refer to it as color screening, and reserve the term
"confinement" for magnetic disorder.

These remarks about gluon confinement also apply to G_2 gauge theory. As already noted, the center of G_2 is trivial, and this has led some authors [10] to argue that since G_2 is a confining gauge theory with a trivial center, center symmetry (and the vortex mechanism) cannot be necessary for confinement. However, since all of the usual order parameters for confinement, such as Wilson loops and Polyakov loops, exhibit a non-confining behavior in the G_2 case, it brings us back to the question of what is meant by the term "confinement." If all that is meant is that the theory has a color-singlet spectrum, then it is certainly true that vortices are not required. As we have seen, a color-singlet spectrum is also obtained in a gauge-Higgs theory. If, on the other hand, by confinement we mean a certain property of the vacuum state, i.e. magnetic disorder at arbitrarily large scales, then G_2 gauge theory does not have this property. From this point of view, the G_2 gauge group is better regarded an example of, rather than a counter-example to, the relevance of vortices to confinement [3]. No center vortices means no asymptotic string tension. The absence of color non-singlet quarks in G_2 gauge theory is no different, in principle, from the absence of colored gluons in $SU(N)$ theory. In both cases, the absence of colored states can be attributed to a color screening mechanism, rather than the large-scale vacuum fluctuations of gauge fields, which are required for magnetic disorder at all distance scales.

7.3 The Random Surface Model

We have seen that center projected field configurations can account for most of the long-range physics associated with Yang-Mills theories; i.e. confinement, chiral symmetry breaking, topological susceptibility, and spatial string tension in the deconfinement phase. This raises a natural question: Can one construct a field theory of vortices alone, which would capture the same physics? Such a theory was introduced by Engelhardt and Reinhardt [11], and further developed by Engelhardt and collaborators [12–15, 18].

Vortices are surfaces in $D = 4$ dimensions, so the Engelhardt–Reinhardt model is a theory of fluctuating random surfaces; i.e. a kind of string theory. In its simplest form, the model contains only one dimensionless coupling c, and a fixed lattice spacing a. The finite lattice spacing corresponds to the fact that center vortices in Yang-Mills configurations have a finite thickness; it therefore makes no sense to take the $a \to 0$ limit, which would describe infinitely thin vortices. The vortex sheets, in the $SU(2)$ gauge theory, are closed surfaces on a hypercubic lattice, and the lattice action is simply

$$S_V = cN_{bend}, \tag{7.14}$$

where N_{bend} is the number of cases in which two neighboring vortex plaquettes share a link, but lie in orthogonal planes. In other words, the action density is proportional to the extrinsic curvature of the vortex worldsheet. This is in contrast to a Nambu-style string action, in which the action is proportional to the area of the

world sheet. In fact one could add such a term, with an associated coupling s, but Engelhardt and Reinhardt have shown that there is a line of constant physics in the c–s coupling plane, so one may as well drop the Nambu term and just work with the extrinsic curvature.

In the $SU(2)$ version of the random surface model, one evaluates Wilson loops in the obvious way, assigning a value $(-1)^n$ to a Wilson loop, where n is the number of vortex piercings of the minimal area of the loop, and averaging this quantity over an ensemble of vortex configurations generated, via Monte Carlo simulation, with the action (7.14). One can vary temperature by varying the extension of the lattice in the time direction, and, in the $SU(2)$ case, a second-order phase transition is found. Then the coupling c is fixed to reproduce the dimensionless quantity $T_C/\sqrt{\sigma}$, where T_C is the deconfinement transition temperature, and σ is the string tension, with both quantities obtained in the random surface model in lattice units. Given c, the lattice spacing a is fixed by setting $\sigma/a^2 = (440 \text{ MeV})^2$.

With coupling c and lattice spacing a in hand, the random surface model becomes predictive, and the following results have been obtained for the $SU(2)$ version of the theory:

- The spatial string tension in the deconfined phase vs. temperature [11]. The values are in good agreement with the full theory.
- The topological susceptibility [12] (Fig. 7.8). Again the values are in good agreement with the full theory, e.g. the topological susceptibility χ in the full theory, at zero temperature, is known to be $\chi^{1/4} \approx 190 \text{ MeV}$.
- The chiral condensate [13].

It was already mentioned that for the $SU(2)$ theory, the model finds a second order phase transition, in agreement with the known result for the full theory. The random surface model can be extended to the $SU(3)$ [14, 15] and larger gauge groups, and for $SU(3)$ it is found that the deconfinement transition is weakly first-order, and, for baryons, the area law behaves as though the electric flux tubes between quarks are in a "Y" configuration, rather than a Δ configuration, as shown

Fig. 7.8 Fourth root of the topological susceptibility vs. temperature, in the random surface model. From Engelhardt [12]
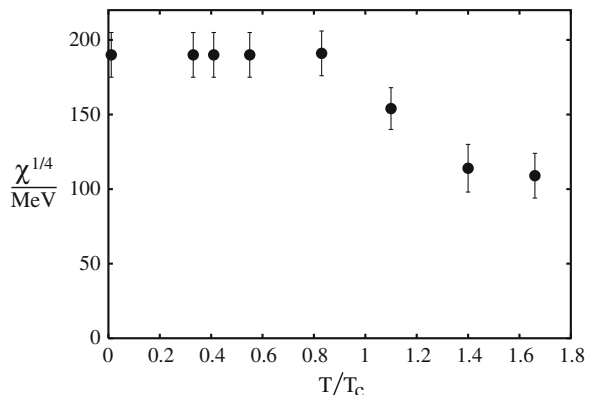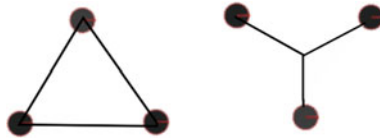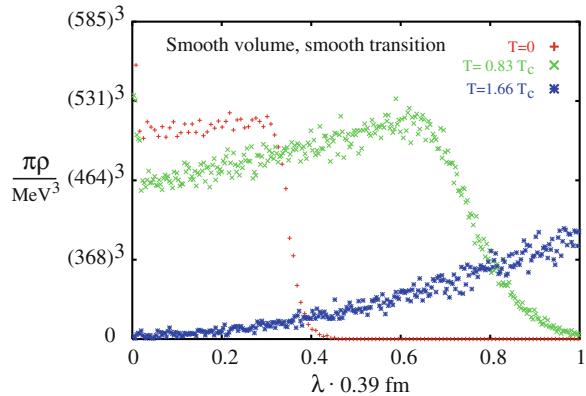
Fig. 7.9 The electric flux tube between three static quarks in an $SU(3)$ gauge theory can have either the "Δ" (*left*) or the "Y" (*right*) configuration shown. The random surface model predicts a "Y" configuration, in agreement with the known result [16, 17] found via standard Monte Carlo simulations

Fig. 7.10 Dirac spectral density in the random vortex model. From Engelhardt [13]

in Fig. 7.9. These features are also in agreement with the known properties of $SU(3)$ lattice gauge theory.

In order to measure the topological susceptibility, it is necessary to measure the topological charge density. As already noted in the previous chapter, a non-zero topological charge density is associated with vortex plaquettes which meet at a point, and which are orthogonal to one another; these are the "writhing" and self-intersection points displayed in Fig. 6.15. However, in the Z_2 random surface model there is no indication of the sign of $F_{\mu\nu}$, which is needed to define the topological charge. This is handled by the surprising procedure of assigning orientations (signs) stochastically to vortex plaquettes; this amounts to inserting monopole lines on the vortex surfaces with some weighting. One would think that this would introduce some new coupling parameter, and that while one might be able to fit a known value of χ, there would be no predictive power. It turns out however, for geometrical reasons explained in [12], that χ is insensitive to the density of monopole lines, and therefore the measurement of this quantity is a genuine prediction of the model, whose result is displayed in Fig. 7.8.

As also noted in the last section, the chiral condensate $\langle \overline{\psi}\psi \rangle$ is proportional, according to the Banks-Casher relation, to the eigenvalue density $\rho(0)$ of the zero modes of the Dirac operator. This has also been computed in the random surface model, by assigning a continuum gauge-field configuration to a vortex surface, and evaluating the continuum Dirac operator in this configuration. The results derived

from the model are displayed in Fig. 7.10, and we see that the model predicts $\rho(0) > 0$ at temperatures below the deconfinement transition, and therefore a non-vanishing chiral condensate and chiral symmetry breaking, while $\rho(0) = 0$ above the transition temperature, which means that chiral symmetry is restored.

In general, the random surface model works remarkably well for $SU(2)$ and $SU(3)$ gauge theories, with only a single dimensionless coupling and lattice spacing required to account for a wide range of non-perturbative phenomena. For larger gauge groups, where there is more than one type of vortex and various vortex joinings are possible, it seems to be necessary to introduce additional couplings to describe the data, and consequently the model is much less predictive in those cases [18].

7.4   Vortices as Solitons

The only classical solutions which are known in pure gauge theories are the instanton solutions discovered by Belavin et al. [19], and their finite temperature variants known as *calorons*, to be discussed in the next chapter. If matter fields are added, there are other interesting possibilities. In particular, in an $SU(N)$ gauge theory, suppose we add enough matter fields in the adjoint color representation to give a mass, via the Higgs mechanism, to all of the gauge bosons. In that case, the theory typically contains center vortices as classical solutions.

A particularly interesting case has long been advocated by Cornwall [20] on the grounds that the dynamics of a pure gauge field should generate a mass for the gluon, and that this effect could be represented by an effective local action with a gauge-invariant mass term. Consider the action

$$S = \int d^4x \left\{ \frac{1}{2g^2} \text{Tr}[F_{\mu\nu}^2] + \frac{m^2}{g^2} \text{Tr}[(D_\mu U)U^{-1}]^2 \right\}, \qquad (7.15)$$

where $U(x)$ is an $N \times N$ unitary-matrix valued field, transforming under an $SU(N)$ gauge transformation G as $U(x) \rightarrow G(x)U(x)$. This theory has center vortex solutions [20, 21], specified by a closed two-surface S, and a gauge field given by the integral over this surface:

$$A_\mu(x) = -2\pi i Q_a \varepsilon_{\mu\nu\alpha\beta} \partial_\nu \int d\sigma_{\alpha\beta} \frac{1}{2} [\Delta_m(x - x'(\sigma)) - \Delta_0(x - x'(\sigma))], \qquad (7.16)$$

where $\Delta_{m,0}(x)$ is the standard propagator for a free scalar of mass m or mass zero, respectively, and Q_a is any one of the diagonal matrices

$$Q_a = diag\left[\frac{1}{N}, \frac{1}{N}, \ldots, -1 + \frac{1}{N}, \frac{1}{N}, \ldots\right] \qquad (7.17)$$

with the $-1 + 1/N$ entry in the a-th position. These matrices have the property that

$$e^{2\pi i Q_a} = e^{2\pi i/N}. \tag{7.18}$$

In the special case that the 2-surface is the $z - t$ plane (so at a fixed time the middle of the center vortex runs along the z-axis), the solution (in cylindrical coordinates z, ρ, ϕ) is

$$\mathbf{A}_\mu(\mathbf{x}) = -i\widehat{\phi}Q_a\left[\frac{1}{\rho} - mK_1(m\rho)\right], \tag{7.19}$$

where K_1 is a modified Bessel function. It is not hard to see that for large Wilson loops (at large ρ, in this example) which are linked topologically to the surface S, the loop holonomy $P\exp[i\oint dx_\mu A_\mu]$ approximates the center element $\exp[2\pi i/N]$. Center vortices carrying larger amounts of center flux are constructed by replacing a single Q_a with a sum of different Q_a's with different values of a.

The effective theory specified by (7.15) is quite rich. In addition to vortex solutions, the theory also contains "nexus" solutions [21] which correspond to magnetic monopoles running along the vortex sheets. These have, in fact, been seen in numerical simulations (as will be discussed in the next chapter) and are required for the generation of topological charge, as realized independently by Engelhardt and Reinhardt [12, 22]. In the case of $SU(3)$, the effective theory predicts a "Y" configuration for baryonic (i.e. three-quark) flux-tube states [23], in agreement with numerical simulations [16, 17].

7.5 Critique

I have presented a fair amount of evidence in favor of the center vortex scenario of confinement. There are still, however, some unsettled issues.

First of all, in numerical simulations, there is the question of what to do about Gribov copies in maximal center gauge. Should these be averaged over, or should one attempt to fix to a "best" copy? Apart from a question of principle, the strategy chosen makes a substantial quantitative difference (perhaps 25%) to the projected string tension, with the "average copy" procedure yielding the better results. Secondly, the vortex scenario gives us no real insight into the origin of the Lüscher term in the static quark potential. Presumably this term is hidden in some subtle correlations among vortex piercings, but we can't claim that it appears naturally. Finally, I have mainly shown numerical results for the $SU(2)$ gauge group. The corresponding numerical results in $SU(3)$ have both good and less good features, from the point of view of the vortex scenario. In $SU(3)$ the data for vortex-limited Wilson loops does trend to the desired result [24]

$$\frac{W_1(C)}{W_0(C)} \rightarrow \exp[\pm 2\pi i/3] \tag{7.20}$$

for large loops, and vortex removal sends the string tension to zero. Both of these findings support the vortex picture. On the other hand, in $SU(3)$ the center-projected string tension comes out to only about $\frac{2}{3}$ the full string tension [24], which is substantially worse than for the $SU(2)$ case.

References

1. de Forcrand, P., Kratochvila, S.: Observing string breaking with Wilson loops. Nucl. Phys. B **671**, 103–132 (2003) [arXiv:heplat/0306011]
2. Kovács, T., Tomboulis, E.: Computation of the vortex free energy in SU(2) gauge theory. Phys. Rev. Lett. **85**, 704–707 (2000). arXiv:hep-lat/0002004
3. Greensite, J., Langfeld, K., Olejník, Š., Reinhardt, H., Tok, T.: Color screening, Casimir scaling, and domain structure in G(2) and SU(N) gauge theories. Phys. Rev. D **75**, 034501-1–034501-25 (2007). arXiv:hep-lat/0609050
4. Faber, M., Greensite, J., Olejnik, S.: Casimir scaling from center vortices: Towards an understanding of the adjoint string tension. Phys. Rev. D **57**, 2603–2609 (1998). arXiv:hep-lat/9710039
5. Deldar, S.: Potentials between static SU(3) sources in the fat center vortices model. JHEP **0101**, 013-1–013-9 (2001). arXiv:hep-ph/9912428
6. Cornwall, J.M.: In: Milton, K.A., Samuel, M.A. (eds.) Proceedings of the Workshop on Non-Perturbative Quantum Chromodynamics. Birkhäuser, Stuttgart (1983)
7. Greensite, J., Olejnik, S.: k-string tensions and center vortices at large N. JHEP **0209**, 039-1–039-22 (2002). arXiv:hep-lat/0209088
8. Liptak, L., Olejnik, S.: Casimir scaling in G(2) lattice gauge theory. Phys. Rev. D **78**, 074501-1–074501-14 (2008). arXiv:0807.1390 [hep-lat]
9. Madsen, J.: Universal charge-radius relation for subatomic and astrophysical compact objects. Phys. Rev. Lett. **100**, 151102-1–151102-4 (2008)
10. Holland, K., Minkowski, P., Pepe, M., Wiese, U.J.: Exceptional confinement in G(2) gauge theory. Nucl. Phys. B **668**, 207–236 (2003). arXiv:hep-lat/0302023
11. Engelhardt, M., Reinhardt, H.: Center vortex model for the infrared sector of Yang–Mills theory: Confinement and deconfinement. Nucl. Phys. B **585**, 591–613 (2000). arXiv:hep-lat/9912003
12. Engelhardt, M.: Center vortex model for the infrared sector of Yang–Mills theory: Topological susceptibility. Nucl. Phys. B **585**, 614–633 (2000). arXiv:hep-lat/0004013
13. Engelhardt, M.: Center vortex model for the infrared sector of Yang–Mills theory: Quenched Dirac spectrum and chiral condensate. Nucl. Phys. B **638**, 81–110 (2002). arXiv:hep-lat/0204002
14. Engelhardt, M.: Center vortex model for the infrared sector of SU(3) Yang–Mills theory—baryonic potential. Phys. Rev. D **70**, 074004-1–074004-8 (2004). arXiv:hep-lat/0406022
15. Quandt, M., Reinhardt, H., Engelhardt, M.: Center vortex model for the infrared sector of SU(3) Yang–Mills theoryvortex free energy. Phys. Rev. D **71**, 054026-1–054026-8 (2005). arXiv:hep-lat/0412033
16. Takahashi, T.T., Suganuma, H.: Detailed analysis of the gluonic excitation in the three-quark system in lattice QCD. Phys. Rev. D **70**, 074506-1–074506-13 (2004). arXiv:hep-lat/0409105
17. Alexandrou, C., de Forcrand, P., Jahn, O.: The ground state of three quarks. Nucl. Phys. Proc. Suppl. **119**, 667–669 (2003). arXiv:hep-lat/0209062
18. Engelhardt, M., Sperisen, B.: Confinement and center vortex dynamics in different gauge groups. AIP Conf. Proc. **892**, 176–179 (2007). arXiv:hep-lat/0610122
19. Belavin, A.A., Polyakov, A.M., Shvarts, A.S., Tyupkin, Yu.S.: Pseudoparticle solutions of the Yang–Mills equations. Phys. Lett. B **59**, 85–87 (1975)

20. Cornwall, J.: Quark confinement and vortices in massive gauge invariant QCD. Nucl. Phys. B **157**, 392–412 (1979)
21. Cornwall, J.M.: On topological charge carried by nexuses and center vortices. Phys. Rev. D **65**, 085045-1–085045-12 (2002). arXiv:hep-th/0112230
22. Engelhardt, M., Reinhardt, H.: Center projection vortices in continuum Yang–Mills theory. Nucl. Phys. B **567**, 249–292 (2000). arXiv:hep-th/9907139
23. Cornwall, J.M.: On the center vortex baryonic area law. Phys. Rev. D **69**, 065013-1–065013-13 (2004). arXiv:hep-th/0305101
24. Langfeld, K.: Vortex structures in pure SU(3) lattice gauge theory. Phys. Rev. D **69**, 014503-1–014503-12 (2004). arXiv:hep-lat/0307030

Chapter 8
Monopoles, Calorons, and Dual Superconductivity

In a type II superconductor, magnetic fields are collimated into flux tubes known as *Abrikosov vortices*. If stable magnetic monopoles existed, and could be pair produced within a type II superconductor, then a magnetic flux tube would run between widely separated monopoles and antimonopoles, and the static monopole potential would rise linearly with monopole separation.[1] It was suggested by 't Hooft [2] and Mandelstam [3] in the mid 1970's that the QCD vacuum might resemble a type II superconductor, except with a reversal of the roles of the electric and magnetic fields. Instead of a condensate of electrically charged particles (the Cooper pairs of a superconductor), there would be a condensate of magnetic charge; and magnetic monopoles were invoked to fill this role. Then, instead of collimation of the color magnetic field, it would be the color electric field that would be squeezed into a flux tube, and color electric, rather than magnetic, charges would be confined by a linear potential.

This idea of "dual" superconductivity is simple and attractive, and has had an enormous influence on people's thinking about confinement, but it is essentially an abelian mechanism, because the magnetic monopoles of interest are normally defined with respect to abelian groups. In the context of a non-abelian gauge theory, it is necessary to single out an abelian subgroup, and this is done with the help of either a Higgs field, or a composite operator, in the adjoint representation of the gauge group. This approach has certain difficulties with N-ality, which I will come to. But let us begin with a success: Polyakov's demonstration, in 1975 [4], that magnetic monopoles are responsible for confinement in compact QED in $D = 3$ dimensions.

---

[1] For a review of magnetic monopole solutions in abelian and non-abelian gauge theories, cf. Coleman [1].

8.1 Magnetic Monopoles in Compact QED

Compact QED in three and four dimensions has monopole excitations, and these excitations are responsible for the confinement of electric charge. The confinement property exists only at strong lattice couplings in $D = 4$ dimensions, but it is found at all lattice couplings in $D = 3$ dimensions. The word "compact" refers to the compactness of the $U(1)$ gauge group in the lattice (as opposed to the continuum) formulation of electrodynamics. In $D = 3$ Euclidean dimensions it is sometimes useful to continue referring to $B_i(x) = \frac{1}{2}\varepsilon_{ijk}F_{jk}$ (rather than only the F_{12} component) as the "magnetic" field, and its integral over an area as the magnetic flux. I will follow that convention in this chapter.

Consider a plaquette variable $U(p) = \exp[i\theta(p)]$, i.e., the product of link variables around a plaquette, in $U(1)$ lattice gauge theory in $D = 3$ dimensions (compact QED_3). Obviously the lattice action is minimized for both magnetic flux zero, $\theta(p) = 0$, and magnetic flux $\theta(p) = 2\pi$. This double minimum is a reflection of the compactness of the gauge group in the lattice formulation. Suppose we have a series of parallel plaquettes with $\theta(p) = 2\pi$ arranged along, e.g., the positive x-axis. A line piercing this series of plaquettes can be thought of as an infinitely thin solenoid, known as a "*Dirac line*," of flux 2π. At $x = 0$ this 2π flux cannot simply disappear, but must spread out through the plaquettes of a cube, one side of which has the incoming magnetic flux 2π (cf. Fig. 8.1). In other words, 2π flux enters through one plaquette of the cube, and then exits through the remaining sides of the cube. The flux which leaves the cube also adds up to 2π, but is less than that amount at any given plaquette. The 2π flux of the Dirac line is undetectable, and in fact the position of the Dirac line can be moved around via $U(1)$ gauge transformations, but the endpoint of the line is gauge invariant, and appears as a point-like source of magnetic field; i.e., a magnetic monopole. The Dirac line either goes off to infinity or, on a finite periodic lattice, ends in an antimonopole. In $D = 3$ dimensions monopoles are point-like objects, i.e., instantons. In $D = 4$ dimensions monopoles trace out trajectories, and are line-like objects.
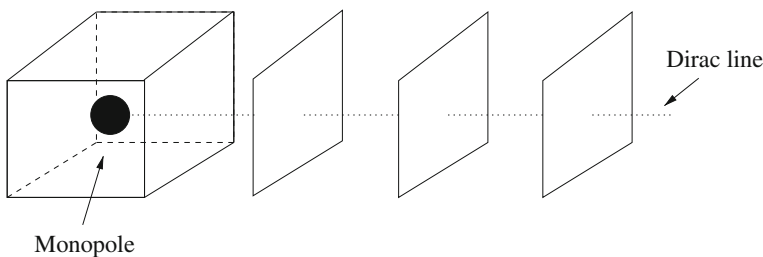

Fig. 8.1 In compact QED$_3$, a *Dirac line* is a line piercing the middle of plaquettes with flux $\Phi_B = 2\pi/e$. The line is bounded at one end by a Dirac monopole, and at the other by an antimonopole

Given a lattice configuration $U_\mu(x) = \exp[i\theta_\mu(x)]$, how do we locate the monopoles and determine their charges? The basic idea is to use Gauss's Law. Let's start in either $D = 3$ dimensions, or in a three-dimensional timeslice of a $D = 4$ dimensional lattice, and define the flux through a plaquette

$$f_{\mu\nu}(x) = \partial_\mu\theta_\nu(x) - \partial_\nu\theta_\mu(x), \tag{8.1}$$

where, in this case, the partial derivatives must be defined as "forward" lattice derivatives, i.e., finite differences

$$\partial_\mu\theta_\nu(x) = \theta_\nu(x + \hat{\mu}) - \theta_\nu(x). \tag{8.2}$$

Now if we would simply add up the flux around a cube (or any closed surface), it would just sum to zero as an identity. The trick is to remove from $f_{\mu\nu}(x)$ any flux which is due to a Dirac string. If that flux is removed, then the remaining magnetic flux is no longer necessarily divergence free. So, following DeGrand and Toussaint [5], let us express $f_{\mu\nu}$ as

$$f_{\mu\nu}(x) = \overline{f}_{\mu\nu}(x) + 2\pi n_{\mu\nu}(x), \quad \overline{f}_{\mu\nu}(x) \in [-\pi, \pi], \tag{8.3}$$

where $n_{\mu\nu}(x)$ is a positive or negative integer. Let $\hat{e}(p;c)$ be the normal vector to plaquette p, pointing away from the center of cube c, and define $\Phi(p;c) = \frac{1}{2}\varepsilon_{ijk}e_i(p;c)\overline{f}_{jk}$. Then we obtain, from $\overline{f}_{\mu\nu}(x)$, the Gauss law for magnetic fields

$$\begin{aligned} 2\pi m &= \sum_{p\in c} \Phi(p;c) \\ &= \frac{1}{2}\varepsilon_{ijk}\partial_i\overline{f}_{jk}(x). \end{aligned} \tag{8.4}$$

It is understood that the location of the magnetic charge m is at the center of the cube. The centers of all cubes on a $D = 3$ dimensional lattice are the sites of the dual lattice (cf. Chap. 6), and the magnetic monopoles are located at the sites of that structure.

It is natural to wonder if it is possible to express the action for compact QED$_3$ in terms of the magnetic monopoles and their interactions, integrating out all other degrees of freedoms. For this purpose, it is useful to use an approximation known as the *Villain* or periodic Gaussian model, in which the partition function is

$$Z = \int_{-\pi}^{\pi} \prod_{x,\mu} d\theta_\mu(x) \sum_{n_{\alpha\beta}(x)=-\infty}^{\infty} \exp\left[-\frac{1}{4g^2 a}\sum_p (f_{\mu\nu}(x) - 2\pi n_{\mu\nu}(x))^2\right], \tag{8.5}$$

where a is the lattice spacing, and $g_L^2 = g^2 a$ is the dimensionless lattice coupling in three dimensions. Like the Wilson action, the action in the Villain model is periodic with respect to $f_{\mu\nu} \to f_{\mu\nu} \pm 2\pi$, and in fact the Villain model approaches the usual Wilson action formulation as charge $g^2 \to 0$. In this model the

integration over θ_μ can be carried out analytically, and after a number of steps that can be found in [4, 6, 7], one arrives at an action, for N monopoles, which involves only the monopole charges and their locations:

$$S_{\text{mon}} = \frac{2\pi^2}{g^2 a} \sum_{i \neq j} m_i m_j G(r_i - r_j) + \frac{2\pi^2}{g^2 a} G(0) \sum_i m_i^2, \qquad (8.6)$$

where the indices i, j run from 1 to N, and $G(r - r')$ is the lattice Coulomb propagator in $D = 3$ dimensions

$$-\nabla^2_{\text{lattice}} G(r - r') = \delta_{r,r'}, \qquad (8.7)$$

going as $1/4\pi|r - r'|$ at large separations. This expression was first derived by Polyakov [4, 6, 7], and it represents the action of a gas of monopoles in three dimensions interacting via Coulombic forces; i.e., a monopole plasma. A Wilson loop, in this formulation, can be thought of as a current loop interacting with the magnetic field due to the monopole background. The loop expectation value can be calculated explicitly in $D = 3$ dimensions, and the result is that for a Wilson loop associated with n units of electric charge [4, 6, 8, 9]

$$\langle U_n(C) \rangle \approx \exp[-n\sigma \, \text{area}(C)], \qquad (8.8)$$

where string tension σ is a calculable function of coupling β, and the loop holonomy $U_n(C) = (U_1(C))^n$ in compact QED$_3$.

A very rough image of what's going on is this: A Wilson loop can be thought of as a current loop which itself is the source of a magnetic field. In a monopole plasma, the free magnetic charges, i.e., the monopoles and antimonopoles, will tend to line up along the minimal area of the loop, and screen out the magnetic field that would have been generated by the Wilson (current) loop source (cf. Fig. 8.2). The area law falloff for Wilson loops is associated with the screening sheet of monopoles and antimonopoles along the minimal loop area.

In $D = 3$ dimensions, monopoles and antimonopoles form a plasma at any lattice coupling, rather than being bound in monopole-antimonopole pairs. This is because of the fact that, in three dimensions, the entropy of monopole location wins over the cost in monopole action. In $D = 4$ dimensions, monopoles trace out

Fig. 8.2 Monopoles and antimonopoles in the plasma line up along the minimal area of the loop (*solid line*), to screen the magnetic field generated by the loop
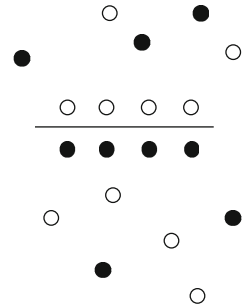
worldlines, and the creation and eventual annihilation of a monopole-antimono-pole pair is associated with a closed worldline loop. At weak couplings, the cost in monopole action of a long loop exceeds the entropy associated with the size and shape of large loops. This means that there are only small loops, and in a timeslice one would find that every monopole was close to an antimonopole, with a sepa-ration on the order of the typical extension of the small loops. In that case there is no monopole plasma and no confinement. But as the coupling increases and the monople action $\sim 1/g^2$ decreases, there is a point where the entropy wins, and monopole loops percolate through the volume. Now if we take a timeslice of the loop, the monopoles and antimonopoles can be very far apart. The system is then in a plasma phase, and confinement of electric charges is the result.

Polyakov's demonstration of confinement in compact QED$_3$ is quite beautiful and it has been very influential, so the calculation is worth displaying (in the $n = 1$ single charge case) in a little more detail. First of all, at weak couplings monopoles are very heavy, with action proportional to the square of magnetic charge, so monopoles with charges $|m_i| > 1$ are suppressed. Then, allowing for the fact that the number N of monopoles + antimonopoles is variable, the partition function for a monopole Coulomb gas is

$$Z_{\text{mon}} = \sum_{N=0}^{\infty} \frac{\xi^N}{N!} \sum_{\{r_n\}} \sum_{\{m_n=\pm 1\}} \exp\left[-\frac{2\pi^2}{g^2 a} \sum_{i \neq j} m_i m_j G(r_i - r_j)\right], \tag{8.9}$$

where

$$\xi = \exp\left[-\frac{2\pi^2}{g^2 a} G(0)\right] \tag{8.10}$$

and $G(0) \approx 0.253$ in lattice units [10]. Introducing a field $\chi(r)$ to localize the action,

$$\begin{aligned} Z_{\text{mon}} &= \int \prod_{r'} d\chi(r') \exp\left[-\frac{g^2 a}{4\pi^2} \sum_r \frac{1}{2}(\partial_\mu \chi(r))^2\right] \sum_{N=0}^{\infty} \frac{\xi^N}{N!} \sum_{\{r_n\}} \sum_{\{m_n=\pm 1\}} e^{i\sum_n m_n \chi(r_n)} \\ &= \int \prod_{r'} d\chi(r') \exp\left[-\frac{g^2 a}{4\pi^2} \sum_r \frac{1}{2}(\partial_\mu \chi(r))^2\right] \sum_{N=0}^{\infty} \frac{\xi^N}{N!} \left[\sum_r 2\cos(\chi(r))\right]^N \\ &= \int \prod_r d\chi(r) \exp\left[-\frac{g^2 a}{4\pi^2} \sum_r \left(\frac{1}{2}(\partial_\mu \chi)^2 - M_0^2 \cos \chi(r)\right)\right], \end{aligned} \tag{8.11}$$

where

$$M_0^2 = \frac{8\pi^2}{g^2 a} \exp\left[-\frac{2\pi^2}{g^2 a} G(0)\right]. \tag{8.12}$$

To make the expressions a little more compact, we switch to continuum notation. The translation is

$$(\partial_\mu)_{\text{lattice}} \rightarrow a(\partial_\mu)_{\text{continuum}}, \quad a^3 \sum_r \rightarrow \int d^3r, \tag{8.13}$$

so that

$$Z_{\text{mon}} \approx \int D\chi(r) \exp\left[-\frac{g^2}{4\pi^2} \int d^3r \left(\frac{1}{2}(\partial_\mu\chi)^2 - M^2 \cos\chi(r)\right)\right], \tag{8.14}$$

where

$$M^2 = a^{-2}M_0^2. \tag{8.15}$$

In the Coulomb gas approximation, the magnetic field due to the monopoles at sites $\{r_n\}$ can be obtained from the monopole density $\rho(r)$

$$\partial_\mu H_\mu(r) = 2\pi\rho(r), \quad \text{i.e.,} \quad H_\mu(r) = \frac{1}{2}\int d^3r' \frac{(r-r')_\mu}{|r-r'|^3}\rho(r'), \tag{8.16}$$

where

$$\rho(r) = \sum_{i=1}^N m_i\delta(r-r_i) \tag{8.17}$$

In particular we find, if C denotes a closed curve and $S(C)$ a surface with C as boundary, that

$$\oint_C dr_\mu A_\mu(r) = \int_{S(C)} dS_\mu(r)H_\mu(r) = \int d^3r \eta_{S(C)}(r)\rho(r), \tag{8.18}$$

where

$$\eta_{S(C)}(r) = -\frac{1}{2}\frac{\partial}{\partial r_\mu}\int_{S(C)} dS_\mu(r')\frac{1}{|r-r'|}. \tag{8.19}$$

If C is a planar curve in the (x, y) plane and $S(C)$ the planar surface with C as boundary curve, then we have

$$\partial^2\eta_{S(C)} = 2\pi\delta'(z)\theta_{S(C)}(x, y), \tag{8.20}$$

where $\theta_{S(C)}(x, y) = 1$ for a point on $S(C)$ inside the boundary C, and $\theta_{S(C)}(x, y) = 0$ for a point outside the boundary C. For points close to the surface S

$$\eta_{S(C)}(r) = \pi \, \text{sign}(z)\theta_{S(C)}(x, y), \tag{8.21}$$

i.e., this quantity jumps by 2π across the surface.

A planar Wilson loop can now be expressed in terms of the monopole charge density, using (8.18), as follows[2]:

$$\langle U_1(C)\rangle \equiv \langle e^{i\oint dr_\mu A_\mu(r)}\rangle = \langle e^{i\int d^3 r \eta_{S(C)}(r)\rho(r)}\rangle. \tag{8.22}$$

Repeating the steps which lead from (8.9) to (8.14), inclusion of the Wilson loop (8.22) leads to a shift of $\chi(r)$ by $\eta_{S(C)}$, i.e.,

$$\langle U_1(C)\rangle = \frac{1}{Z_{\text{mon}}}\int D\chi(r)\exp\left[-\frac{g^2}{4\pi}\int d^3 r\left(\frac{1}{2}(\partial_\mu(\chi - \eta_{S(C)}))^2 - M^2\cos\chi(r)\right)\right]. \tag{8.23}$$

The dominant contribution to the expectation value (8.23) comes from the classical solution to the effective action in (8.23); the saddlepoint approximation is to estimate the functional integral by evaluating the exponent at the classical value. Choosing the planar loop to lie in the (x, y) plane we obtain from (8.20):

$$\partial^2\chi = 2\pi\delta'(z)\theta_{S(C)}(x, y) + M^2\sin\chi. \tag{8.24}$$

The solution of this equation is discontinuous across the minimal surface of the loop, and at large M falls rapidly to zero away from the surface. Substituting this solution into the effective action results in an action density peaked in the neighborhood of the minimal area, and the increase in the action due to the Wilson loop source is proportional to the minimal area of the loop. In this way Polyakov obtained an the area law falloff for large Wilson loops

$$\langle U_1(C)\rangle \approx \exp[-\sigma\text{area}(C)], \tag{8.25}$$

from the three dimensional Coulomb gas of monopoles characteristic of compact QED in three Euclidean dimensions.[3]

8.2 The Georgi–Glashow Model in $D = 3$ Dimensions

Confinement by monopoles in three-dimensional electrodynamics is a beautiful mechanism, but is it also relevant to non-abelian theories? The first place to look is in non-abelian theories, such as the Georgi–Glashow model, which have magnetic monopoles as solitonic solutions of the equations of motion.

The Georgi–Glashow model, also known as the $SU(2)$ adjoint Higgs model, is an $SU(2)$ gauge theory with a Higgs field in the adjoint representation. In $D = 3$ dimensions the action is

---

[2] A factor of electric charge g has been absorbed into the definition of A_μ.

[3] A rigorous proof of confinement in compact QED$_3$ was later published in [11].

$$S = \int d^4x \left[\frac{1}{2} Tr[F_{\mu\nu}^2] + \frac{1}{2}(D_\mu \phi^a)^2 + \frac{1}{4} \lambda (\phi^a \phi^a - v^2)^2 \right], \qquad (8.26)$$

where ϕ^a has three color components, and D_μ is the usual covariant derivative. The potential is obviously minimized at $|\phi| = \sqrt{\phi^a \phi^a} = v$. Let us define

$$\widehat{\phi}^a = \frac{\phi^a}{|\phi|}. \qquad (8.27)$$

If we transform to a unitary gauge, where $\widehat{\phi} = \delta^{a3}$, then there is still a residual $U(1)$ local gauge invariance, which leaves $\widehat{\phi}$ invariant. It is tempting, in this gauge, to regard the model as a gauge theory of "photons," i.e., the gauge potentials A_μ^3, which are neutral with respect to the $U(1)$ subgroup, coupled to electrically charged "W-bosons" $W_\mu^{\pm} = (A_\mu^1 \mp i A_\mu^2)/\sqrt{2}$. The corresponding abelian field strength due to the photons would be $f_{\mu\nu}^3 = \partial_\mu A_\nu^3 - \partial_\nu A_\mu^3$. 't Hooft [12] found a gauge invariant field strength

$$\begin{aligned}
\mathscr{F}_{\mu\nu} &= \partial_\mu(\widehat{\phi}^a A_\nu^a) - \partial_\nu(\widehat{\phi}^a A_\mu^a) - \frac{1}{g} \varepsilon^{abc} \widehat{\phi}^a \partial_\mu \widehat{\phi}^b \partial_\nu \widehat{\phi}^c \\
&= F_{\mu\nu}^a \widehat{\phi}^a - \frac{1}{g} \varepsilon^{abc} \widehat{\phi}^a D_\mu \widehat{\phi}^b D_\nu \widehat{\phi}^c,
\end{aligned} \qquad (8.28)$$

which obviously reduces to the abelian field strength in unitary gauge. This tensor is used to define the electromagnetic field strength gauge-invariantly in the Georgi–Glashow model.

't Hooft [12] and Polyakov [13] found that the equations of motion of the Georgi–Glashow model contain magnetic monopole solutions which are instantons, in $D = 3$ dimensions, or static solutions, in $D = 3 + 1$ dimensions. A monopole centered at the origin has, in a convenient gauge, the "hedgehog" form

$$\begin{aligned}
\phi^a(x) &= v \frac{x^a}{r} f(r), \\
A_i^a(x) &= \varepsilon_{aij} \frac{x^j}{gr^2} a(r),
\end{aligned} \qquad (8.29)$$

where $f(r)$, $a(r)$ satisfy certain differential equations, with $f(0) = a(0) = 0$, and $f(\infty) = a(\infty) = 1$. This solution can be identified as a magnetic monopole because the abelian magnetic field, derived from the electromagnetic field strength $\mathscr{F}_{ij}$, has the monopole form at large r

$$B_i = \frac{1}{2} \varepsilon_{ijk} \mathscr{F}_{jk} \rightarrow -\frac{1}{g} \frac{x^i}{r^3}, \qquad (8.30)$$

so that the surface integral of the B-field, evaluated on a large closed surface (e.g. a sphere at large r) enclosing the origin will be

$$\int d\mathbf{S} \cdot \mathbf{B} = -\frac{4\pi}{g}. \tag{8.31}$$

Since the flux integrated over a closed surface is equal (Heaviside–Lorentz convention) to the charge contained inside, we conclude that the magnetic charge of the magnetic monopole is $-4\pi/g$. The mass of a magnetic monopole, in $D = 3 + 1$ dimensions, is

$$M_{\text{mon}} = \frac{4\pi\text{v}}{g} h\left(\frac{\lambda}{g^2}\right), \tag{8.32}$$

where $h(x)$ is a slowly varying function of order one.

In a unitary gauge, where ϕ^a is rotated to point in the positive color 3 direction, the A_μ^3 field is the photon, and the charged W bosons are the $W_\mu^\pm = (A_\mu^1 \mp iA_\mu^2)/\sqrt{2}$ fields. In this gauge the 't Hooft–Polyakov monopole lies in the color 3 direction, and has the form of a Dirac monopole at large r. In other words, if we consider a surface enclosing the monopole, then there is a physically undetectable Dirac string, carrying a magnetic flux of $4\pi/g$ in the color 3 direction, which enters the surface, and an equal magnetic flux, diverging from the center of the monopole, which leaves the surface. As in compact QED$_3$, the location of the Dirac string can be varied by a gauge transformation, although its endpoint on the monopole is fixed. A point to note is that the Higgs field vanishes at the center of the monopole in any gauge; this will be important when we come to the abelian projection proposal for identifying monopoles in pure gauge theories.

't Hooft–Polyakov monopoles again interact via a Coulombic potential, and the partition function for the corresponding monopole Coulomb gas can be brought to the form of Eq. 8.14, except that now the parameter M, which diverges in compact QED$_3$ in the lattice spacing $a \to 0$ limit, is replaced by a finite constant, dependent on the coupling g and W-boson mass. Wilson loops can be computed exactly as in compact QED$_3$, and a finite string tension $\sigma \sim \exp[-S_{\text{mon}}]$ is obtained [7]. It would *seem* that in the Georgi–Glashow model, the goal of separating the gauge field into a piece $\mathscr{A}$, which carries only the confining fluctuations, and the remaining non-confining fluctuations $\widetilde{A} = A - \mathscr{A}$ has been accomplished unambiguously: The instruction is to go to unitary gauge, identify the monopole positions $\{x_k^\mu\}$ from the zeros of the Higgs fields, and let $\mathscr{A}_\mu^a(x) = \delta^{a3}\mathscr{A}_\mu$ be the vector potential of Dirac monopoles at those positions. The corresponding long-range magnetic field at point x, due to monopoles and antimonopoles centered at points $\{x_k\}$ is then

$$B_\mu^3(x) = \sum_k q_k \frac{(x - x_k)_\mu}{|\mathbf{x} - \mathbf{x_k}|^3}, \tag{8.33}$$

where $q_k = \pm 1/g$. This identification of the confining fluctuations is certainly correct for compact QED$_3$. It is a little problematic, however, in the Georgi–Glashow model.

The problem is connected with loops of multiple electric charge. Consider, in unitary gauge, loops of the form

$$U_n(C) = \exp\left[in\frac{1}{2}\oint_C A_\mu^3(x)dx^\mu\right].\tag{8.34}$$

If the confining fields are those generated by a monopole Coulomb gas, then the charge-dependence of the string tension is the same as for compact QED$_3$, namely

$$\langle U_n(C)\rangle \sim e^{-n\sigma A(C)}.\tag{8.35}$$

with $A(C)$ the minimal area. But this answer cannot be correct in the Georgi–Glashow model, as pointed out in ref. [9], because there are charged $W^\pm$ bosons carrying two units of the minimum electric charge. This means that static charges carrying n units of electric charge are screened to n mod 2 units of charge, and therefore

$$\langle U_n(C)\rangle \sim \begin{cases} e^{-\sigma A(C)} & n \text{ odd} \\ e^{-\mu P(C)} & n \text{ even} \end{cases}.\tag{8.36}$$

The vector potential obtained from the monopole Coulomb gas action leads to the wrong answer for these observables, and from this it follows that confining magnetic flux derived from $\mathscr{A}_\mu^a(x)$ is not distributed in accordance with a monopole plasma, at least, not at very large distance scales.

What we learn from this is that the infrared effects of W bosons cannot be neglected in constructing an effective long-range action for the Georgi–Glashow theory. The usual justification for ignoring the W bosons is that these are massive objects, and therefore cannot affect the dynamics at large scales. For a confining theory, that argument is simply wrong. Moreover, confining fluctuations must organize themselves at large scales in such a way that $n =$ even charged loops have a vanishing string tension. The most obvious way to satisfy this condition is for the confining flux to be collimated into Z_2 vortices, rather than being distributed as in a monopole Coulomb gas [9].

On the other hand, the monopole Coulomb gas approximation ought to have *some* range of validity in the D $= 3$ Georgi–Glashow model. On the basis of simple energetics arguments, the electric flux tube between static sources carrying two units of U(1) electric charge will break by W pair creation only when the energy of the flux tube is greater than the combined mass of the two W bosons; i.e., when $L\sigma_2 > 2\,m_W$. The double-charge string tension σ_2 is zero beyond the string-breaking distance; however, up to that point, the monopole Coulomb gas result that $\sigma_2 = 2\sigma_1$ should hold. Then the string-breaking distance is $L \approx m_W/\sigma_1$, and up to that length scale a monopole Coulomb gas analysis ought to give a good account of the dynamics.

What about D $= 4$ dimensions? In four dimensions the Georgi–Glashow model has a confining phase and a non-confining phase [14–16], and only in the

non-confining phase, where it is possible to perturb (in, e.g., unitary gauge) around a non-zero value of the scalar field, are there stable magnetic monopole solutions. Monopoles and antimonopoles trace out worldlines in four dimensions, and monopole antimonopole pairs trace out closed loops. In the non-confining phase these loops are small and do not percolate (in the Euclidean theory) throughout the spacetime volume. In a time slice, one would find that every monopole was paired with an antimonopole. The monopoles do not form a plasma, and it appears that the Polyakov calculation is not applicable.

A possible way to extend the $D = 3$ physics a little way into $D = 4$ dimensions is to consider a pure gauge theory in a four-dimensional space in which one of the space dimensions is compactified, i.e., we apply periodic boundary conditions in one of the dimensions, so space is not R^4 but $R^3 \times S^1$. Polyakov lines through the compact direction then play a role somewhat analogous to the Higgs field in the Georgi–Glashow model. However, the expectation value of the trace of the Polyakov line is non-zero if the radius of S^1 is small, so from the 4D viewpoint the compactified model is in the deconfined phase, and a phase transition separates the large and small radius regimes. It has been pointed out by Shifman and Ünsal [17] that this deconfinement transition could be avoided by adding to the action of an $SU(N)$ gauge theory the following term

$$\Delta S = \int\limits_{R^3} d^3x \frac{2}{\pi L^3} \sum_{n=1}^{[N/2]} d_n \left| Tr U^n(\mathbf{x}) \right|^2, \qquad (8.37)$$

where L is the extension of the compact direction, $U(\mathbf{x})$ is a Polyakov loop holonomy winding around the compact direction, $\mathbf{x}$ denotes coordinates in R^3, and $[N/2]$ is the integer part of quantity in brackets. For small L and judiciously chosen values d_n, the eigenvalues of the Polyakov line $U(\mathbf{x})$ are driven as far apart as possible, so that the action is minimized at $P(\mathbf{x}) = Tr U(\mathbf{x}) = 0$, and center symmetry is unbroken. It is then possible to demonstrate confinement via a monopole mechanism, along the same lines as the $D = 3$ Georgi–Glashow model. At large L, of course, center symmetry arises in a rather different fashion; it is due to large and random fluctuations in the eigenvalues of the Polyakov line. At large L there is no obvious abelian confinement mechanism, and the analytical control available at small L is lost. Nevertheless, the Shifman-Ünsal approach does demonstrate confinement in non-supersymmetric theories in $D = 4$ dimensions for small L, and L can be regarded as a parameter which may interpolate continuously between an abelian monopole mechanism, and (since ΔS is negligible at large L) whatever mechanism is operative in R^4. The approach has been extended to theories with fermions in various representations. In fact, with massless fermions in the adjoint representation, ΔS can be dispensed with, because a similar effect is found at the level of the effective action, obtained after integrating out the massive Kaluza–Klein excitations that arise due to the compact direction [18].

8.3 Dual Superconductivity, and the Seiberg–Witten Model

When a type II superconductor is placed in an external magnetic field, the magnetic field can only penetrate the superconductor in cylindrical regions known as Abrikosov vortices, which carry only a certain fixed amount of magnetic flux. Abrikosov vortices are magnetic flux tubes, and in principle such a flux tube could begin at a monopole, of appropriate magnetic charge, and end on an antimonopole. The constant energy density along an Abrikosov vortex then implies a linear potential between the monopole and antimonopole. In other words, a type II superconductor, which is a system in which bosons (the Cooper pairs) of electric charge $2e$ are condensed, is a system which squeezes magnetic fields into tubes of quantized flux, and confines magnetic charge via a linear potential. So if there *were* real magnetic monopoles and antimonopoles present in a material of this kind, we would observe that these objects were confined.

The example of type II superconductors becomes even more suggestive in view of the electric/magnetic *duality* symmetry of Maxwell's equations. If we are willing to allow for the existence of elementary magnetic charges g_M, then the corresponding Maxwell's equations

$$\partial^\mu F_{\mu\nu} = j_\nu^e, \quad \partial^{\mu*} F_{\mu\nu} = j_\nu^m \tag{8.38}$$

are symmetric with respect to the interchange of fields $F_{\mu\nu} \to {}^*F_{\mu\nu}$, ${}^*F_{\mu\nu} \to -F_{\mu\nu}$, and currents, i.e.,

$$\mathbf{E} \to \mathbf{B}, \quad \mathbf{B} \to -\mathbf{E}, \quad j_\mu^e \leftrightarrow j_\mu^m, \quad j_\mu^m \leftrightarrow -j_\mu^e, \tag{8.39}$$

where ${}^*F_{\mu\nu}$ is the dual field strength tensor

$$^*F_{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}. \tag{8.40}$$

Conservation of both the electric and magnetic currents implies separate electric and magnetic $U(1)$ symmetries. It occurred independently to 't Hooft [2] and Mandelstam [3], in 1976, that quark confinement in QCD might be due to some type of *dual superconductivity*, in which the QCD vacuum state can be regarded as a condensate of a magnetically charged boson field (magnetic monopoles), confining electrically charged particles (the quarks). In other words, a dual superconductor is like an ordinary superconductor in which the roles of the $\mathbf{E}$ and $\mathbf{B}$ fields, and electric and magnetic charges, have been interchanged. This approach entails, as in the Georgi–Glashow model, singling out an abelian subgroup of the full gauge group.

The idea is expressed more concretely in a relativistic generalization of the Landau-Ginzburg theory of superconductivity. The relativistic action is the *abelian Higgs model*

$$S = \int d^D x \left(\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \left| (\partial_\mu + ieA_\mu)\varphi \right|^2 + \frac{\lambda}{4} (\varphi^* \varphi - v^2)^2 \right). \tag{8.41}$$

The abelian Higgs model has magnetic flux-tube solutions analogous to Abrikosov vortices, which are known as *Nielsen–Olesen vortices* [19]. For a vortex along the z-axis, the Nielsen–Olesen vortex in cylindrical (r, ϕ, z) coordinates has the following asymptotic form:

$$\begin{aligned} \varphi(r, \phi) &= v f(r) e^{in\phi} \\ A_\phi(r) &= \frac{n}{er} a(r), \end{aligned} \tag{8.42}$$

with $f(0) = a(0) = 0$ and $f(\infty) = a(\infty) = 1$. The topological stability of these vortices is related to the fact that the Higgs potential has its minimum away from $\varphi = 0$, and that the complex phase of the Higgs field along a closed loop around the vortex has a non-zero integer winding number n. The magnetic flux carried by a Nielsen–Olesen vortex is $2\pi n/e$. In the dual version of the abelian Higgs model, the Higgs field is magnetically charged, and it couples, not to the usual vector potential A_μ, but rather to a "dual" photon field C_μ whose associated field strength is $^*F_{\mu\nu}$. Then the Nielsen–Olesen vortex is an electric, rather than magnetic flux tube, and it is electric charge which is confined.

The dual-superconductor idea is not only very attractive, but it can also be demonstrated to be operating in a particular theory: the Seiberg–Witten model. The important work of Seiberg and Witten on duality and confinement in supersymmetric gauge theories is beyond the scope of this book, and the reader is referred to the original articles [20, 21] and reviews [22–24] for an exposition. A few comments, however, are relevant at this point.

The Seiberg–Witten work is concerned with supersymmetric Yang–Mills theory. Supersymmetry is an extension of Poincare symmetry, adding to the usual group generators (translations, rotations, and boosts) of the Poincare Lie algebra a set of $\mathcal{N}$ supersymmetry generators. These enlarge the Lie algebra to include certain anticommutator relations among the supersymmetry generators. The number $\mathcal{N}$ of different supersymmetry generators is, in general, a power (≥ 0) of 2. Irreducible representations of supersymmetry contain both bosonic and fermionic particle states, with equal numbers of bosonic and fermionic degrees of freedom in the supermultiplet. The (bosonic + fermionic) field content of a supersymmetric field theory, as well as the interaction terms in such theories, are highly constrained.

$\mathcal{N} = 2$ super Yang–Mills theory, like the Georgi–Glashow model, has scalar fields in the adjoint representation of the gauge group, which can be used to single out a compact abelian subgroup, which is $U(1)^{N-1}$ for the $SU(N)$ gauge group. The theory does not have a unique vacuum; rather there is a continuous set of inequivalent vacua, which are specified by the values ("moduli") of certain scalar field operators. Seiberg and Witten were able to show that there is a point in

moduli space at which the magnetic monopoles of this theory become massless. On adding a soft supersymmetry breaking term, which reduces the $\mathcal{N} = 2$ supersymmetry to $\mathcal{N} = 1$, the theory goes into a confining phase associated with a condensation of the monopole field. Seiberg and Witten derive in this case an effective low energy action which is a supersymmetric generalization of the dual abelian Higgs model. This is a truly remarkable result: the dual abelian Higgs model is expressed in terms of dual (magnetically charged) fields, even though such fields were invisible in the Lagrangian of the starting point, which is $\mathcal{N} = 2$ super Yang–Mills theory. Once a dual abelian Higgs model has been obtained, electric flux tube formation and confinement of electric charge follows, along the same lines as magnetic flux tube confinement of magnetic charge in the ordinary abelian Higgs model, and in type II superconductors. The Sine Law rule for k-string tensions, already mentioned in Chap. 5, was motivated by the $SU(N)$ version of the Seiberg–Witten model, where the Sine Law was derived by Douglas and Shenker [25]. These authors also pointed out that the $SU(N)$ gauge group leads to $N - 1$ separate copies of the dual abelian Higgs model, and thus $N - 1$ distinct types of electric flux tube, which is not what one expects in a pure $SU(N)$ gauge theory.

The validity of the effective low energy theory at very large distance scales is questionable, however. A flux tube will also form between objects which carry two units of the $U(1)$ electric charge, and there is nothing in the effective dual abelian Higgs model which would allow that flux tube to break. On the other hand, in the $SU(2)$ theory these electric charges have zero N-ality, and the flux tube *must* break at some point, in this case via pair creation of W-bosons. Just as in Polyakov's monopole gas treatment of the $D = 3$ dimensional Georgi–Glashow model, the Seiberg–Witten low-energy effective action neglects the W bosons of the theory, on the grounds that they are massive and should not contribute to long-range physics. But we have already seen that the W bosons are very relevant to long-distance physics; the N-ality dependence of string tensions σ_r cannot be obtained without them. This means that the Seiberg–Witten effective action cannot be the whole story at sufficiently large distance scales. To put it another way: the "low-energy" effective action is not necessarily the long-distance effective action. In the Seiberg–Witten theory, the effective action is obtained keeping only local terms with no more than two derivatives of the bosonic fields. It is not obtained by actually integrating out the massive W bosons, for otherwise the effective action would have the correct N-ality dependence built in. At distance scales on the order of the color screening length, $L \approx m_W/\sigma$, the dual abelian Higgs action arrives at the wrong N-ality dependence for abelian Wilson loops, and on those grounds it is inadequate to describe the large-scale vacuum fluctuations in the theory. There is, as in the Georgi–Glashow model, an intermediate range of distances up to the screening scale for which the Seiberg–Witten effective action gives a good account of the physics. Beyond that scale the contributions of virtual W bosons are important, and can no longer be ignored.

There have been interesting efforts in recent years to construct supersymmetric models in which the confining dual Higgs theory, and the confining flux tube, are essentially non-abelian in character; cf. [26–28] and references therein for reviews. For this purpose it seems crucial to add some number of matter "hypermultiplets," containing both Fermi and Bose fields in the fundamental representation of the gauge field (which would imply string breaking, as in real QCD). The first models of this kind were found in $\mathcal{N} = 2$ super Yang–Mills theory with a gauge group $SU(N + 1)$ broken to $U(1) \times SU(N)$, and $N = N_f$ equal-mass matter hypermultiplets containing complex scalar fields ("squarks") and Weyl fermions in the fundamental representation. Squark condensates "break" (in the sense of breaking a global subgroup, cf. Chap. 3) both the $SU_C(N)$ color and $SU_F(N)$ flavor groups, leaving an unbroken diagonal $SU(N)_{C+F}$ global subgroup. This pattern of symmetry breaking is known as "color–flavor locking." The model has magnetic flux tube solutions in which the chromomagnetic flux is free to rotate in the $SU(N)$ color directions, and it is on these grounds that the flux tube solutions are referred to as "non-abelian" vortices. Many other dual models with non-abelian vortices have been constructed, for which the reader is referred to the cited references. It is not clear that these constructions, which rely on a special choice of fundamental representation matter fields and pattern of symmetry breaking, would be relevant to the flux tubes encountered in simpler confining theories, such as pure gauge theories with no fundamental matter fields, and having no string breaking by such fields.

8.4 The Abelian Projection

The example of monopole confinement in compact QED, in the Georgi–Glashow model, and in the Seiberg–Witten model, raises the hope that some similar mechanism might exist in pure gauge theories, or gauge theories without fundamental matter fields. To make this idea work, one has to identify an abelian subgroup of the gauge group, associated with the abelian magnetic monopole field. Since there is no elementary Higgs field in the adjoint representation, whose unitary gauge would single out an appropriate abelian subgroup, 't Hooft [29] suggested that a composite operator involving only gluonic fields, and behaving under a gauge transformation $G(x)$ like a matter field in the adjoint representation, might serve the same purpose. For a matrix-valued operator $X(x)$, the gauge transformation law would be

$$X(x) \rightarrow G(x)X(x)G^{\dagger}(x). \tag{8.43}$$

An example of such an operator would be any spatial component of the non-abelian field strength tensor, e.g. $F_{12}(x)$. The analog of a unitary gauge would be the gauge in which the operator X is diagonal. In an $SU(N)$ theory X is an $N \times N$ matrix, and in the gauge where

$$X = diag[\lambda_1, \lambda_2, \ldots, \lambda_N] \quad \text{with} \quad \lambda_1 < \lambda_2 < \cdots < \lambda_N \qquad (8.44)$$

there is left unfixed an abelian subgroup $U(1)^{N-1}$, whose generators belong to the Cartan subalgebra of the gauge group. The Cartan subalgebra is the largest subset of group generators which commute with one another, but the choice is not unique. For example, the generators of $SU(2)$ are the three Pauli matrices ($\times \frac{1}{2}$). The $U(1)$ subgroup generated by any one of the Pauli matrices can be taken as the Cartan subgroup. For $SU(3)$, one could choose, e.g., the third component of isospin I_3, and hypercharge Y, forming the subgroup $U(1) \times U(1)$. In general, for any $SU(N)$ group, the Cartan subgroup is $U(1)^{N-1}$.

In a gauge in which the operator X is diagonal, the gauge theory can be thought of as an abelian $U(1)^{N-1}$ gauge theory, with "photon" and monopole excitations, the photons corresponding to the generators of the Cartan subalgebra. In addition there are "matter" fields (the remaining gluons), which are electrically charged with respect to the Cartan subgroup. The idea is that confinement is due to monopole condensation, as in the dual abelian Higgs model. The locations of the monopoles are taken to be at those sites where the unitary gauge fixing is ambiguous, and this is where two of the eigenvalues of the X operator coincide (recall that the Higgs field, which is used to define the unitary gauge, vanishes at the center of a 't Hooft–Polyakov monopole). Of course, the locations of such monopoles are entirely dependent on the choice of operator X; it is not at all obvious that there is any real physical content associated with points or worldlines located in this way. Indeed, the gauge based on the operator $X = F_{12}$ has not led to anything interesting.

Gauges which fix the gauge symmetry only up to the residual symmetry of the Cartan subgroup are known as *abelian projection gauges*, and they do not necessarily require that we know the operator X explicitly. In particular, since the goal is to identify an abelian confinement mechanism inside a non-abelian gauge theory, it is reasonable that the idea might work best, in an $SU(N)$ gauge theory, in a gauge where the theory looks most like a $U(1)^{N-1}$ gauge theory. This would be a gauge where the quantum fluctuations of "charged" gluon degrees of freedom are much suppressed, compared to the quantum fluctuations of the "photon" degrees of freedom. So we look for a gauge in which the link variables are as diagonal as possible. This is known as the *maximal abelian gauge* [30], and in $SU(2)$ gauge theory, it is specified by the condition that

$$R = \sum_{x,\mu} \text{Tr}[U_\mu(x)\sigma_3 U_\mu^\dagger(x)\sigma_3] \qquad (8.45)$$

is maximized.

Fixing to the maximal abelian gauge leaves a residual $U(1)$ gauge symmetry

$$U_\mu(x) \to e^{i\phi(x)\sigma_3} U_\mu(x) e^{-i\phi(x+\widehat{\mu})\sigma_3}. \qquad (8.46)$$

Let $u_\mu(x)$ be the diagonal part of $U_\mu(x)$, rescaled to restore unitarity

$$U_\mu(x) = a_0 \mathbb{1} + i\mathbf{a} \cdot \boldsymbol{\sigma}$$

$$u_\mu(x) = \frac{1}{\sqrt{a_0^2 + a_3^2}} \left[a_0 \mathbb{1} + i a_3 \sigma_3 \right]$$

$$= \begin{bmatrix} e^{i\theta_\mu(x)} & \\ & e^{-i\theta_\mu(x)} \end{bmatrix}.$$

(8.47)

Then, make the decomposition

$$U_\mu(x) = C_\mu u_\mu(x)$$

$$= \begin{bmatrix} \left(1 - |c_\mu(x)|^2\right)^{1/2} & c_\mu(x) \\ -c_\mu^*(x) & \left(1 - |c_\mu(x)|^2\right)^{1/2} \end{bmatrix} \begin{bmatrix} e^{i\theta_\mu(x)} & \\ & e^{-i\theta_\mu(x)} \end{bmatrix}.$$

(8.48)

What is interesting is that under the remnant $U(1)$ gauge symmetry, $\theta_\mu(x)$ transforms like an abelian gauge field, and $c_\mu(x)$ transforms like a matter field carrying two units of the abelian charge, i.e.,

$$e^{i\theta_\mu(x)} \rightarrow e^{i\phi(x)} e^{i\theta_\mu(x)} e^{-i\phi(x+\widehat{\mu})}$$

$$c_\mu(x) \rightarrow e^{2i\phi(x)} c_\mu(x),$$

(8.49)

so the particle content of the theory in maximal abelian gauge consists of "photons" coupled to double-charged matter fields. The *abelian projected lattice* is obtained by the replacement $U_\mu(x) \rightarrow \exp[i\theta_\mu(x)]$.

The $U(1)$ gauge symmetry is compact, and there are monopole currents $k_\mu(x)$ in the theory, which can be extracted from the abelian-projected links $\exp[i\theta_\mu(x)]$, using the generalization, to $D = 4$ dimensions, of (8.4)[4]

$$k_\mu(x) = \frac{1}{4\pi} \varepsilon_{\mu\alpha\beta\gamma} \partial_\alpha \overline{f}_{\beta\gamma}(x).$$

(8.50)

One can then construct link variables from the Coulombic fields of those monopoles alone, discarding the "photon" contributions, just from the Dirac string variables $n_{\mu\nu}$ [31]

$$u_\mu^{\text{mon}}(x) = \exp[i\theta_\mu^{\text{mon}}(x)]$$

$$\theta_\mu^{\text{mon}}(x) = -2\pi \sum_y G(x-y) \partial_\nu' n_{\mu\nu}(y),$$

(8.51)

---

[4] The monopole currents actually live on the links of the dual lattice (cf. Chap. 6) in four dimensions.

where $G(x - y)$ is again the lattice Coulomb propagator (inverse of the lattice Laplacian), and this time ∂'_μ denotes the backward lattice derivative. Replacing link variables by the $u_\mu^{\mathrm{mon}}(x)$, and calculating observables (such as Wilson loops) in that ensemble, is known as the *monopole dominance* approximation [32, 33].

One suggested application of abelian projection is to help define a monopole creation operator $\mu(\mathbf{x})$, whose expectation value could serve as an order parameter for dual superconductivity in a non-abelian gauge theory [34, 35]. Recall that in a compact $U(1)$ gauge theory the conserved magnetic current is associated with a dual $U(1)$ gauge symmetry, and the total magnetic charge generates a global subgroup of this dual gauge symmetry. If we define an operator $\mu(\mathbf{x})$ which acts on states in the Schrodinger representation by inserting a monopole field configuration $A_i^M(\mathbf{y})$ centered at the point $\mathbf{y} = \mathbf{x}$, i.e.,

$$\mu(\mathbf{x})|A_i\rangle = |A_i + A_i^M\rangle, \tag{8.52}$$

then μ does not commute with the total magnetic charge, and $\langle\mu\rangle \neq 0$ is a signal that the associated global symmetry is spontaneously broken. In a non-abelian gauge theory, an abelian projection gauge is introduced to single out an abelian $U(1)^{N-1}$ subgroup, and μ is defined in terms of the gauge fields associated with that abelian subgroup. If the expectation value of μ is non-zero, it means that the system is in the phase of "dual superconductivity." The proposal is that a gauge theory is in a confinement phase if and only if $\langle\mu\rangle > 0$, and a phase transition to $\langle\mu\rangle = 0$ is therefore a transition from the confined to the deconfined phase. In fact, in very many cases it has been shown numerically that the high-temperature deconfinement transition does coincide with $\langle\mu\rangle \to 0$. The problem with the proposal, however, is that the transition $\langle\mu\rangle \to 0$ is also found in cases where there is a bulk transition that is *not* associated with a transition to the deconfined phase, and there are even cases where the transition $\langle\mu\rangle \to 0$ occurs in the absence of any thermodynamic transition whatever [36]. This appears to be further evidence that the breaking of a remnant gauge symmetry is an unreliable guide to the presence or absence of a confined phase, as already argued in Chap. 3.

8.4.1 Monopoles and Vortices

The monopole dominance approximation has been investigated numerically [31–33] in SU(2) lattice gauge theory, and in this approximation the string tensions come out about right (i.e., almost equal to the tension of gauge invariant loops) for single charged ($n = 1$) Wilson loops. We write

$$\begin{aligned} W_n^{\mathrm{abel}}(C) &= \left\langle \exp[in\theta(C)] \right\rangle, \\ W_n^{\mathrm{mon}}(C) &= \left\langle \exp[in\theta^{\mathrm{mon}}(C)] \right\rangle. \end{aligned} \tag{8.53}$$

as the Wilson loops corresponding to n units of electric charge in the abelian projected lattice $U_\mu(x) \rightarrow \exp[i\theta_\mu(x)]$, and in the monopole dominance approximation $U_\mu(x) \rightarrow \exp[i\theta_\mu^{\text{mon}}(x)]$, respectively, where $\theta(C)$ is the lattice "loop integral" of link angles $\theta_\mu(x)$ around the loop C. Because of screening by the double-charged (off-diagonal) gluons, we should have

$$\sigma_n = \begin{cases} \sigma_1 & n \text{ odd} \\ 0 & n \text{ even} \end{cases}. \tag{8.54}$$

However, charge screening is difficult to check numerically in Wilson loop operators. Instead, the expectation values of double-charged Polyakov loops have been computed in both the abelian projection, and the monopole dominance approximation. Because of charge-screening, we expect

$$\langle P_n(\mathbf{x}) \rangle = \begin{cases} = 0 & n \text{ odd} \\ \neq 0 & n \text{ even} \end{cases}. \tag{8.55}$$

For single-charged $n = 1$ loops, below the deconfinement transition temperature, both the abelian projected and monopole dominance Polyakov loops have expectation values consistent with zero, as they should. However, the abelian projection and monopole dominance approximation give drastically different results for double-charged loops [37]. For the $n = 2$ loops, the expectation values of abelian-projection Polyakov loops are non-zero, as they should be. In contrast, the corresponding expectation values in the monopole dominance approximation are found to be consistent with zero, and this conflicts with the requirements of N-ality, and the existence of charge screening. It follows that, even if the confining disorder is dominated (in some gauge) by abelian configurations, the distribution of abelian flux *cannot* be that of a monopole Coulomb gas; such a distribution disorders both $n = 1$ and 2 loops. A similar argument applies to dual superconductors: if there is confinement for $n = 1$ electrically charged objects, then $n = 2$ charge is also confined, and this is incompatible with the existence of charge screening. In a type-II dual superconductor there would simply be two, rather than one, electric flux tubes running between doubly-charged sources.

On the other hand, the abelian projection does get some things right, such as the string tension of the fundamental representation Wilson loop on an abelian projected lattice [31, 38]. To see what may be going on, let's think of how vortices would look in maximal abelian gauge, at some fixed time t. Figure 8.3 is a sketch of a thick center vortex in the absence of gauge-fixing. In this case the vortex field strength $F_{\mu\nu}^a$ points in random directions in the Lie algebra. Fixing to maximal abelian gauge, the field tends to line up in the $\pm \sigma_3$ direction, but there will still be regions where the field strength rotates in the group manifold, from $+ \sigma_3$ to $- \sigma_3$ (Fig. 8.4). Now, if we keep only the abelian part of the link variables $u_\mu(x)$ (i.e., abelian projection), we get a monopole-antimonopole chain (Fig. 8.5), with $\pm \pi$ magnetic flux running between a monopole and neighboring antimonopole. The total monopole flux is $\pm 2\pi$, as it should be. Then a typical vacuum configuration at a fixed time, after abelian projection, looks something like the sketch in Fig. 8.6.
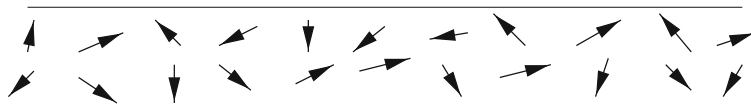
Fig. 8.3 Vortex field strength before gauge fixing. The *arrows* indicate direction in color space
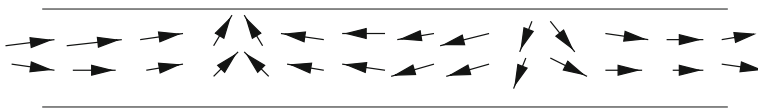

Fig. 8.4 Vortex field strength after maximal abelian gauge fixing. Vortex strength is mainly in the $\pm \, \sigma_3$ direction
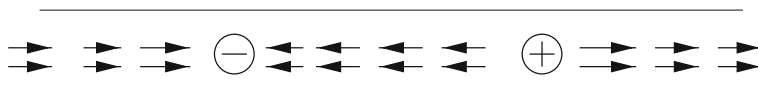
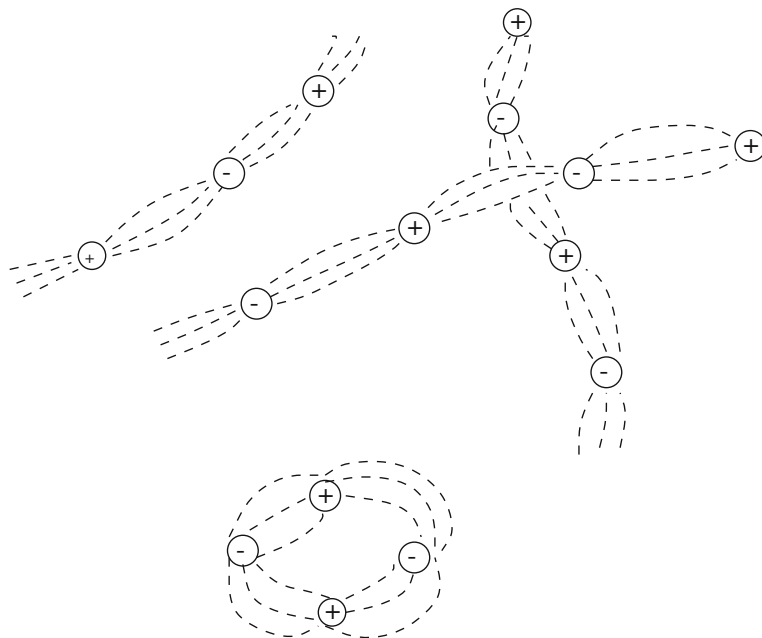

Fig. 8.5 Vortex field after abelian projection

Fig. 8.6 Hypothetical collimation of monopole/antimonopole flux into center vortex tubes on the abelian-projected lattice
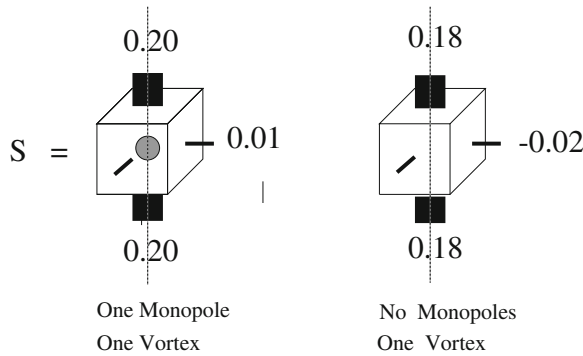
Fig. 8.7 The *left figure* shows the plaquette excess action distribution on an isolated monopole cube (no nearest neighbor monopole currents) pierced by a single P-vortex. In the illustration the bottom and top plaquettes are pierced by the vortex. For comparison, the excess action distribution is also shown for a no-monopole cube (*right figure*) pierced by a P-vortex. The simulation is at $\beta = 2.4$ in $SU(2)$ lattice gauge theory. From Ambjørn et al. [37]

Double-charged Wilson loops and Polyakov loops, in the abelian projection, are insensitive to linking with such vortices, and this will ensure that the string tensions σ_n will satisfy (8.54) as required.

The picture outlined above can be tested by numerical simulation. We work in the indirect maximal center gauge, which uses maximal abelian gauge as an intermediate step, and identify the locations of both monopoles (by abelian projection) and vortices (by center projection). We also measure the excess action (above the average plaquette value S_0), on plaquettes belonging to monopole cubes, and on plaquettes pierced by vortex lines. The following results are obtained for $SU(2)$ lattice gauge theory at $\beta = 2.4$ [37]:

1. Almost all monopoles and antimonopoles (97%) lie on vortex sheets.
2. At fixed time, the monopoles and antimonopoles alternate on the vortex lines, in a chain.
3. Excess (gauge-invariant) plaquette action is highly directional (cf. also Gubarev et al. [39]), and lies mainly on plaquettes pierced by vortex lines as indicated in Fig. 8.7. The presence or absence of a monopole in the cube is not so important to the excess action. Very similar results are obtained from 2^3 and 3^3 cubes.

8.5 Calorons

As we saw earlier in this chapter, Polyakov managed to derive confinement analytically in compact QED_3, by summing over classical solutions, in this case monopoles, of the Euclidean field equations. Monopoles are instantons in compact QED_3, and it is natural to wonder if confinement could be derived from some

semiclassical treatment of Yang–Mills theory based on the instanton solutions of non-abelian gauge theories. The standard instantons, introduced by Belavin et al. [40], do not seem to work; their field strengths fall off too rapidly to produce the desired magnetic disorder in the vacuum. In recent years, however, it has been realized that instanton solutions at finite temperature, known as *calorons*, might do the job. These caloron solutions were introduced independently by Kraan and van Baal [41, 42] and Lee and Lu [43] (KvBLL), and they have the remarkable property of containing monopole constituents which may, depending on the type of caloron, be widely separated.

The monopole constituents of the KvBLL calorons are also called *dyons* because they are a source of both electric and magnetic fields. For an $SU(N)$ gauge group there are N types of dyons, and in a suitable gauge, far from the dyon center, their field strengths are abelian (i.e., taking values in the Cartan subalgebra). Asymptotically, for the n-th dyon, we may write [44]

$$\pm E_k^{(n)} = B_k^{(n)} = \frac{1}{2}\frac{x_k}{r^3}C^{(n)}, \tag{8.56}$$

where the $C^n = \text{diag}[0,\ldots,1,-1,0,\ldots,0]$ (with 1 in the n-th position) for n from 1 to $N-1$, and $C^{(N)} = \text{diag}[-1,0,\ldots,1]$. The dyon solutions were discovered by Bogolmolny, Prasad, and Sommerfield [45, 46], and are known as BPS monopoles. A naive superposition of dyon fields is in general not a saddlepoint of the classical action; calorons, however, are not naive at all. A (topological) charge one caloron has N dyon constituents, and is indeed a solution of the Euclidean field equations.

A characteristic feature of the KvBLL caloron is the fact that the Polyakov loop holonomy (i.e., the Wilson line closed by periodicity in the time direction, before taking the trace) need not be a center element far away from the caloron. In general we can write the Polyakov loop holonomy as

$$\begin{aligned}\mathscr{P}(\mathbf{x}) &= P\exp\left[i\int dt A_0(\mathbf{x},t)\right]\\ &= S\text{diag}[e^{2\pi i\mu_1}, e^{2\pi i\mu_2},\ldots,e^{2\pi i\mu_N}]S^{-1},\end{aligned} \tag{8.57}$$

where S is some $SU(N)$ matrix which diagonalizes $\mathscr{P}$, and $\sum_n\mu_n = 0$. There are different caloron solutions, depending on the holonomy far from the caloron center. If $\mathscr{P}$ is a center element, then all the μ_n are equal; this is the "trivial" holonomy which characterizes the finite temperature instanton solutions found by Harrington and Shepard [47, 48]. Caloron solutions allow for non-trivial holonomy. If we order the $\{\mu_n\}$ such that $\mu_i \leq \mu_{i+1}$, then the maximally "non-trivial" holonomy is where the $\{\mu_n\}$ are equally spaced and as far apart as possible, i.e.,

$$\mu_n^{\text{max}} = -\frac{1}{2} - \frac{1}{2N} + \frac{n}{N}. \tag{8.58}$$

In this special case, $\text{Tr}[\mathscr{P}] = 0$. In a vacuum dominated by calorons of this special type, a Polyakov loop at a given location would tend to fluctuate around
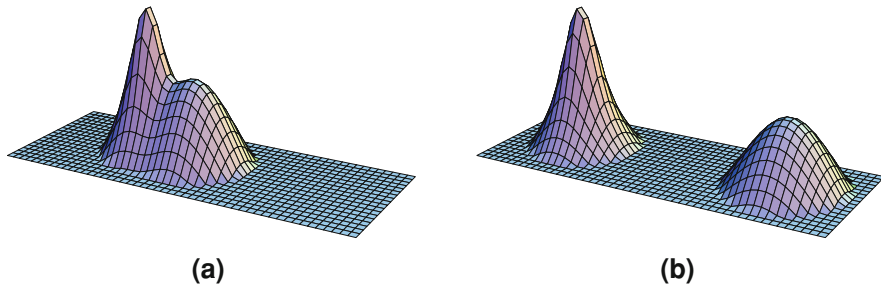
Fig. 8.8 Action density of an *SU*(2) caloron in a timeslice $t = 0$, with $\mu_2 = -\mu_1 = 0.125$. The parameter ρ determines the separation of the consituent dyons; the figure on the left is at $\rho = 0.8$, and the figure on the right has $\rho = 1.6$. From Kraan and van Baal [49]

$\text{Tr}[\mathscr{P}] = 0$, with the the probability density peaked at $\text{Tr}[\mathscr{P}] = 0$. This is in sharp contrast to the center vortex confinement scenario, where the probability density would be peaked at center elements, and in fact Polykov loop holonomies are *only* center elements in center-projected configurations. The vanishing of the Polyakov loop expectation value in the vortex scenario is due to an unbroken center symmetry, leading to an exact cancellation between positive and negative contributions, whereas in a purely caloron configuration, with maximally non-trivial holonomy, the vanishing of the Polyakov loop is a property of the configuration itself, rather than arising from quantum fluctuations among such configurations.

One of the features of the KvBLL caloron is that the positions of dyon constituents are set by the parameters ("moduli") of the caloron solution. The constituents can be widely separated; in fact they can be placed anywhere, depending on the choice of moduli. An example for the *SU*(2) group is shown in Fig. 8.8, which displays the action density of two *SU*(2) caloron solutions in a timeslice at $t = 0$. Because the caloron is an instanton solution, one might think that the dyon constituents appear and then disappear in a certain time interval. In fact, this is true for small dyon separations. When the dyons are widely separated, however, they persist throughout the entire periodic time interval, as seen in Fig. 8.9.

The fact that calorons are saddlepoints of the Euclidean Yang–Mills action, and can have both non-trivial holonomy and widely separated monopole constituents, has led to the conjecture that these constituents form a kind of dyon gas, which gives rise to the necessary magnetic disorder for confinement. Then the Yang–Mills functional integral might be approximated by summing over dyon configurations. To formulate this proposal precisely, one really needs the integration measure for the moduli of multi-dyon configurations, and this measure is unknown. Diakonov and Petrov, generalizing from the known measure of a single caloron, have proposed a measure for the many dyon case, and with this choice have been able to derive a number of interesting results [44]:

1. At finite but low temperatures, the free energy of the dyon gas is minimized at the maximally non-trivial Polyakov holonomy, where $\text{Tr}[\mathscr{P}] = 0$.
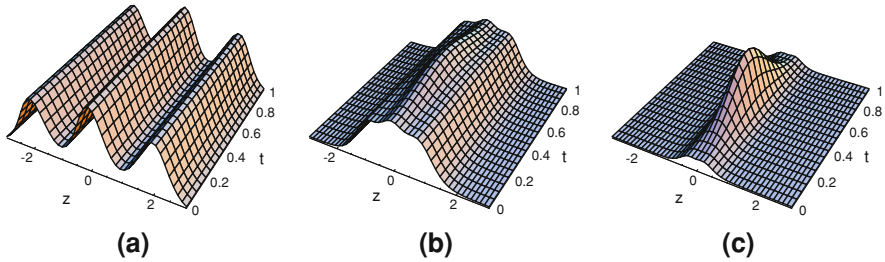
Fig. 8.9 Action density of $SU(3)$ calorons as a function of time and one space coordinate. Each caloron has three dyon constituents, and the dyon separations are increasing from *right* to *left*. Note the persistence in time of the widely separated consituent dyons. From Diakonov [50]

2. k-string tensions extracted from Polyakov line correlators, and from spacelike Wilson loops, agree with one another, and follow the Sine Law. The calculation of spacelike string tensions is similar to Polyakov's determination of the string tension in compact QED_3.
3. At high temperatures, the free energy is minimized at the trivial holonomy (deconfinement), and there is a phase transition between the confined and deconfined phases at a critical temperature T_c, with $T_c/\sqrt{\sigma}$ in impressive agreement with lattice simulations.
4. $\text{Tr}[\mathscr{P}] = 0$ is also obtained in G_2 gauge theory, leading to a claim [50] that center symmetry is of no fundamental importance to confinement.

While the Diakonov-Petrov calculation is impressive, their proposed measure is not positive definite, as pointed out by Bruckmann et al. [51], which violates a known property of the exact measure. A different approach, in which the measure for the dyon gas is parametrized and the system is then simulated numerically, has been followed by Gerhold, Ilgenfritz, and Mueller-Preussker [52]. Their results also indicate, not too surprisingly, that a dyon gas gives rise to a confining potential.

The caloron idea is probably the most promising current version of monopole confinement in pure non-abelian gauge theories, but it is basically (in certain gauges) a superposition of monopoles with spherically symmetric abelian fields, and this leads to the same questions raised in connection with monopole Coulomb gases. In a gauge in which the dyons are asymptotically abelian, what is the behavior of a double charged abelian loop? In a monopole Coulomb gas, the string tension of a double charged loop is simply twice the string tension of a single charged loop, yet this cannot be correct in a non-abelian theory, due to screening by gluons. In numerical simulations it has been found that monopole fields are concentrated along vortex lines, and this distribution avoids generating a string tension for double-charged abelian Wilson loops. If the dyon gas, like the monopole Coulomb gas, predicts an area law for double-charged abelian loops, then the spherically symmetric magnetic field distribution assumed for dyons cannot be right.

Another issue is the probability distribution for Polyakov holonomies. If the Polyakov loop vanishes for a loop in the fundamental representation, then the corresponding Polyakov loop in the adjoint representation is negative. In fact, if $\mathscr{P}(\mathbf{x})$ is the Polyakov holonomy, one can express the probability density $\rho(g)$ that $\mathscr{P}(\mathbf{x})$ is equal to group element g via a character expansion. For, e.g., the case of $SU(2)$, we have

$$
\begin{aligned}
\rho[g] &= \langle \delta[\mathscr{P}(\mathbf{x}) - g] \rangle \\
&= \Big\langle \sum_j \chi_j[\mathscr{P}(\mathbf{x})] \chi_j[g] \Big\rangle \\
&\approx 1 + a_1 \chi_1[g],
\end{aligned}
\tag{8.59}
$$

where $a_1 = \langle \chi_1[\mathscr{P}(\mathbf{x})] \rangle$ is the expectation value of the Polyakov line in the adjoint ($j = 1$) representation. The $j =$ half-integer representations can be dropped, since they should vanish if center symmetry is unbroken, and the expectation values of $\mathscr{P}(\mathbf{x})$ in representations $j > 1$ are negligible. If $a_1 < 0$ then the maximally non-trivial holonomy is the most probable, in conformity with the dyon gas idea. If, on the other hand, $a_1 > 0$, then $\rho[\mathscr{P}]$ is peaked at $\mathscr{P} = $ a center element, as would be expected in the center vortex picture. The expectation value of an adjoint Polyakov line in the confined phase is actually quite small and difficult to measure, nevertheless there are measurements of this quantity in $SU(3)$ gauge theory just below the deconfinement transition, and the result is that the expectation value is small but positive [53].

Finally, if it were really true that the Polyakov line expectation value vanishes *exactly* in G_2 gauge theory at low temperatures, then center symmetry is indeed irrelevant to confinement, as claimed in [50]. However, it is much more likely that this prediction of the dyon gas approximation is just incorrect. In G_2 it is necessary to bind at least three gluons to screen a Polyakov line in the fundamental representation, and a gluelump of this sort will be very heavy. As a result, the Polyakov line in G_2 gauge theory will be very small indeed, and it will be exceedingly difficult to measure this tiny non-zero value by numerical simulation. But unless current ideas about charge screening are completely wrong, the Polyakov line in G_2 must be non-zero. For the same reason, any large G_2 Wilson loop can be screened by gluons, and the asymptotic string tension must vanish. This means that the lack of center symmetry implies the absence, at sufficiently large scales, of magnetic disorder. The fact that the dyon gas approximation misses this fact should not be taken as implying that center symmetry is irrelevant; instead it means that the dyon gas approximation is itself missing something which is important to the asymptotic physics.

8.6 Critique: Monopoles and N-ality

The principal critique of the monopole confinement mechanism, at least in the form of a monopole Coulomb gas or a dual abelian Higgs model, has been stated

repeatedly in this chapter. Confinement of electric charge by monopole conden-
sation is an abelian mechanism, and objects with n units of the fundamental
electric charge will be confined, for any $n > 1$. In a suitable gauge, the expectation
value of any abelian Wilson loop, or any power of any abelian loop, will have an
area law falloff. This prediction is at odds with N-ality, and with color-screening
by gluons. The implication is that the distribution of magnetic monopoles, and the
magnetic flux associated with those monopoles, cannot be as predicted by the
simple theories just mentioned.

Let us consider, as an example, a Wilson loop in the adjoint representation of
$SU(2)$ pure gauge theory. In all of the monopole confinement theories we have
discussed there is some gauge, defined by either an adjoint Higgs field or (as in the
abelian projection and caloron cases) by some composite operator, in which the
monopole fields, far from the monopole locations, all point in the same direction in
color space. Let us suppose that this is the color 3 direction. The approximation
which yields confinement is to replace the full gauge field by the field due to the
monopoles, in which case

$$
\begin{aligned}
W_A[C] &= \left\langle \chi_{j=1}\left[P \exp\left\{ ig \oint_C dx^\mu A_\mu^a \frac{\sigma_a}{2} \right\} \right] \right\rangle \\
&\approx \left\langle \chi_{j=1}\left[\exp\left\{ ig \oint_C dx^\mu A_\mu^3 \frac{\sigma_3}{2} \right\} \right] \right\rangle \\
&= \sum_{m=-1}^{1} \left\langle \exp\left[img \oint_C dx^\mu A_\mu^3 \right] \right\rangle .
\end{aligned}
\tag{8.60}
$$

This quantity does not have an area-law falloff, due to the $m = 0$ contribution.
From this fact it would seem that the monopole confinement mechanism is
compatible with the N-ality condition.

That conclusion is mistaken, however. Consider instead the trace, in the fun-
damental representation, of the square of the Wilson loop holonomy, i.e.,

$$
W^{(2)}(C) = \left\langle Tr\left[\{U(C)\}^2 \right] \right\rangle ,
\tag{8.61}
$$

where $U(C)$ is the holonomy

$$
U(C) = P \exp\left[ig \oint_C dx^k A_k^a \frac{\sigma_a}{2} \right] .
\tag{8.62}
$$

The square of $U(C)$ is a zero N-ality operator in $SU(2)$ gauge theory, and cannot
have an area law falloff. This should also be clear from the identity

$$
W^{(2)}(C) = \langle Tr_A[U(C)] \rangle - 1 .
\tag{8.63}
$$

However, the same logic as above leads to the estimate, in the monopole confinement picture

$$W^{(2)}(C) = \sum_{m=\pm 1} \left\langle \exp\left[img \oint_C dx^\mu A^3_\mu \right] \right\rangle. \qquad (8.64)$$

This time the $m = 0$ component is absent, and the operator corresponds to a sum of two loops, with opposite orientations, each carrying two units of the abelian charge. If a Wilson loop with one unit of abelian charge has an area law, in the monopole confinement picture, so does a loop with two units of abelian charge, as can be verified explicitly in compact QED$_3$ [9]. This means that the monopole Coulomb gas scenario is simply not compatible with the N-ality condition.

To obtain an effective long-distance theory involving *only* the abelian degrees of freedom, it is necessary to integrate out the charged degrees of freedom (the "W-bosons"); these degrees of freedom cannot be ignored simply because they are massive. When this is done, the distribution of magnetic fields emanating from the monopoles in the effective theory must bring the abelian theory into compliance with N-ality. The simplest possibility is that the monopole worldlines lie on vortex sheets, and monopole magnetic fields are collimated along thick center vortices, as illustrated in Figs. 8.5 and 8.6 above. As we have noted, there is numerical evidence for this kind of monopole-vortex connection in pure lattice Yang–Mills theory, from abelian projection in maximal abelian gauge.

References

1. Coleman, S.: The magnetic monopole fifty years later. In: Zichichi, A. (ed.) The Unity of the Fundamental Interactions. Kluwer Academic Publishers, Dordrecht (1983)
2. 't Hooft, G.: High energy physics. In: Zichichi A (ed.) Gauge theories with unified weak, electromagnetic, and strong interactions. EPS International Conference, Palermo (1975)
3. Mandelstam S.: Vortices and quark confinement in non-abelian gauge theories. Phys. Reports **23C**, 245–249 (1976)
4. Polyakov, A.: Gauge Fields and Strings. Harwood Academic Publishers, Chur (1987)
5. DeGrand, T., Toussaint, D.: Topological excitations and Monte Carlo simulation of Abelian gauge theory. Phys. Rev. D **22**, 2478–2489 (1980)
6. Polyakov, A.: Compact gauge fields and the infrared catastrophe. Phys. Lett. B **59**, 82–84 (1975)
7. Polyakov A.: Quark confinement and topology of gauge groups. Nucl. Phys. B **120**, 429–458 (1977)
8. Banks, T., Myerson, R., Kogut, J.B.: Phase transitions in Abelian lattice gauge theories. Nucl. Phys. B **129**, 493–510 (1977)
9. Ambjørn, J., Greensite, J.: Center disorder in the 3D Georgi–Glashow model. JHEP **05**, 1–23 (1998) [arXiv: hep-lat/9804022]
10. Watson, G.N.: Three triple integrals. Quart. J. Math. **10**, 266–276 (1939)
11. Gopfert, M., Mack, G.: Proof of confinement of static quarks in 3-dimensional U(1) lattice gauge theory for all values of the coupling constant. Commun. Math. Phys. **82**, 545–606 (1981)

12. 't Hooft, G.: Magnetic monopoles in unified gauge theories. Nucl. Phys. B **79**, 276–284 (1974)
13. Polyakov, A.M.: Particle spectrum in the quantum field theory. JETP Lett. **20**, 194–195 (1974)
14. Bornyakov, V.G., Ilgenfritz E.M., Mitrjushkin V.K., Zadorozhnyi, A.M., Muller-Preussker, M.: Investigation of the vacuum structure of the Georgi-Glashow model on the lattice. Z. Phys. C **42**, 633–640 (1989)
15. Schierholz, G., Seixas, J., Teper, M.: Patterns of symmetry restoration in gauge-scalar theories: A Monte Carlo simulation of the SO(3) Georgi–Glashow model. Phys. Lett. B **157**, 209–215 (1985)
16. Brower, R.C., Kessler, D.A., Schalk, T., Levine, H., Nauenberg, M.: The SU(2) adjoint Higgs model. Phys. Rev. D **25**, 3319–3324 (1982)
17. Shifman, M., Ünsal, M.: QCD-like theories on R(3) × S(1): A smooth journey from small to large r(S(1)) with double-trace deformations. Phys. Rev. D **78**, 065004-1–065004-20 (2008) [arXiv:0802.1232 [hep-th]].
18. Ünsal, M.: Magnetic bion condensation: A new mechanism of confinement and mass gap in four dimensions. Phys. Rev. **D80**, 065001-1–065001-21 (2009) [arXiv:0709.3269 [hep-th]]
19. Nielsen, H.B., Olesen, P.: Vortex line models for dual strings. Nucl. Phys. B **61**, 45–61 (1973)
20. Seiberg, N., Witten, E.: Electric–magnetic duality, monopole condensation, and confinement in N = 2 supersymmetric Yang–Mills theory. Nucl. Phys. B **426**, 19–52 (1994) [Erratum-ibid. B **430**, 485 (1994)] [arXiv:hep-th/9407087]
21. Seiberg, N., Witten, E.: Monopoles, duality and chiral symmetry breaking in N = 2 supersymmetric QCD. Nucl. Phys. B **431**, 484–550 (1994) [arXiv:hep-th/9408099]
22. Bilal, A.: Duality in N = 2 SUSY SU(2) Yang–Mills theory: A pedagogical introduction to the work of Seiberg and Witten. arXiv:hep-th/9601007
23. Alvarez-Gaume, L., Hassan, S.F.: Introduction to *S*-duality in *N* = 2 supersymmetric gauge theories (A pedagogical review of the work of Seiberg and Witten). Fortsch. Phys. **45**, 159–236 (1997) [arXiv:hep-th/9701069]
24. Lerche, W.: Introduction to Seiberg-Witten theory and its stringy origin. Nucl. Phys. Proc. Suppl. **55B**, 83–117 (1997) [Fortsch. Phys. **45**, 293 (1997)] [arXiv:hep-th/9611190]
25. Douglas, M.R., Shenker, S.H.: Dynamics of SU(N) supersymmetric gauge theory. Nucl. Phys. B **447**, 271–296 (1995)
26. Shifman, M., Yung, A.: Supersymmetric solitons and how they help us understand non-abelian gauge theories. Rev. Mod. Phys. **79**, 1139–1196 (2007) [arXiv:hep-th/0703267]
27. Konishi, K.: Advent of non-abelian vortices and monopoles—further thoughts about duality and confinement. Prog. Theor. Phys. Suppl. **177**, 83–98 (2009) [arXiv:0809.1370 [hep-th]]
28. Shifman, M.: Understanding confinement in QCD: elements of a Big Picture. Int. J. Mod. Phys. A **25**, 4015–4031 (2010) [arXiv:1007.0531 [hep-th]]
29. 't Hooft, G.: Topology of the gauge condition and new confinement phases in nonabelian gauge theories. Nucl. Phys. B **190** [FS3], 455–478 (1981)
30. Kronfeld, A., Laursen, M., Schierholz, G., Wiese, U.-J.: Monopole condensation and color confinement. Phys. Lett. B **198**, 516–520 (1987)
31. Bali, G.S., Bornyakov, V., Muller-Preussker, M., Schilling, K.: Dual superconductor scenario of confinement: A systematic study of Gribov copy effects. Phys. Rev. D **54**, 2863–2875 (1996) [arXiv:hep-lat/9603012]
32. Shiba, H., Suzuki, T.: Monopoles and string tension in SU(2) QCD. Phys. Lett. B **333**, 461–466 (1994), hep-lat/9404015
33. Stack, J., Neiman, S., Wensley, R.: String tension from monopoles in SU(2) lattice gauge theory. Phys. Rev. D **50**, 3399–3405 (1994), hep-lat/9404014
34. Del Debbio, L., Di Giacomo, A., Paffuti, G.: Detecting dual superconductivity in the ground state of gauge theory. Phys. Lett. B **349**, 513–518 (1995) [arXiv:hep-lat/9403013]

35. Di Giacomo, A., Lucini, B., Montesi, L., Paffuti, G.: Color confinement and dual superconductivity of the vacuum. 1. Phys. Rev. D **61**, 034503-1–034503-10 (2000) [arXiv:hep-lat/9906024]
36. Greensite, J., Lucini, B.: Is confinement a phase of broken dual gauge symmetry? Phys. Rev. D **78**, 085004-1–085004-7 (2008) [arXiv:0806.2117 [hep-lat]]
37. Ambjørn, J., Giedt, J., Greensite, J.: Vortex structure versus monopole dominance in Abelian projected gauge theory. JHEP **02**, 1–39 (2000) [arXiv: hep-lat/9907021]
38. Suzuki, T., Yotsuyanagi, I.: A possible evidence for Abelian dominance in quark confinement. Phys. Rev. D **42**, 4257–4260 (1990)
39. Gubarev, F.V., Kovalenko, A.V., Polikarpov, M.I., Syritsyn, S.N., Zakharov, V.I.: Fine tuned vortices in lattice SU(2) gluodynamics. Phys. Lett. B **574**, 136–140 (2003) [arXiv:hep-lat/0212003]
40. Belavin, A.A., Polyakov, A.M., Shvarts, A.S., Tyupkin, Yu.S.: Pseudoparticle solutions of the Yang–Mills equations. Phys. Lett. B **59**, 85–87 (1975)
41. Kraan, T.C., van Baal, P.: Periodic instantons with nontrivial holonomy. Nucl. Phys. B **533**, 627–659 (1998) [arXiv:hep-th/9805168]
42. Kraan, T.C., van Baal, P.: Exact T duality between calorons and Taub-NUT spaces. Phys. Lett. B **428**, 268–276 (1998) [arXiv:hep-th/9802049]
43. Lee, K.M., Lu, C.h.: SU(2) calorons and magnetic monopoles. Phys. Rev. D **58**, 025011-1–025011-7 (1998) [arXiv:hep-th/9802108]
44. Diakonov, D., Petrov, V.: Confining ensemble of dyons. Phys. Rev. D **76**, 056001-1–056001-22 (2007) [arXiv:0704.3181 [hep-th]]
45. Bogomol'ny, E.B.: Stability of classical solutions. Sov. J. Nucl. **24**, 449–458 (1976)
46. Prasad, M.K., Sommerfield, C.M.: An exact classical solution for the 't Hooft monopole and the Julia–Zee dyon. Phys. Rev. Lett. **35**, 760–762 (1975)
47. Harrington, B.J., Shepard, H.K.: Thermodynamics of the Yang–Mills gas. Phys. Rev. D **18**, 2990–2994 (1978)
48. Harrington, B.J., Shepard, H.K.: Periodic Euclidean solutions and the finite temperature Yang–Mills gas. Phys. Rev. D **17**, 2122–2125 (1978)
49. Kraan, T.C., van Baal, P.: Constituent monopoles without gauge fixing. Nucl. Phys. Proc. Suppl. **73**, 554–556 (1999) [arXiv:hep-lat/9808015]
50. Diakonov, D.: Topology and confinement. Nucl. Phys. Proc. Suppl. **195**, 5–45 (2009) [arXiv:0906.2456 [hep-ph]]
51. Bruckmann, F., Dinter, S., Ilgenfritz, E.M., Muller-Preussker, M., Wagner, M.: Cautionary remarks on the moduli space metric for multi-dyon simulations. Phys. Rev. D **79**, 116007-1–116007-12 (2009) [arXiv:0903.3075 [hep-ph]]
52. Gerhold, P., Ilgenfritz, E.M., Muller-Preussker, M.: An SU(2) KvBLL caloron gas model and confinement. Nucl. Phys. B **760**, 1–37 (2007) [arXiv:hep-ph/0607315]
53. Gupta, S., Huebner, K., Kaczmarek, O.: Renormalized Polyakov loops in many representations. Phys. Rev. D **77**, 034503-1–034503-16 (2008) [arXiv:0711.2251 [hep-lat]]

Chapter 9
Coulomb Confinement

In Coulomb gauge, the longitudinal electric field $\mathbf{E}_L^a = -\nabla \phi^a$ associated with a static charge is obtained by solving the Gauss Law $\mathbf{D} \cdot \mathbf{E}^a = \rho^a$ for ϕ^a, and the energy density E_L^2 associated with this electric field, integrated over all space, is the Coulomb energy. There is an ongoing effort, pioneered originally by Gribov [1] and Zwanziger [2, 3], to demonstrate that the Coulomb potential associated with quark–antiquark sources is confining. The argument that the Coulomb potential might have this property relies, in an essential way, on the fact that in Coulomb gauge the functional integration cannot range over all possible gauge fields satisfying the gauge condition, but instead must be restricted to a subspace in the space of all gauge fields, bounded by a hypersurface known as the *Gribov horizon*.

9.1 The Gribov Horizon

We have already seen that the Coulomb and Landau gauge conditions do not fix the gauge uniquely. In each case there is some remnant symmetry left, which may or may not be spontaneously broken, depending on the gauge theory and the gauge and matter coupling constants. Beyond this, however, there is a further ambiguity, first pointed out by Gribov [1]: In a gauge theory, each gauge field is a member of a set of gauge-equivalent configurations known as a *gauge orbit*. If we imagine that the gauge-fixing condition restricts the functional integral to a certain subspace $\mathscr{S}$ of the space of all gauge fields, then it turns out that a typical gauge orbit intersects this gauge-fixing hypersurface many times (not counting the multiplicity due to remnant symmetries), and these multiple intersections are known as *Gribov copies*. At each Gribov copy, the Faddeev–Popov (F–P) determinant may be positive or negative, which raises the alarming possibility that by summing over all copies, the functional integral may actually vanish! In fact, something of the sort is exactly what happens in BRST quantization of lattice gauge theories at the non-perturbative level, according to a famous result [4] known as

Neuberger's Theorem Let Q be a gauge or BRST invariant observable in a lattice gauge theory with compact link variables, with S_{gf} the gauge-fixing part of the action, and let $c, \overline{c}, b$ denote the ghost, antighost, and auxiliary fields, respectively, that arise in the BRST quantization procedure.[1] Then

$$\langle Q \rangle = \frac{\int DU Dc D\overline{c} Db \; Q e^{-(S+S_{gf})}}{\int DU Dc D\overline{c} Db \; e^{-(S+S_{gf})}} \tag{9.1}$$
$$= \frac{0}{0}.$$

This means, of course, that BRST quantization is not really defined at the non-perturbative level. The problem can be traced, in the Faddeev–Popov approach, to the exact cancellation of contributions due to Gribov copies with F–P determinants of opposite sign [4].

It was suggested by Gribov that the domain of functional integration in Coulomb and Landau gauges should be restricted to a subspace, now known as the *Gribov region*, where the eigenvalues of the F–P operator, and hence the F–P determinant, are all positive. In this proposal the different Gribov copies cannot cancel. Consider, in the continuum, the gauge which minimizes the quantity

$$R[A] = \int d^4x \; \text{Tr}[A_\mu(x)A_\mu(x)]. \tag{9.2}$$

If $A_\mu(x)$ is a (local) minimum of $R[A]$, and

$$^\theta A_\mu(x) = e^{i\theta^a(x)L_a} A_\mu(x) e^{-i\theta^a(x)L_a} - \frac{i}{g} e^{i\theta^a(x)L_a} \partial_\mu e^{-i\theta^a(x)L_a} \tag{9.3}$$

is a gauge transformation of $A_\mu(x)$ (with $\{L_a\}$ the group generators), then it must be that $R[^\theta A]$ is minimized at $\theta(x) = 0$. This requires, first of all, stationarity with respect to infinitesimal variations of $\theta(x)$, i.e. that

$$\left(\frac{\delta R[^\theta A]}{\delta \theta^a(x)} \right)_{|\theta=0} = \frac{1}{g} \; \partial_\mu A_\mu^a(x) \tag{9.4}$$

must vanish, which is the usual Landau gauge condition. Secondly, for a local minimum the Hessian

$$\left(\frac{\delta^2 R[^\theta A]}{\delta \theta^a(x) \delta \theta^b(y)} \right)_{|\theta=0} = \frac{1}{g^2} \; M_{xy}^{ab} \tag{9.5}$$

---

[1] BRST quantization is described in most modern textbooks on quantum field theory, e.g. [5, 6].

must have only positive semidefinite eigenvalues, where

$$M_{xy}^{ab} = -\partial_\mu D_\mu^{ab} \delta^4(x-y) \tag{9.6}$$

is the F–P operator, with $D_\mu^{ab} = \delta^{ab}\partial_\mu - gf^{acb}A^c$ the covariant derivative. From this it is clear that the Landau gauge Gribov region consists of all local minima of $R[A]$ for each gauge orbit.

The lattice version of Landau gauge, restricted to the Gribov region, is the gauge which minimizes

$$R = -\sum_x \sum_\mu \mathrm{ReTr}[U_\mu(x)]. \tag{9.7}$$

Denoting gauge-transformed links by

$$^\theta U_\mu(x) = e^{i\theta^a(x)L_a} U_\mu(x) e^{-i\theta^a(x+\widehat{\mu})L_a}, \tag{9.8}$$

the stationarity condition for $R[^\theta U]$ is

$$\left(\frac{\partial R[^\theta U]}{\partial \theta^a(x)}\right)_{|\theta=0} = i\sum_\mu \mathrm{Tr}[L_a(U_\mu(x) + U_\mu^\dagger(x-\hat{\mu})] = 0, \tag{9.9}$$

which becomes the usual Landau gauge condition $\partial_\mu A_\mu^a(x) = 0$ in the continuum limit. The lattice F–P operator M_{xy}^{ab} is again obtained from the Hessian matrix and, as in the continuum, the condition that the field configuration is a local *minimum* of R is that the eigenvalues of M_{xy}^{ab} are all positive (i.e. all directions on the gauge orbit away from the stationary point increase the value of R).

In Coulomb gauge, the sum over the spacetime index $\mu = 0, 1, 2, 3$ in the above expressions is replaced by a sum over a space index $k = 1, 2, 3$, and the four-dimensional delta function in (9.6) is replaced by a three-dimensional delta function.

In lattice simulations one generates configurations in the usual way, and observables are computed after fixing these configurations to the required gauge. But the techniques which are used to gauge fix on the lattice only obtain local minima of $R[U]$, rather than arbitrary stationary points, and this guarantees that the eigenvalues of the F–P operator M_{xy}^{ab} are almost all positive. The exceptions are the trivial zero modes, whose origin is the remnant global symmetry $U_\mu(x) \to GU_\mu(x)G^\dagger$, where $G \in SU(N)$ is any position-independent group element. The remnant symmetry implies that at any stationary point of R there must be flat directions along the gauge orbit, corresponding to the generators of the global transformations, and these flat directions mean that there must be zero modes of the F–P operator. These are the trivial eigenstates of the F–P operator M, which are easily seen, by inspection of (9.6), to be

$$\phi_n^a(x) = \frac{1}{\sqrt{V}}\delta_{na}, = 1, 2, \ldots, N^2 - 1. \tag{9.10}$$

The statement that the F–P determinant is positive in the Gribov region really refers to the determinant of the operator in the subspace orthogonal to these trivial zero modes.

The problem with defining the path integral, in Landau or Coulomb gauge, as a sum over all copies in the Gribov region is that it cannot be guaranteed that all gauge orbits are weighted equally; it may be that different gauge orbits intersect the Gribov region different numbers of times. Ideally, one would like to restrict the region of functional integration to a subset of the Gribov region, which each gauge trajectory intersects only once. An example is the *Fundamental Modular Region* suggested by Zwanziger [3], which consists, among all local minima of R, of configurations which give the absolute minimum of R along the gauge orbit. These configurations, presumably, are unique. Unfortunately there is no known procedure for locating such absolute minima in practice, nor is there any reason to think that, if we ordered the local minima in order of their value of R, the gauge copy with the lowest minimum in the set is more physically relevant than, say, the copy with the ten-thousandth minimum in the sequence. Lattice simulations in Coulomb and Landau gauges generally ignore this issue, in the hope that the choice of gauge copy will not make a serious difference to the final results. Some simulations, however, apply a technique known as simulating annealing [7], which supplies a local minimum of R that is almost always lower than that of a gauge copy chosen at random. But it is not entirely clear, on physical grounds, why a lower minimum should be preferred.

The Gribov region is bounded in its entirety, and the Fundamental Modular region is bounded in part, by a hypersurface known as the first Gribov horizon, as illustrated in Fig. 9.1. Consider the eigenvalue equation for the F–P matrix
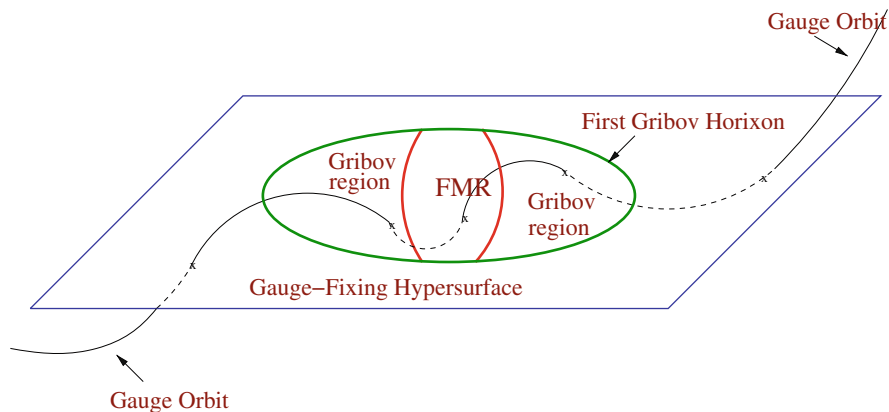
Fig. 9.1 The rectangular region indicates the subspace, in the space of all gauge configurations satisfying the (Coulomb or Landau) gauge condition. A typical gauge orbit intersects the subspace many times; the intersections (marked "x") are known as "Gribov copies." Within the Gribov region the non-trivial eigenvalues of the F–P operator are all positive; this region is bounded by the first Gribov horizon, where one non-trivial eigenvalue is zero. One can also define subregions within the Gribov region, such as the Fundamental Modular region (FMR), which are intersected only once by each gauge orbit

$$M_{xy}^{ab} \phi_y^{b(n)} = \lambda_n \phi_x^{a(n)}, \tag{9.11}$$

where summation (lattice) or integration (continuum) over the repeated y-coordinate is implied. The Gribov region, by definition, is the region in which all the λ_n are positive, apart from (in $SU(N)$) the first $N^2 - 1$ zero eigenvalues of the trivial zero modes. The *first* Gribov horizon is the boundary of the Gribov region, where the lowest non-trivial eigenmode also has a zero eigenvalue, while all other eigenvalues are positive (here I will use the terms "Gribov horizon" and "first Gribov horizon" interchangeably).

The speculation that all this has something to do with confinement is based on the fact that the Coulomb potential, as we will see below, involves the inverse of the F–P operator M. Zwanziger [3] has argued that most of the volume of the Gribov region is concentrated near the boundary, i.e. at the first Gribov horizon, in much the same way that the volume of a sphere of radius R, in d dimensions, is concentrated at $r \approx R$ for large d. Since the first non-trivial eigenvalue of M vanishes at the first Gribov horizon, it is possible that the near-zero eigenvalues of M, for configurations very near the Gribov horizon, will considerably enhance the Coulomb potential, and possibly strengthen it to a confining form. In the next sections we will see evidence that this speculation is probably correct.

9.2 Coulomb Potential on the Lattice

Coulomb gauge, like axial and temporal gauge (but *unlike* the covariant gauges) is a "physical" gauge, in the sense that it is possible to formulate a ghost-free Hamiltonian operator, and avoid the introduction of propagating ghost fields. The classical Coulomb-gauge Hamiltonian of Yang–Mills theory is $H = H_{glue} + H_{coul}$, where [8]

$$H_{glue} = \frac{1}{2} \int d^3x (\mathbf{E}_T^2 + \mathbf{B}^2)$$

$$H_{coul} = \frac{1}{2} \int d^3x d^3y \, \rho^a(x) K^{ab}(x,y) \rho^b(y) \tag{9.12}$$

$$K^{ab}(x,y;A) = \left[(M^{-1})_{\mathbf{xz}}^{ac} (-\nabla^2)_{\mathbf{z}} (M^{-1})_{\mathbf{zy}}^{cb} \right]$$

$$\rho^a = \rho_m^a - g f^{abc} A_k^b E_{Tk}^c,$$

and where E_T denotes the transverse E-field $\nabla \cdot \mathbf{E}_T^a = 0$.[2] The color-charge density due to matter fields is given by $\rho_m(x)$. If this charge density is due to static

---

[2] At the quantum level, the Hamiltonian contains factors of $(\det M)^{1/2}$ and $(\det M)^{-1/2}$, which cancel out in the classical limit where the ordering of operators is irrelevant [9]. These factors also cancel out in the expression for the Coulomb energy due to static color sources.

quark–antiquark color charge sources separated by a distance R, then the corresponding Coulomb interaction + self-energies are given by

$$\mathcal{E}_{Coul}(R) = \frac{1}{2} \int d^3x d^3y \langle \rho_m^a(x) K^{ab}(x,y) \rho_m^b(y) \rangle. \qquad (9.13)$$

In the path-integral formulation it is useful to introduce Faddeev–Popov ghost fields, but in Coulomb gauge these ghost fields do not propagate in time. Their propagator at equal times is given by $\langle (M^{-1})_{\mathbf{xy}}^{ab} \rangle$.

A glance at the operator $K(x, y; A)$ in (9.12) shows that it involves two factors of the inverse F–P operator M^{-1}. If the functional integration is dominated by configurations at or near the first Gribov horizon, then the eigenvalue spectrum of M begins at or near zero. This leads to the suggestion, long advocated by Gribov [1] and Zwanziger [3], that $\langle \rho_m K \rho_m \rangle$ is so greatly enhanced in the infrared by the near-zero eigenvalues of M that a linear potential

$$\langle K(x,y;A) \rangle \sim \sigma |\mathbf{x} - \mathbf{y}| \qquad (9.14)$$

is obtained. Since $\langle K(x, y; A) \rangle$ is essentially the instantaneous piece of the correlator $\langle A_0(x) A_0(y) \rangle$, confinement could then be understood as arising from one-gluon exchange with a (highly non-perturbative) dressed gluon propagator.

We would like to study the validity of this Gribov–Zwanziger scenario numerically, and the first step is to figure out how to compute the Coulomb energy on the lattice. Let us begin with a physical state containing a heavy quark–antiquark pair, separated by distance R:

$$|\Psi_{q\bar{q}}\rangle = \bar{q}(0) q(R) |\Psi_0\rangle, \qquad (9.15)$$

where Ψ_0 is the vacuum state of Yang–Mills theory in Coulomb gauge. Note that this is not necessarily the lowest energy state containing the two static sources, but it *is* the state in which the $\bar{q}(0) q(R)$ interaction energy is entirely Coulombic, i.e.

$$\begin{aligned} \mathcal{E}_{Coul}(R) &= \langle \Psi_{q\bar{q}} | H | \Psi_{q\bar{q}} \rangle - \langle \Psi_0 | H | \Psi_0 \rangle \\ &= V_{coul}(R) + E_{se}, \end{aligned} \qquad (9.16)$$

where the Coulomb potential $V_{coul}(R)$ comes from the non-local $\rho_m K \rho_m$ term in the Hamiltonian, and E_{se} is some R-independent self-energy term. We would like to know, first of all, if $V_{coul}(R)$ is confining. If it is confining, then is it asymptotically linear? If linear, does the Coulomb string tension σ_{coul} equal the string tension σ of the static quark potential? Finally, is there any connection to the center vortex mechanism? What happens to the Coulomb potential if vortices are removed?

In order to compute the Coulomb potential numerically, we begin with

$$G(R,T) = \langle \Psi_{q\bar{q}} | e^{-(H-E_0)T} | \Psi_{q\bar{q}} \rangle, \qquad (9.17)$$

where E_0 is the vacuum energy. For very massive quarks in an $SU(N)$ gauge theory, by the same that reasoning that led to introducing Polyakov lines, and up to an overall constant of $O(m^{2T})$,

$$G(R,T) = \frac{1}{N}\langle \text{Tr}[L^\dagger(0,T)L(R,T)]\rangle, \tag{9.18}$$

where

$$L(\mathbf{x},T) = P \exp\left[i \int_0^T dt\, A_0(\mathbf{x},t)\right] \tag{9.19}$$

is a Wilson line in the time direction. Here T is some finite time lapse, but it is *not* the full extension of the lattice in the time direction. Inserting a complete set of states in (9.17)

$$G(R,T) = \sum_n |\langle \Psi_n | \Psi_{q\bar{q}}\rangle|^2 e^{-(E_n - E_0)T}. \tag{9.20}$$

Denote

$$V(R,T) = -\frac{d}{dT}\log[G(R,T)]. \tag{9.21}$$

Then it is easy to see that

$$\begin{aligned}
\mathcal{E}_{coul}(R) &= \langle \Psi_{q\bar{q}} | H - E_0 | \Psi_{q\bar{q}}\rangle \\
&= V_{coul}(R) + E_{se} \\
&= V(R,0),
\end{aligned} \tag{9.22}$$

while

$$\begin{aligned}
\mathcal{E}_{min}(R) &= V(R) + E'_{se} \\
&= \lim_{T\to\infty} V(R,T),
\end{aligned} \tag{9.23}$$

where $\mathcal{E}_{min}$ is the lowest energy that can be obtained among states which contain the static $q\bar{q}$ system, and $V(R)$ is the static quark potential. In general there is no reason that the minimal energy state should be the $\Psi_{q\bar{q}}$ state defined in (9.15), which is used to extract the Coulomb potential. With lattice regularization, E_{se} and E'_{se} are negligible at large R, compared to $V(R)$. Then, since $\mathcal{E}_{min} \leq \mathcal{E}_{coul}$, it follows that

$$V(R) \leq V_{coul}(R), \tag{9.24}$$

as first noted by Zwanziger in [10]. Therefore, if confinement exists at all, there must also be a confining Coulomb potential.

On the lattice

$$L(x, T) = U_0(x, 1) U_0(x, 2) \cdots U_0(x, T)$$

$$V(R, T) = \frac{1}{a} \log \left[\frac{G(R, T)}{G(R, T + 1)} \right]. \tag{9.25}$$

Then

$$V(R, 0) = V_{coul}(R) + \text{const.}$$

$$\lim_{T \to \infty} V(R, T) = V(R) + \text{const.}, \tag{9.26}$$

where, using the fact that $L(\mathbf{x}, 0) = \mathbb{1}$ by definition,

$$V(R, 0) = -\log[G(R, 1)] \tag{9.27}$$

in lattice units. This relation allows us to derive the lattice Coulomb potential (= $V(R,0)$ up to a constant) from the timelike link-link correlator $G(R,1)$ at equal times, and compare to the static potential $V(R)$ on the lattice.

We may also consider the Coulomb energy of a one-particle state $\Psi_q^a = q^a(x) \Psi_0$. By the same reasoning as above, the Coulomb energy computed on the lattice will be

$$\mathscr{E}_q \propto -\log[\langle \text{Tr}[U_0(x)] \rangle], \tag{9.28}$$

and therefore the Coulomb energy of this one-particle state is infinite if the expectation value of a timelike link vanishes in Coulomb gauge. Recall that this was also the condition, discussed in Chap. 3, that the remnant gauge symmetry $U_0(\mathbf{x}, t) \to g(t) U_0(\mathbf{x}, t) g^\dagger(t + 1)$ is unbroken. It should be understood, however, that $\mathscr{E}_q$ is the energy of a specific state, and as such it is only an upper bound on the minimal energy of a states containing a single static quark. Therefore $\mathscr{E}_q = \infty$ is a necessary but not sufficient condition for the non-existence of finite-energy isolated charge states.

Returning to the lattice Coulomb potential for static $q\bar{q}$ sources, we begin with a check: Define $\sigma(T)$ from a fit of $V(R,T)$ to

$$V(R, T) = c(T) - \frac{\pi}{12R} + \sigma(T)R, \tag{9.29}$$

where $-\pi/12R$ is the Lüscher term, and see if $\sigma(T) \to \sigma$ in the large-T limit. This seems to work out as expected, as seen in Fig. 9.2. Figure 9.3 shows the data for $V(R,4)$, and Fig. 9.4 shows the data for $V(R,0)$, at $\beta = 2.5$. At this coupling $V(R,4)$ is not far off the static quark potential $V(R)$, while $V(R,0)$ gives us the Coulomb potential $V_{coul}(R)$ up to some constant. Each plot also displays the result (lower line) obtained from lattice configurations with vortices removed by the de Forcrand-D'Elia procedure, which was introduced in Section 6.3.3. There is a very substantial difference in the slope of the Coulomb potential (Fig. 9.4) and of the static quark potential (Fig. 9.3). Figure 9.5 shows the ratio of σ_{coul}/σ of the Coulomb and

Fig. 9.2 Falloff of $\sigma(T)$ with increasing T at $\beta = 2.3, 2.4, 2.5$. *Solid lines* indicate the accepted values of the asymptotic string tension at each β value, with *dashed lines* indicating the error bars. From [11]
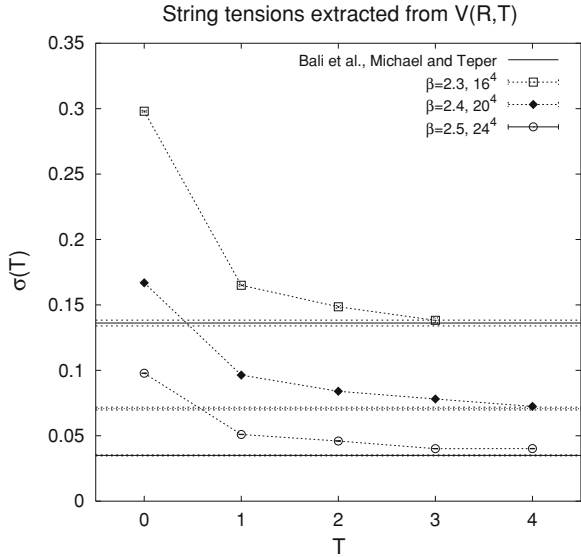
Fig. 9.3 $V(R, 4)$ at $\beta = 2.5$. The "without vortices" data points are obtained on lattices with vortices removed by the de Forcrand–D'Elia procedure. From [11]
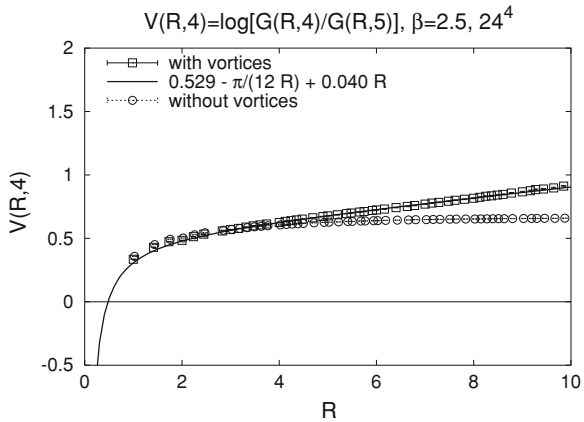
asymptotic string tensions vs. coupling β, where σ_{coul} is extracted from a fit to (9.29) both with and without the Lüscher term. At $\beta = 2.5$ we find that

$$\sigma_{coul} \approx 3\sigma. \tag{9.30}$$

The evidence, then, is that the Coulomb potential is linear but overconfines, in agreement with the Zwanziger inequality.[3]

---

[3] Another approach is to calculate $\langle K^{ab}(x, y; A) \rangle$ directly, via Monte Carlo simulations. This computationally more demanding procedure has been followed in [12] for the $SU(3)$ gauge group, with the result that the Coulomb potential is linearly confining, with a string tension that is greater (by a factor estimated at 1.6) than the asymptotic string tension.

Fig. 9.4 $V(R, 0)$ at $\beta = 2.5$, which is related (by 9.22) to the Coulomb energy. The *solid (dashed) line* is a fit to a linear potential with (without) the Lüscher term. From [11]
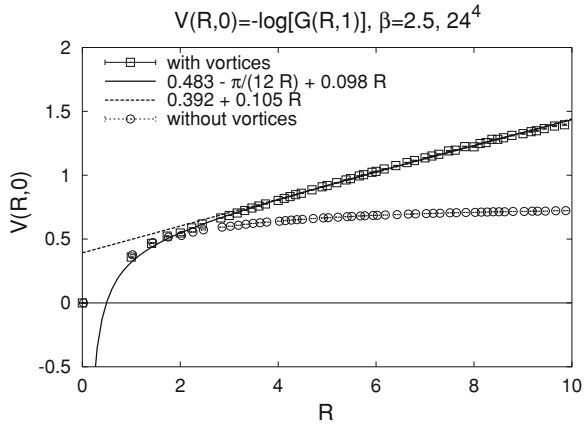
Fig. 9.5 The ratio $\sigma(0)/\sigma$, at various β, from fits which include ($b = \pi/12$) or do not include ($b = 0$) the Lüscher term. This ratio should equal the ratio of the Coulomb to asymptotic string tension. From [13]
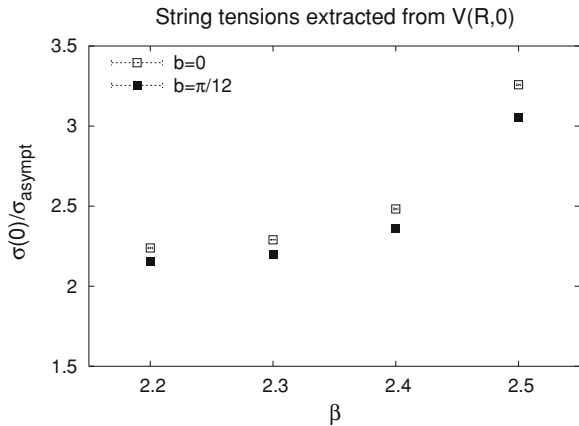
When vortices are removed, the string tensions of both the Coulomb potential and the static quark potential vanish, as shown by the lower data set in Figs. 9.3 and 9.4. This brings up two interesting questions. First, what property of Yang–Mills vacuum configurations is responsible for the confining Coulomb potential? Secondly, how is this property related to the existence of center vortices?

9.3 Faddeev–Popov Eigenvalue Density, and the Coulomb Self-Energy

In a confining theory, the energy of a color-nonsinglet state is infinite. In Coulomb gauge, such a state is, e.g.

$$\Psi^a[A; x] = q^a(x)\Psi_0[A]. \tag{9.31}$$

There is nothing intrinsically unphysical (or gauge non-invariant) about such a state, at least in an infinite volume. Gauge invariance in a Hamiltonian formulation amounts to satisfying the Gauss Law in one way or another. In Coulomb gauge only physical degrees of freedom appear, and Gauss's Law is solved explicitly, resulting in the H_{coul} term in the Hamiltonian. In $A_0 = 0$ gauge, Gauss's Law is an operator constraint on physical states, requiring that such states are invariant under infinitesimal gauge transformations. In QED, the state in $A_0 = 0$ gauge which corresponds to the state (9.31) in Coulomb gauge is

$$\Psi[A; x] = \exp\left[ie \int d^3z\, A(z) \cdot \nabla \frac{1}{4\pi|x - z|}\right] q(x) \Psi_0[A]. \qquad (9.32)$$

This conversion of the charged Coulomb gauge state to temporal gauge can be extended to non-abelian theories [14].

While the state corresponding to an isolated color non-singlet particle (9.31) qualifies as a physical state in Coulomb gauge, we would nevertheless expect that the self-energy of such a state is infinite in a confining theory, even with an ultraviolet cutoff. Before studying this Coulombic self-energy in Yang–Mills theory, it is instructive to do the calculation in an abelian theory, but in a way which will easily generalize to the non-abelian case. In the abelian theory we have

$$H = \int d^3x\, (\mathbf{E}_T^2 + (\nabla \times \mathbf{A})^2) + H_{coul}$$
$$H_{coul} = \frac{1}{2} \int d^3x\, d^3y\, \rho(x) K(x, y) \rho(y) \qquad (9.33)$$
$$K(x, y) = M_{xz}^{-1}(-\nabla^2)_z M_{zy}^{-1},$$

and also

$$\mathscr{E} = e^2 K(x, x), \qquad (9.34)$$

where $\mathscr{E}$ is the Coulomb self-energy of a static charge at point x. As in the non-abelian theory, the abelian F–P operator is obtained from the stationarity of the integral over $A_\mu A_\mu$, but in this case M is independent of the gauge field

$$M_{xy} = -\nabla^2 \delta(x - y). \qquad (9.35)$$

The eigenstates of the F–P operator, satisfying

$$M_{xy} \phi_y^{(n)} = \lambda_n \phi_x^{(n)}, \qquad (9.36)$$

are simply the plane wave states, with eigenvalues $\lambda_n = \mathbf{k}_n \cdot \mathbf{k}_n$ for wavenumber $\mathbf{k}_n$. These states are discrete in a finite volume, and integration over position become summation over position in a lattice regularization. We can write the Green's function in the form

$$G_{xy} = \left[M^{-1}\right]_{xy} = \sum_n \frac{\phi_x^{(n)} \phi_y^{(n)*}}{\lambda_n}. \tag{9.37}$$

With some simple manipulations, on a lattice of spatial volume $V = L^3$,

$$\begin{aligned}
\mathcal{E} &= \frac{e^2}{L^3} \left(M^{-1}(-\nabla^2)M^{-1}\right)_{xx} \\
&= \frac{e^2}{L^3} \sum_x \sum_{y_1 y_2} G_{xy_1} (-\nabla^2)_{y_1 y_2} G_{y_2 x} \\
&= \frac{e^2}{L^3} \sum_x \sum_{y_1 y_2} \sum_m \sum_n \frac{\phi_x^{(m)} \phi_{y_1}^{(m)*}}{\lambda_m} (-\nabla^2)_{y_1 y_2} \frac{\phi_{y_2}^{(n)} \phi_x^{(n)*}}{\lambda_n} \\
&= \frac{e^2}{L^3} \sum_{y_1 y_2} \sum_n \frac{\phi_{y_1}^{(n)*} (-\nabla^2)_{y_1 y_2} \phi_{y_2}^{(n)}}{\lambda_n^2},
\end{aligned} \tag{9.38}$$

we can express the Coulomb self-energy in the form

$$\mathcal{E} = \frac{e^2}{L^3} \sum_n \frac{F(\lambda_n)}{\lambda_n^2}, \tag{9.39}$$

where

$$F(\lambda_n) = \left(\phi^{(n)}\big|(-\nabla^2)\big|\phi^{(n)}\right). \tag{9.40}$$

Let $\rho(\lambda)$ denote the density of eigenvalues, scaled so that

$$\int d\lambda \, \rho(\lambda) = 1. \tag{9.41}$$

Then at large volumes the sum over eigenstates can be approximated by an integral

$$\mathcal{E} = e^2 \int d\lambda \, \frac{\rho(\lambda)F(\lambda)}{\lambda^2}. \tag{9.42}$$

In QED it is easy to show (see Eq. 6.56 with $D = 3$) that

$$\rho(\lambda) = \frac{\sqrt{\lambda}}{4\pi^2}, \quad F(\lambda) = \lambda. \tag{9.43}$$

The minimum eigenvalue in a volume of extension L is $\lambda_{min} \sim 1/L^2$, and the maximum eigenvalue $\lambda_{max} \sim 1$. Putting it all together, we find (in lattice units)

$$\begin{aligned}
\mathcal{E} &= e^2 \int d\lambda \frac{\rho(\lambda)F(\lambda)}{\lambda^2} \\
&\sim e^2 \left(1 - \frac{1}{L}\right).
\end{aligned} \tag{9.44}$$

In physical units (divide the lattice expression by lattice spacing a), this energy has the familiar UV divergence as $a \to 0$. The second term is a finite volume effect, which goes to zero in the infinite volume limit. However, finiteness at $L \to \infty$ at fixed a clearly depends on the small λ behavior of $\rho(\lambda)F(\lambda)$. If we had instead

$$\lim_{\lambda \to 0} \frac{\rho(\lambda)F(\lambda)}{\lambda} > 0, \tag{9.45}$$

then the Coulomb energy would be divergent in the large volume limit.

Typical configurations in the Gribov region are expected to approach the Gribov horizon in the infinite-volume limit. This is true even at the perturbative level, where $\lambda_{min} \sim 1/L^2$. But what counts for confinement is the density of eigenvalues $\rho(\lambda)$ near $\lambda = 0$, and the lack of smoothness of these near-zero eigenvalues, as measured by $F(\lambda)$. This is what determines whether the Coulomb confinement criterion (9.45) is satisfied.

In Yang–Mills theory the calculation of the self-energy of a static charge goes in much the same way, except that the F–P operator (9.6) depends on the gauge field, and for an arbitrary gauge group, with static charge in color representation r,

$$\mathscr{E}_r = \frac{g^2 C_r}{d_A} \mathscr{E}, \tag{9.46}$$

where C_r is the quadratic Casimir of representation r, d_A is the dimension of the adjoint representation, and

$$\mathscr{E} = \int d\lambda \left\langle \frac{\rho(\lambda)F(\lambda)}{\lambda^2} \right\rangle,$$
$$F(\lambda_n) = \left(\phi^{(n)} \big| (-\nabla^2) \big| \phi^{(n)} \right). \tag{9.47}$$

In order to test if $\mathscr{E}$ is infrared divergent, we can calculate $\rho(\lambda)$, $F(\lambda)$, numerically, on finite-size lattices, and extrapolate to infinite volume. This was done in [15], at a coupling $\beta = 2.1$ corresponding to a rather large lattice spacing, which allows us to probe fairly large volumes in physical units. The procedure is to generate field configurations by the lattice Monte–Carlo method, and fix each of them to Coulomb gauge (i.e. to a gauge copy in the Gribov region). Then, for each configuration, the first 200 eigenstates of the lattice F–P operator on each lattice time-slice are computed numerically, via the Arnoldi algorithm [16]. From these we calculate the average, over all gauge-fixed configurations, of $\rho(\lambda)$, $F(\lambda)$. The results, for lattice sizes 8^4 to 20^4 are shown in Fig. 9.6 for $\rho(\lambda)$, and Fig. 9.7 for $F(\lambda)$. From the scaling of this data with lattice size at small λ (cf. [15] for details), it is found that

$$\rho(\lambda) \sim \lambda^{0.25}, \quad F(\lambda) \sim \lambda^{0.4}, \tag{9.48}$$

and therefore

$$\int d\lambda \frac{\rho(\lambda)F(\lambda)}{\lambda^2} \to \infty \tag{9.49}$$

Fig. 9.6 The F–P eigenvalue density at $\beta = 2.1$, on $8^4 - 20^4$ lattice volumes. From [15]
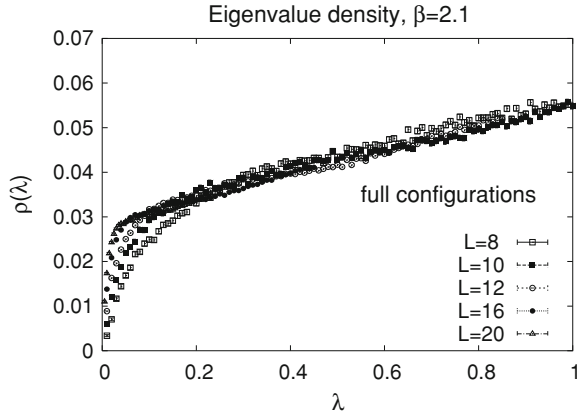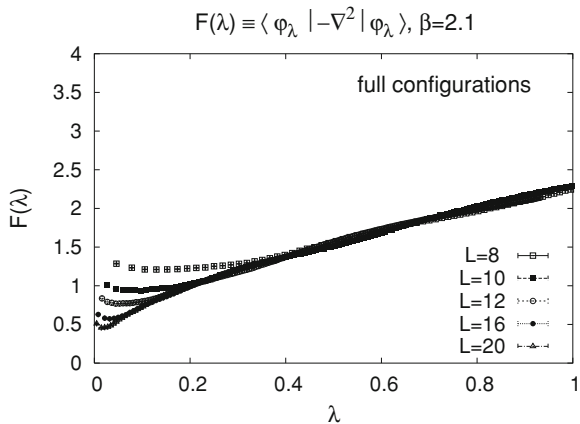

Fig. 9.7 $F(\lambda)$, the diagonal matrix element of $(-\nabla^2)$ in F–P eigenstates, plotted versus F–P eigenvalue. From [15]

is divergent in the infinite volume limit, due to the degree of divergence of the integrand at the lower ($\lambda = 0$) end of the integration region. This means that the Coulomb energy of an isolated color charge, in the state represented by Ψ^a in (9.31), is infinite, even when the usual ultraviolet divergence is regulated on the lattice.

9.3.1 The Role of Center Vortices

Next we can ask whether center vortices are related, in some way, to the divergent Coulomb energy of an isolated charge. The simplest thing to do is to repeat the above calculation for center-projected configurations ("vortex-only"), and for configurations with the vortices removed via the de Forcrand-D'Elia prescription [17]. Figures 9.8 and 9.9 show the results for $\rho(\lambda)$ and $F(\lambda)$ in the center-projected lattices. From a lattice size scaling analysis it is found that, at small λ,

Fig. 9.8 F–P eigenvalue
density in vortex-only
configurations. From [15]
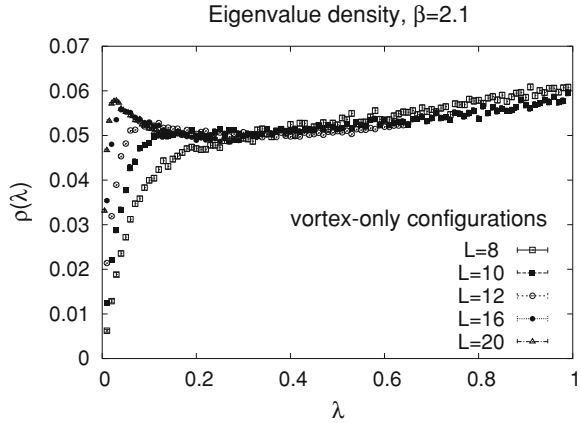
Fig. 9.9 $F(\lambda)$, the diagonal
matrix element of $(-\nabla^2)$ in
F–P eigenstates, for
vortex-only configurations.
From [15]
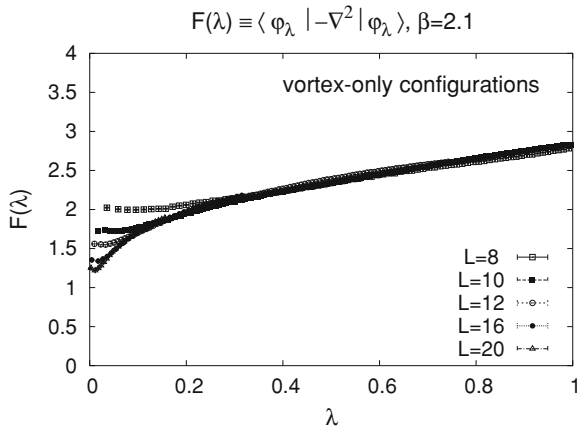
$$\rho(\lambda) \sim \lambda^{0\pm0.05}, \quad F(\lambda) \approx 1, \qquad (9.50)$$

and again the expression (9.49) is divergent.

   Figures 9.10 and 9.11 (left) show the corresponding results for $\rho(\lambda)$ and $F(\lambda)$ in vortex-removed configurations. This time there is a drastic change, as compared to the unmodified, and center-projected lattices; the density of states shows a series of peaks, and $F(\lambda)$ has a "band" structure. The peaks and bands correspond, in the sense that the eigenvalues belonging to the same peak of ρ also belong to the same band of F.

   It turns out that the structure of peaks in $\rho(\lambda)$ is essentially perturbative in origin. On a finite lattice, the density of states of the ordinary Laplacian $-\nabla^2$ (the $g = 0$ approximation to the covariant Laplacian) is a series of delta functions, each of which is singular at one of the eigenvalues of the Laplacian. The degeneracy of each eigenvalue is computable, and one finds that the numbers of eigenvalues in the peaks of $\rho(\lambda)$ precisely match the degeneracies of the low-lying eigenvalues

Fig. 9.10 F–P eigenvalue densities for vortex-removed configurations, on a 20^4 lattice volume. From [15]
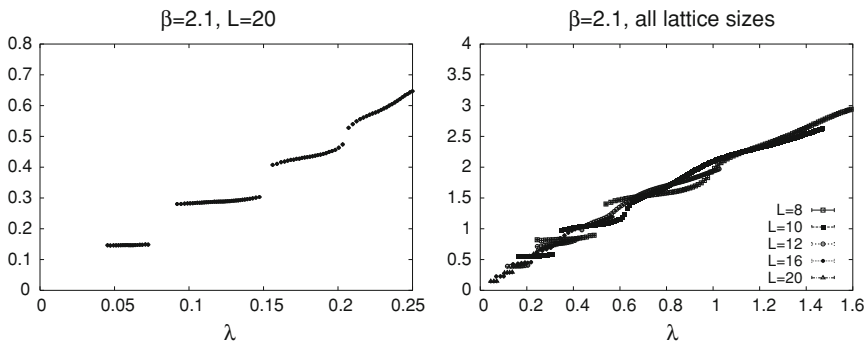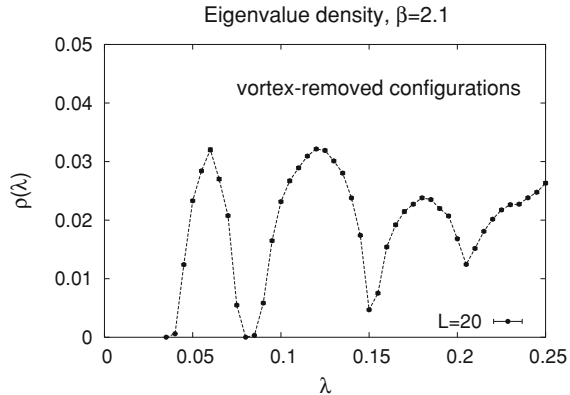
Fig. 9.11 $F(\lambda)$ versus λ in the vortex-removed configurations. We display results for the 20^4 volume alone (*left*), and a variety of lattice volumes (*right*). From [15]

of $-\nabla^2$. The effect of the vortex-removed lattice fields is to broaden the delta functions into peaks of finite width. The impression that the vortex-removed configurations are only a perturbation of the free theory is strengthened by the fact that the lowest-lying eigenvalues go to zero like $1/L^2$, where the L is the lattice extension, just as in a free theory. Moreover, if we plot our data for $F(\lambda)$ vs. λ at all volumes, as in Fig. 9.11 (right), we see that $F(\lambda) \propto \lambda$, again as in a free theory.

From all of this data, it appears that the enhancement of $\rho(\lambda)$ and $F(\lambda)$ near $\lambda = 0$, which is responsible for the infrared divergent Coulomb self-energy of an isolated charge, is correlated in some way with the presence of a center vortex ensemble.

9.4 Critique

Although all available numerical evidence favors a confining Coulomb force, and indeed this is a necessary condition for the existence of a confining static quark potential, there are also good reasons to doubt that the confining force between static

color charges is simply the Coulomb force in Coulomb gauge. In the first place, a necessary condition is not a sufficient condition, and there is also numerical evidence that the Coulomb force is confining at high temperatures, past the deconfinement transition [13]. We have also seen that even at zero temperature, the Coulomb string tension is several times larger than the asymptotic string tension. Finally, there is no reason to expect that the Coulombic color electric field, which depends on color charge density ρ according to $\mathbf{E} = \nabla(\nabla \cdot D)^{-1}\rho$, will be collimated into a flux tube, and this would lead to long-range dipole or van der Waals forces between color-singlet states. Nevertheless, it is possible that the confining Coulomb potential is an important element in understanding the formation of confining flux tubes, because this formation might be understood as a response which brings the Coulombic string tension down to the asymptotic value. The gluon chain model, to be discussed in Chap. 11, is one proposal along these lines.

References

1. Gribov, V.: Quantization of Nonabelian Gauge Theories. Nucl. Phys. B **139**, 1–19 (1978)
2. Zwanziger, D.: Vanishing of zero-momentum lattice gluon propagator and color confinement. Nucl. Phys. B **364**, 127–161 (1991)
3. Zwanziger, D.: Renormalization in the Coulomb gauge and order parameter for confinement in QCD. Nucl. Phys. B **518**, 237–272 (1998)
4. Neuberger, H.: Nonperturbative BRS Invariance and the Gribov Problem. Phys. Lett. B **183**, 337–340 (1987)
5. Srednicki, M.: Quantum Field Theory. Cambridge University Press, New York (2007)
6. Peskin, M., Schroeder, D.: An Introduction to Quantum Field Theory. Perseus Books, Cambridge (1995)
7. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, New York (2007)
8. Abers, E.S., Lee, B.W.: Gauge Theories. Phys. Rep. **9**, 1–141 (1973)
9. Christ, N.H., Lee, T.D.: Operator ordering and Feynman Rules in Gauge Theories. Phys. Rev. D **22**, 939–958 (1980)
10. Zwanziger, D.: No confinement without Coulomb confinement. Phys. Rev. Lett. **90**, 102001-1–102001-4 (2003) [arXiv:hep-th/0209105]
11. Greensite, J., Olejník, Š.: Coulomb energy, vortices, and confinement. Phys. Rev. D **67**, 094503-1–094503-6 (2003) [arXiv: hep-lat/0302018]
12. Voigt, A., Ilgenfritz, E.M., Muller-Preussker, M., Sternbeck, A.: The Effective Coulomb potential in SU(3) lattice Yang–Mills theory. Phys. Rev. D **78**, 014501-1–014501-11 (2008) [arXiv:0803.2307 [hep-lat]]
13. Greensite, J., Olejnik, S., Zwanziger, D.: Coulomb energy, remnant symmetry, and the phases of nonAbelian gauge theories. Phys. Rev. D **69**, 074506-1–074506-8 (2004) [arXiv:hep-lat/0401003]
14. Lavelle, M., McMullan, D.: Hadrons without strings. Phys. Lett. B **471**, 65–71 (1999) [arXiv: hep-ph/9910398]
15. Greensite, J., Olejnik, S., Zwanziger, D.: Center vortices and the Gribov horizon. JHEP **0505**, 070-1–070-38 (2005) [arXiv:hep-lat/0407032]
16. Lehoucq, R.B., Sorensen, D.C., Yang, C.: Arpack User's Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. SIAM, Philadelphia (1998)
17. de Forcrand, P., D'Elia, M.: On the relevance of center vortices to QCD. Phys. Rev. Lett. **82**, 4582–4585 (1999) [arXiv:hep-lat/9901020]

Chapter 10
Ghosts, Gluons, and Dyson–Schwinger Equations

Shortly after the recognition, in the 1970s, that the static quark potential ought to be linear with quark separation, there was an effort to show that the momentum space gluon propagator, in covariant gauges, goes as $1/k^4$ as $k \to 0$. The reason this was considered desirable is that if we naively evaluate the static quark potential from one-gluon exchange, then it is not hard to show that a $1/k^4$ limit at low momenta leads to a linear potential. Attempts to derive this $1/k^4$ behavior were based on trying to solve a truncated set of Dyson–Schwinger equations (DSEs), but because of various ambiguities, and the doubtful validity of the truncation, the effort was abandoned after a few years.

In the late 1990s a series of articles by Alkofer and co-workers [1–4] revived interest in the Dyson–Schwinger approach, and there has been much work on the topic since then (cf. the review article by Fischer [5]). In the new approach the gluon propagator was not found to go as $1/k^4$; in fact the prediction was that the gluon propagator actually vanishes in the $k \to 0$ limit, and it is instead the ghost propagator which should have an enhanced singularity (i.e. more singular than $1/k^2$) in the infrared. This solution of the Dyson–Schwinger equations is referred to as the *scaling solution*. The enhanced infrared singularity of the ghost propagator, found in the scaling solution, is necessary for the Kugo–Ojima confinement criterion (3.12) to be satisfied [6]; it is also the "horizon condition" advocated—together with the vanishing of the gluon propagator at $k = 0$—by Zwanziger [7, 8]. In recent years there has been a concerted effort to check these predictions by calculating ghost and gluon propagators numerically, on the lattice. Surprisingly, the lattice simulations appear to support the scaling solution in Landau gauge in $D = 2$ dimensions [9], but not in $D = 3$ and 4 dimensions [10–13]. Further investigation has shown that there exists another consistent solution of the Dyson–Schwinger equations, known as *decoupling solution* [14–16], and this solution does appear to be in harmony with the $D = 3$ and 4 dimensional lattice results.

The DSE effort is ongoing, and not everyone is in agreement that the current numerical results are decisive in favoring the decoupling solution. At the very least, we are seeing a very intriguing interplay between analytical arguments and

numerical simulations, and much has been learned, in the last few years, about the infrared behavior of ghost and gluon propagators on the lattice. In this chapter I will briefly review the main developments in this area.

10.1 Dyson–Schwinger Equations and the Scaling Solution

Dyson–Schwinger equations are derived in many standard textbooks on quantum field theory (see, e.g., [17]); they follow from the fact that the functional integral of a total derivative is zero. Consider any field theory involving a set of fields $\{\phi_k\}$. Then

$$
\begin{aligned}
0 &= \frac{1}{Z} \int \left\{ \prod_k D\phi_k \right\} \frac{\delta}{\delta\phi_i(x)} \exp\left[-S + \int d^D x \sum_i j_i(x)\phi_i(x) \right] \\
&= \left\langle -\frac{\delta S}{\delta\phi_i(x)} + j_i(x) \right\rangle.
\end{aligned}
\tag{10.1}
$$

Further differentiation with respect to the sources $\{j_k\}$ yields the Dyson–Schwinger equations.[1] As an example, the Dyson–Schwinger equation for the ghost propagator is derived from setting $\phi_i(x)$ equal to the antighost field $\bar{c}(x)$ and then differentiating with respect to the corresponding current source at point y, yields the Dyson–Schwinger equation for the ghost propagator

$$
\left\langle \frac{\delta S_{gf}}{\delta \bar{c}^a(x)} \bar{c}^b(y) \right\rangle = \delta^{ab} \delta^D(x-y),
\tag{10.2}
$$

where S_{gf} is the gauge-fixing part of the action of a non-abelian gauge theory quantized in a covariant gauge. With the help of some manipulations that can be found in Ref. [18], the above equation can be brought into a form which is easiest to indicate diagramatically, in Fig. 10.1. The left hand side of the illustrated equation is the inverse of the full (or "dressed") ghost propagator, the line on the right hand side is the inverse of the bare ghost propagator, and the rightmost loop diagram contains dressed ghost and gluon propagators connected by one bare and one dressed ghost-gluon vertex. This is one of the simplest of the Dyson–Schwinger equations, the other being the equation for the ghost-gluon vertex shown in Fig. 10.2. The important insight of Refs. [1–4] is that if the momentum on either of the external ghost lines of the loop diagram in Fig. 10.2 goes to zero, and the loop integral is finite, then the whole diagram vanishes. This means that in the infrared limit the dressed ghost-gluon vertex should resemble the bare vertex.

---

[1] In a gauge theory quantized in Landau or Coulomb gauge, where the functional integration is restricted to the Gribov region, the integration does not result in a boundary term, because the Fadeev-Popov determinant vanishes on the boundary (i.e. on the Gribov horizon).
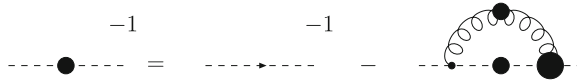
Fig. 10.1 Diagrammatic representation of the Dyson–Schwinger equation for the ghost propagator. Lines with (without) filled circles represent dressed (bare) propagators, the dashed lines are ghost propagators (or inverse propagators), and the wavy line is a dressed gluon propagator. The loop diagram contains one bare and one dressed ghost-gluon vertex. From Fischer [5]
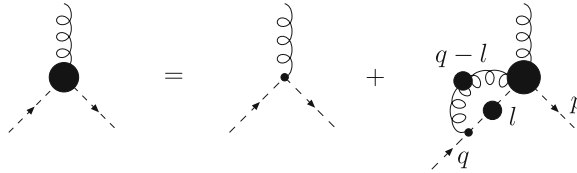

Fig. 10.2 Diagrammatic representation of the Dyson–Schwinger equation for the ghost-gluon vertex, with the same conventions as in Fig. 10.1. From Fischer [5]

The simplicity of the ghost-gluon vertex was then used in the Dyson–Schwinger equation for the ghost propagator, to derive an interesting relationship between the infrared behavior of the ghost and gluon propagators in the infrared limit. Let us write the full ghost and gluon propagators in the form

$$G^{ab}(q) = -\delta^{ab} \frac{J(q^2)}{q^2},$$
$$D^{ab}_{\mu\nu}(q) = \delta^{ab} \left(\delta_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) \frac{Z(q^2)}{q^2}, \tag{10.3}$$

and suppose that, as $q \to 0$

$$J(q^2) \propto (q^2)^{-\kappa_{ghost}}, \quad Z(q^2) \propto (q^2)^{\kappa_{gluon}}. \tag{10.4}$$

If the external momenta are very small in the Dyson–Schwinger equation for the ghost propagator, then the loop integral is dominated by momenta of the same magnitude as the external momenta, and one can replace the dressed propagators by their assumed power law behavior in the infrared. Matching powers of external momenta leads to a coupling of the ghost and gluon exponents

$$\kappa_{gluon} = 2\kappa_{ghost}. \tag{10.5}$$

This kind of analysis has been extended to other n-point functions. The result is as follows: Let $\Gamma^{(n,m)}(q^2)$ represent the dressing function for a one-particle irreducible Green's function with $2n$ external ghost lines and m external gluon lines, and all external momenta set equal to q. Then, as $q \to 0$,

$$\Gamma^{(n,m)}(q) \propto (q^2)^{(n-m)\kappa_{ghost}}. \tag{10.6}$$

The exponent κ_{ghost} has been determined from a further analysis of the gluon Dyson–Schwinger equation to have the value

$$\kappa_{ghost} = \frac{93 - \sqrt{1201}}{98} \approx 0.595. \tag{10.7}$$

For a derivation of these relations, cf. [19, 20]. With this value of κ_{ghost}, the scaling solution is making a clear prediction that the ghost propagator is more singular, in the infrared, than a simple $1/q^2$ pole, while the gluon propagator should actually vanish as $q \to 0$. This prediction for the ghost propagator is in perfect harmony with the Kugo–Ojima confinement criterion, which also requires an infrared singular behavior for the ghost dressing function. Given the infrared behavior of propagators and vertices, one can calculate the various running couplings associated with three gluon, four gluon, and ghost-gluon vertices, and these all run to finite non-zero values in the infrared.

One might wonder how a gluon propagator which vanishes at zero momentum, and running couplings which are finite in the infrared, will ever produce a confining static quark potential. Alkofer et al. [21] have addressed this question, and they argue that the quark-gluon vertex, unlike the ghost-gluon vertex, is singular in the infrared, and that this singularity overwhelms the suppression factor of the gluon propagator to produce a confining potential. Specifically, the prediction is that at low momentum

$$\Gamma^{qg}(q^2) \sim (q^2)^{-0.5 - \kappa_{ghost}}, \quad Z_f(q^2) \sim \text{const.}, \quad Z(q^2) \sim (q^2)^{2\kappa_{ghost}}, \tag{10.8}$$

where Γ^{qg} is the quark-gluon vertex, and Z_f is the dressing function for the quark propagator. This leads to a running coupling

$$\alpha^{qg}(q^2) \sim [\Gamma^{qg}(q^2)]^2 [Z_f(q^2)]^2 Z(q^2) \sim \frac{1}{q^2}. \tag{10.9}$$

Figure 10.3 shows the skeleton expansion of the four-quark Green's function. Already the first term has a $1/q^4$ behavior, leading to a linear potential. Note the contrast to earlier treatments, which tried to derive from the Schwinger–Dyson equations a gluon propagator behaving as $1/q^4$. In the scaling solution, it is the combination of the quark-gluon vertices with the (non-singular) gluon propagator that produces a $1/q^4$ behavior, which implies a linear potential.

One objection that could be made at this point is that the same kind of analysis seems inevitably to lead to long range interactions, i.e. van der Waals forces,
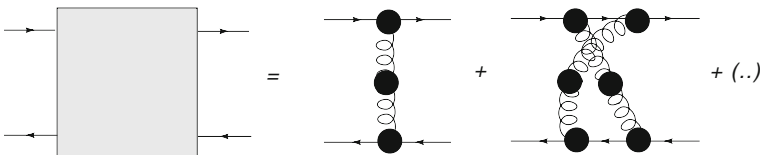

Fig. 10.3 The four quark one-particle irreducible Green's function. From Alkofer et al. [21]

between color singlet bound states. If the combination of quark-gluon vertices and gluon propagators leads to a long range interaction between color non-singlets, it would seem that the same combination would generate long-range forces also between color singlets. Such a long-range force would contradict the existence of a mass gap, which requires that there are only short-range forces between color singlet objects. It has not yet been shown that the scaling solution is consistent with this condition. A closely related issue is that it is not at all obvious, in this analysis, that a color electric flux tube forms between heavy, widely separated quarks. But however serious these problems may be, the scaling solution must first contend with a more direct challenge, coming from the lattice data.

10.2 Numerical Results for Ghost and Gluon Propagators

In view of the strong and remarkable predictions derived from the Dyson–Schwinger equations, there has naturally been an effort to check them via lattice Monte Carlo simulation. The simulations have been carried out independently, on very large lattices, by Cucchieri and Mendes [10–12], and by Bogolubsky et al. [13]. Both groups report similar results, and these appear to contradict the predictions of the scaling solution.

The gluon propagator on the lattice is defined in the usual way:

$$
\begin{aligned}
D_{\mu\nu}^{ab}(q) &= \langle \tilde{A}_\mu^a(k)\tilde{A}_\nu^b(-k) \rangle \\
&= \left(\delta_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) \delta^{ab} D(q^2),
\end{aligned}
\tag{10.10}
$$

where $q_\mu = (2/a)\sin(\pi k_\mu/L)$ with $k_\mu \in [-\frac{1}{2}L, \frac{1}{2}L]$. The variable $\tilde{A}_\mu^a(k)$ is the Fourier transform, in a finite lattice volume, of the gauge field defined from link variables

$$
A_\mu^a(x) = \frac{1}{2iag}\left(U_\mu(x) - U_\mu^\dagger(x) \right).
\tag{10.11}
$$

The lattice ghost propagator, in momentum space, is given by

$$
\begin{aligned}
G^{ab}(q) &= a^2 \left\langle \sum_{x,y} e^{-2\pi i k\cdot(x-y)/L} \left(M^{-1}\right)_{xy}^{ab} \right\rangle \\
&= \delta^{ab} G(q),
\end{aligned}
\tag{10.12}
$$

where M_{xy}^{ab} is the lattice F-P operator in Landau gauge. The calculation of $D(q)$ and $G(q)$ in lattice Landau gauge has been carried out for both the SU(2) gauge group, in $D = 3$ and $D = 4$ dimensions, and also for SU(3) gauge theory, in $D = 4$ dimensions. The results in all cases are quite similar; here we will just display the results of Bogolubsky et al. [13] for the SU(3) gauge group in $D = 4$ dimensions,

Fig. 10.4 The lattice gluon propagator in momentum space, $D(q)$ vs. q^2, at $\beta = 5.7$ and lattice volumes up to 96^4. From Bogolubsky et al. [13]
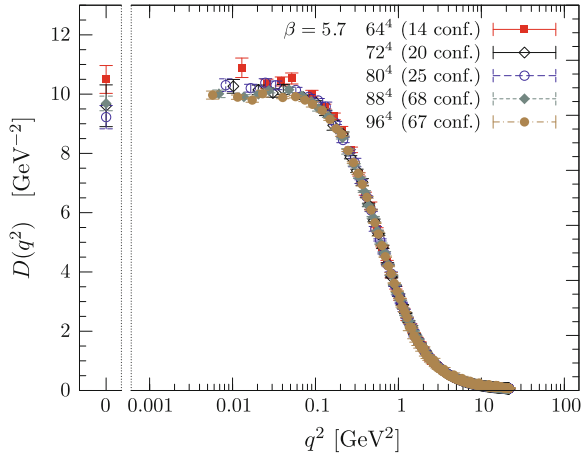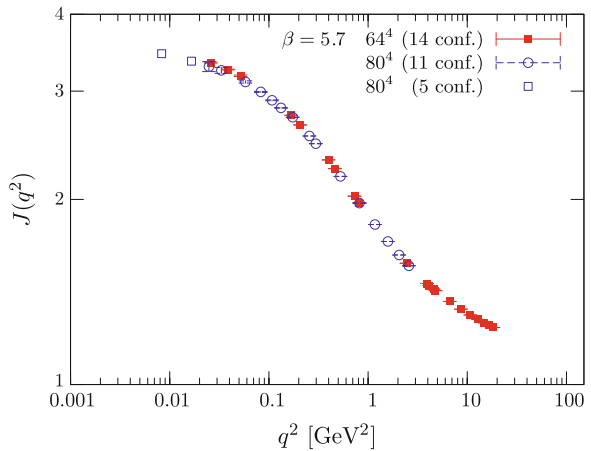

Fig. 10.5 The ghost dressing function in momentum space, $J(q)$ vs. q^2, at $\beta = 5.7$ and lattice volumes of 64^4 and 80^4. From Bogolubsky et al. [13]


at Wilson lattice coupling $\beta = 5.7$, with lattice volumes up to 96^4. Using the appropriate lattice spacing $a = 0.17$ fm at $\beta = 5.7$, all quantities can be expressed in physical units.

The result for the gluon propagator is shown in Fig. 10.4. There is no sign that the gluon propagator vanishes at zero momentum; rather it appears to reach a plateau. The ghost dressing function $J(q^2) = q^2 G(q^2)$ is displayed in Fig. 10.5, and in this case there is no evident power-law singularity on the log-log plot. Rather, the indication is that $J(q^2)$ also reaches a plateau. Putting these results together, one can calculate a renormalization-group invariant running coupling associated with the ghost-gluon vertex

$$\alpha_s(q^2) = \frac{g^2}{4\pi} J^2(q^2) Z(q^2), \qquad (10.13)$$

Fig. 10.6 The running coupling $\alpha_s(q^2)$ associated with the ghost-gluon vertex. From Bogolubsky et al. [13]
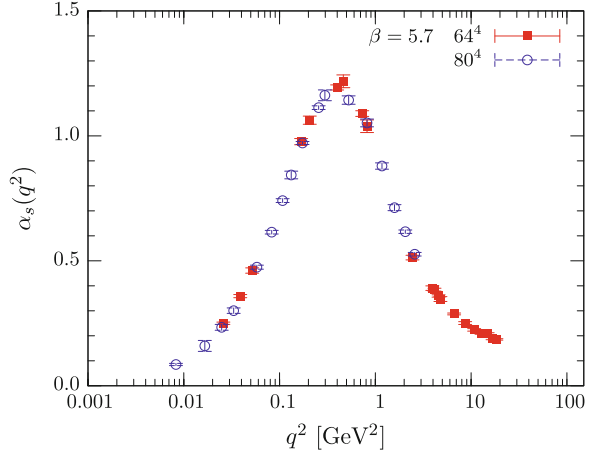
and this coupling is seen in Fig. 10.6 to run to zero at small momentum. The results displayed in all three of these figures clearly contradict the predictions of the scaling solution.

To make matters a little more confusing, it turns out that the situation is different in $D = 2$ dimensions in Landau gauge. Lattice Monte Carlo simulations have shown that in $D = 2$ dimensions the Landau gauge gluon propagator *does* vanish, and the ghost dressing function *is* singular, in the $q^2 \to 0$ limit. In $D = 2$ dimensions the scaling solution predicts, for the exponents of the ghost and gluon dressing functions,

$$\kappa_{gluon} = 2\kappa_{ghost} + 1 \quad \text{and} \quad \kappa_{ghost} = 1/5, \tag{10.14}$$

and these predictions appear to be quite consistent with the $D = 2$ lattice Monte Carlo data obtained by Maas [9].

10.3 The Decoupling Solution

The discouraging results of the numerical simulations in $D > 2$ dimensions have led to a re-examination of the Dyson–Schwinger equations. If the lattice Monte Carlo data is not misleading us, then the scaling solution cannot be the only consistent solution of the Dyson–Schwinger equations. There must be another solution, more in line with the lattice results.

Such a solution was soon found [14–16], and it is known as the *decoupling solution*. The term "decoupling" refers to the fact that the exponent κ_{ghost} of the ghost dressing function is no longer tied to the exponent κ_{gluon} of the gluon dressing function, as expressed in Eqs. 10.5. Instead, the decoupling solution has

$$J(p^2) \sim \text{const.}, \quad Z(p^2) \sim p^2, \tag{10.15}$$

or $\kappa_{ghost} = 0$, $\kappa_{gluon} = -1$ in the infrared, which clearly violates (10.15). This is not a very exotic solution; the condition on the ghost dressing function just says that the ghost propagator behaves as a simple $1/p^2$ pole near $p = 0$, while $Z(p^2) \sim p^2$ means that the gluon propagator is finite at $p = 0$, as is the case for any massive propagator. To see what is happening, we have to go back to the Dyson–Schwinger equation for the ghost propagator, with couplings and renormalization constants defined with respect to some renormalization point μ. The equation displayed diagrammatically in Fig. 10.1 is, in its full glory [22],

$$J^{-1}(p^2, \mu^2) = \tilde{Z}_3 - \tilde{Z}_1 \frac{g^2(\mu^2)N}{(2\pi)^4} \int d^4 q \Gamma_\mu^{(2,1)(0)}(p, q; \mu)\left(\delta_{\mu v} - \frac{k_\mu k_v}{k^2}\right)$$
$$\times \frac{Z(k^2; \mu^2)J(q^2; \mu^2)}{k^2 q^2} \Gamma_v^{(2,1)}(q, p; \mu). \tag{10.16}$$

In this equation, $k \equiv p + q$, $\tilde{Z}_3$ and $\tilde{Z}_1$ are ghost and ghost-gluon vertex renormalization constants, $\Gamma_\mu^{(2,1)(0)}$ is the bare ghost-gluon vertex, and $\Gamma_\mu^{(2,1)}$ is the dressed ghost-gluon vertex. Schematically,

$$J^{-1}(p^2, \mu^2) = \tilde{Z}_3 - \tilde{Z}_1 g^2(\mu^2) I(p; \mu). \tag{10.17}$$

If the ghost dressing function is infrared singular, so that the left hand side of this equation vanishes at $p = 0$, then it must be that the two terms on the right hand side exactly cancel at $p^2 = 0$. But a low-momentum behavior of this type is not the only possibility. One can obtain perfectly valid solutions of the Dyson–Schwinger equations, and these are the decoupling solutions, in which there is not perfect cancellation on the right hand side at $p^2 = 0$ [14–16]. Another way to put it is that the choice of $J^{-1}(0, \mu^2)$ is a boundary condition for the solution of the Dyson–Schwinger equations [22]. If we subtract, from (10.17), the same equation at $p^2 = 0$, then we have

$$J^{-1}(p^2, \mu^2) = J^{-1}(0, \mu^2) - \tilde{Z}_1 g^2(\mu^2)(I(p; \mu) - I(0; \mu)). \tag{10.18}$$

The choice $J^{-1}(0, \mu^2) = 0$ generates the scaling solution, and $J^{-1}(0, \mu^2) > 0$ gives the decoupling solution. From this point of view, the scaling solution is only the endpoint of a family of solutions of the Dyson–Schwinger equations.

   If the decoupling solution is the correct choice in $D > 2$ dimensions, it means that the Kugo-Ojima criterion cannot be correct (as already argued in Chap. 3), and it also means that we cannot hope to understand the confining static quark potential from a few combinations of vertices and propagators in a covariant gauge. The decoupling solution does, however, allow for the possibility that the gluon propagator violates positivity, so that one doesn't expect to see color charged gluons showing up in the asymptotic spectrum. This is a condition of color confinement, which, as explained in Chap. 3, also holds in a Higgs theory, and so it is a little different from the phenomenon of confinement as it is understood in this book.

There remains, as of this writing, a considerable ongoing effort in the Dyson–Schwinger approach. There has been speculation that perhaps the scaling solution would be realized if it were possible to gauge fix to the Fundamental Modular region, and some studies appear to show the gluon propagator becoming a little smaller in the deep infrared, when the gauge fixing algorithm is adjusted to select copies with smaller values of R than would be obtained at random [23, 24].[2] There is also an argument, by Fischer, Maas, and Pawlowski [22], that the decoupling solution is at odds with global BRST symmetry in the confining phase. But even if the scaling solution were obtained in some very special subset of gauge copies, there is a question of whether that is really important, if the vast majority of gauge copies do not support the scaling solution, and yet *any* gauge copy exhibits the confinement property. If the scaling solution were crucial to understanding confinement in covariant gauges, one would be hard pressed to explain the existence of the confinement property in the vast majority of Landau gauge Gribov copies. In any case, lattice results so far favor the decoupling solution. Whether this is the final answer, time will tell.

References

1. von Smekal, L., Alkofer, R., Hauck, A.: The Infrared behavior of gluon and ghost propagators in Landau gauge QCD. Phys. Rev. Lett. **79**, 3591–3594 (1997), hep-ph/9705242
2. von Smekal, L., Alkofer, R., Hauck, A.: A solution to coupled Dyson-Schwinger equations for gluons and ghosts in Landau gauge. Ann. Phys. **267**, 1–60 (1998) [Erratum-ibid. **269**, 182 (1998)], hep-ph/9707327
3. Fischer, C.S., Alkofer, R., Reinhardt, H.: The Elusiveness of infrared critical exponents in Landau gauge Yang-Mills theories. Phys. Rev. D **65**, 094008-1–094008-19 (2002), hep-ph/0202195
4. Fischer, C.S., Alkofer, R.: Infrared exponents and running coupling of SU(N) Yang-Mills theories. Phys. Lett. B **536**, 177–184 (2002), hep-ph/0202202
5. Fischer, C.S.: J. Infrared properties of QCD from Dyson-Schwinger equations. Phys. G **32**, R253–R291 (2006) [arXiv:hep-ph/0605173]
6. Kugo, T.: The Universal renormalization factors Z(1)/Z(3) and color confinement condition in nonAbelian gauge theory. arXiv:hep-th/9511033
7. Zwanziger, D.: Vanishing of zero momentum lattice gluon propagator and color confinement. Nucl. Phys. B **364**, 127–161 (1991)
8. Zwanziger, D.: Renormalizability of the critical limit of lattice gauge theory by BRS invariance. Nucl. Phys. B **399**, 477–513 (1993)
9. Maas, A.: Two and three-point Green's functions in two-dimensional Landau-gauge Yang-Mills theory. Phys. Rev. D **75**, 116004-1–116004-14 (2007) [arXiv:0704.0722 [hep-lat]]
10. Cucchieri, A., Mendes, T.: What's up with IR gluon and ghost propagators in Landau gauge? A puzzling answer from huge lattices. PoS **LAT2007**, 297-1–297-7 (2007) [arXiv:0710.0412 [hep-lat]]

---

[2] However, as noted in [25], fixing to the Fundamental Modular region ought to make the ghost propagator less, rather than more, singular.

11. Cucchieri, A., Mendes, T.: Constraints on the IR behavior of the ghost propagator in Yang-Mills theories. Phys. Rev. D **78**, 094503-1–094503-6 (2008) [arXiv:0804.2371 [hep-lat]]
12. Cucchieri, A., Mendes, T.: Constraints on the IR behavior of the gluon propagator in Yang-Mills theories. Phys. Rev. Lett. **100**, 241601-1–241601-4 (2008) [arXiv:0712.3517 [hep-lat]]
13. Bogolubsky, I.L., Ilgenfritz, E.M., Muller-Preussker, M., Sternbeck, A.: Lattice gluodynamics computation of Landau gauge Green's functions in the deep infrared. Phys. Lett. B **676**, 69–73 (2009) [arXiv:0901.0736 [hep-lat]]
14. Boucaud, Ph., Leroy, J.P., Yaouanc, A.L., Micheli, J., Pene, O., Rodriguez-Quintero, J.: On the IR behaviour of the Landau-gauge ghost propagator. JHEP **0806**, 012-1–012-17 (2008) [arXiv:0801.2721 [hep-ph]]
15. Aguilar, A.C., Binosi, D., Papavassiliou, J.: Gluon and ghost propagators in the Landau gauge: Deriving lattice results from Schwinger-Dyson equations. Phys. Rev. D **78**, 025010-1–025010-6 (2008) [arXiv:0802.1870 [hep-ph]]
16. Dudal, D., Gracey, J.A., Sorella, S.P., Vandersickel, N., Verschelde, H.: The Landau gauge gluon and ghost propagator in the refined Gribov-Zwanziger Framework in 3 dimensions. Phys. Rev. D **78**, 065047-1–065047-22 (2008) [arXiv:0806.4348 [hep-th]]
17. Itzykson, C., Zuber, J.-B.: Quantum Field Theory. McGraw-Hill, New York (1980)
18. Fischer, C.S.: Nonperturbative propagators, running coupling and dynamical mass generation in ghost-anti-ghost symmetric gauges in QCD. arXiv:hep-ph/0304233
19. Lerche C., von Smekal, L.: On the infrared exponent for gluon and ghost propagation in Landau gauge QCD. Phys. Rev. D **65**, 125006-1–125006-22 (2002) [arXiv:hep-ph/0202194]
20. Zwanziger, D.: Nonperturbative Landau gauge and infrared critical exponents in QCD. Phys. Rev. D **65**, 094039-1–094039-13 (2002) [arXiv:hep-th/0109224]
21. Alkofer, R., Fischer, C.S., Llanes-Estrada, F.J., Schwenzer, K.: The Quark-gluon vertex in Landau gauge QCD: Its role in dynamical chiral symmetry breaking and quark confinement. Ann. Phys. **324**, 106–172 (2009) [arXiv:0804.3042 [hep-ph]]
22. Fischer, C.S., Maas, A., Pawlowski, J.M.: On the infrared behavior of Landau gauge Yang-Mills theory. Ann. Phys. **324**, 2408–2437 (2009) [arXiv:0810.1987 [hep-ph]]
23. Bornyakov, V.G., Mitrjushkin V.K., Muller-Preussker, M.: SU(2) lattice gluon propagator: Continuum limit, finite-volume effects and infrared mass scale. Phys. Rev. D **81**, 054503-1–054503-10 (2010) [arXiv:0912.4475 [hep-lat]]
24. Maas, A.: More on Gribov copies and propagators in Landau-gauge Yang-Mills theory. Phys. Rev. D **79**, 014505-1–014505-23 (2009) [arXiv:0808.3047 [hep-lat]]
25. Cucchieri, A.: Gribov copies in the minimal Landau gauge: The Influence on gluon and ghost propagators. Nucl. Phys. B **508**, 353–370 (1997) [arXiv:hep-lat/9705005]

Chapter 11
Large-N, Planar Diagrams, and the Gluon-Chain Model

The Lagrangian of a pure $SU(N)$ gauge theory

$$S = \int d^4x \frac{1}{2} \text{Tr}[F_{\mu\nu}F_{\mu\nu}]$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \tag{11.1}$$

appears to contain only one parameter, namely the coupling g, and one would think that the Feynman diagrammatic expansion for Green's functions, n-point correlation functions, Wilson loops, and other observables can only be organized in powers of g. But it was pointed out by 't Hooft [1], in 1974, that the number of colors N of the $SU(N)$ gauge group can also be regarded as a parameter, and the Feynman diagram series can be reorganized as a double expansion in powers of $1/N^2$, and in powers of the *'t Hooft coupling*

$$\lambda \equiv g^2 N, \tag{11.2}$$

so that the diagrammatic series for any observable has the form

$$\langle O \rangle = N^p \sum_{n=0}^{\infty} N^{-2n} T_n(\lambda), \tag{11.3}$$

where each of the $T_n(\lambda)$ is an (infinite) series in powers of λ. If we take the $N \to \infty$ limit with the 't Hooft coupling λ held fixed, then only the leading term T_0 is important, and this term can usually be represented by *planar diagrams*, which are Feynman diagrams that can be drawn on a sheet of paper, such that no gluon line crosses over any other gluon line.

The planar contribution seems to dominate in $SU(N)$ theories for many observables, even down to $N = 2$. Lattice calculations for, e.g, the string tension σ, from $N = 2$ to $N = 5$, indicate that the tension is almost independent of N when the 't Hooft coupling λ is held fixed, while the dimensionless ratio of glueball mass m to $\sqrt{\sigma}$ varies only a little as N is increased from 2 to 5 [2]. The great hope is that

certain remarkable simplifications, that are inherent in the $N = \infty$ limit, may provide new insights into (or perhaps even allow a solution of) non-abelian gauge theories in this limit.

11.1 Double-Line Notation and Factorization

The rationale for the double expansion in λ and $1/N$ begins with the observation that the gluon field is a matrix-valued field, i.e. $A_\mu^{ab} = A_\mu^c L_c^{ab}$, where the $\{L_c\}$ are $SU(N)$ group generators, and a, b are matrix (or "color") indices which run from 1 to N. The gluon propagator is then

$$\langle A_\mu^{ab}(x) A_\nu^{cd}(y) \rangle_0 = \left(\delta^{ad} \delta^{bc} - \frac{1}{N} \delta^{ab} \delta^{cd} \right) D_{\mu\nu}(x - y). \qquad (11.4)$$

When N is very large, the term proportional to $1/N$ on the right hand side is unimportant, and can be dropped. Alternatively, we may consider the gauge group $U(N)$ as an approximation to $SU(N)$. This gauge group is essentially a product $U(1) \times SU(N)$, since one gluon decouples from the others, and for most gauge-invariant observables that single gluon, out N^2 gluons, makes a negligible contribution when N is very large. For the $U(N)$ gauge group the $1/N$ term in Eq. (11.4) is absent from the start, and the color indices at the beginning of a gluon propagator match the indices at the end. This fact motivates 't Hooft's double-line notation for the gluon propagator, in which each line carries a matrix index. Then, because the interaction terms are just products of matrices (i.e. matrix-valued fields), there is a simple flow of color indices through the interaction vertices which can also be represented in double line notation, as shown in Fig. 11.1. Using this notation, it is quite straightforward to count the power of N associated with any particular diagram. The general rule is that any single line which forms a

Fig. 11.1 't Hooft double-line notation for propagators and vertices. Only the index dependence is shown; vector indices are suppressed
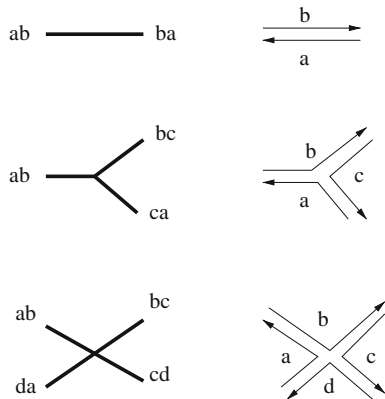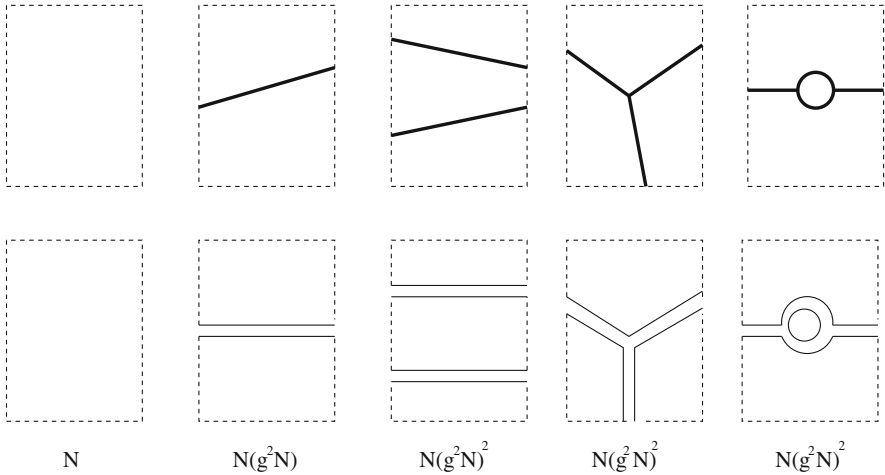
Fig. 11.2 Planar Feynman diagram contributions to a rectangular Wilson loop, in the usual (*upper row*) and double-line (*lower row*) notation

closed loop contributes a factor of N, because it represents a color index which is summed from 1 to N.

As an important example, let us consider the perturbative evaluation of a Wilson loop, assuming that the loop is small enough to be amenable (because of asymptotic freedom) to such methods. Fig. 11.2 shows a few low-order diagrams in both ordinary and double-line notation, and for purposes of index counting the dashed lines representing the Wilson loop are to be counted as part of the index loops. The first, zeroth-order diagram is just equal to the number N of colors. The second diagram, with one gluon exchange, has two powers of the gauge coupling, and two index loops which contribute a factor of N^2. So the diagram is of order g^2N^2, which we write as $N \lambda$. The next three diagrams are fourth order in the coupling, but have three index loops, and are therefore of order $g^4 N^3$, equal to $N \lambda^2$. The pattern (an overall factor of N, times a power of λ) is becoming clear, and it can be shown that the contribution of any planar diagram (no crossing gluon lines) is of order $N \lambda^p$. These diagrams contribute to the term $NT_0(\lambda)$ in the series Eq. (11.3).

On the other hand, consider the leading non-planar contribution shown in Fig. 11.3. This diagram is fourth-order in the coupling, but has only a single index loop, so it is of order $g^4 N$ which can be expressed as $(1/N^2)(N \lambda^2)$. That is the leading contribution to the term NT_1 in the large-N series expansion of the Wilson loop. In general if we take the simultaneous limits $N \to \infty$ and $g \to 0$ in such a way that the 't Hooft coupling $\lambda = g^2 N$ is non-zero and finite in the limit, then all non-planar contributions are negligible, i.e. down by a power of at least $1/N^2$ as compared to the planar contribution.[1] A similar reorganization can be carried out

---

[1] The power of N associated with any graph actually depends on certain topological properties of the graph, cf. [1] and [3].

Fig. 11.3 The leading
non-planar contribution to a
rectangular Wilson loop, due
to two-gluon exchange, in the
usual and in double-line
notation

$$N\,g^4$$
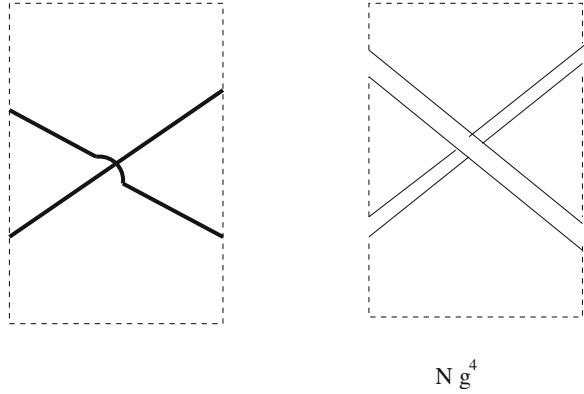
in the strong-coupling expansion of lattice gauge theory; in that case the double expansion is in powers of $1/N^2$ and $1/\lambda$.

If it would be possible to sum up all the planar Feynman diagrams for large Wilson loops at $N = \infty$, and demonstrate confinement in that limit, that would be a spectacular advance. This has so far proved impossible. The hope that it *might* be possible is based in part on a striking property of the $N = \infty$ limit which seems quite unlike the case for $N = 2$ or $N = 3$, and this is the property of *factorization*. Let O_1 and O_2 be any two gauge-invariant operators. The factorization property says that, to leading order in $1/N^2$,

$$\langle O_1 O_2 \rangle = \langle O_1 \rangle \langle O_2 \rangle. \tag{11.5}$$

For example, let O_1 and O_2 be Wilson loops. Feynman diagrams which spoil the equality of the left and right hand sides of (11.5) are diagrams in which gluon lines connect the two loops, and these diagrams are always down by (at least) a factor of $1/N^2$, compared to planar diagrams in which there are no gluon lines connecting the loops.

An immediate consequence of the factorization property is Casimir scaling. Let $U(C)$ be a Wilson loop holonomy. Then the trace of the holonomy in group representation r (the group character $\chi_r[U(C)]$) can be always be expressed in terms of products of characters in the fundamental defining representation. To leading order in $1/N^2$

$$\chi_r[U(C)] = \left(\chi_F[U(C)]\right)^n \left(\chi_F^*[U(C)]\right)^{\overline{n}} + \text{sub-leading terms.} \tag{11.6}$$

Taking the expectation value of both sides, and applying the factorization property (11.5), it follows that as $N \to \infty$ the string tension σ_r of loops in representation r is an integer multiple of the string tension σ_F in the fundamental representation, i.e.

$$\sigma_r = (n + \overline{n})\sigma_F. \tag{11.7}$$

In particular, the string tension of the adjoint loop is simply twice the string tension of the fundamental loop $\sigma_{adj} = 2\sigma_{fund}$, Now in the large-N limit, the quadratic Casimir of representation r is $C_r = (n + \bar{n})N/2$. Then (11.7) is simply the statement of Casimir scaling.

In the $N = \infty$ limit, the Casimir scaling regime extends from the confinement scale out to infinity, and this is in contrast to any theory at finite N, where the asymptotic string tensions depend only on N-ality. The reason for this difference is that string breaking is a $1/N^2$-suppressed process [4], and therefore the loop size where string-breaking occurs, and the asymptotic string tension is obtained, runs off to infinity as $N \to \infty$.[2]

The factorization property also implies, since $\langle O^2 \rangle = \langle O \rangle^2$ and by definition $\Delta O^2 = \langle O^2 \rangle - \langle O \rangle^2$, that the rms deviation ΔO of any gauge-invariant observable from its mean value vanishes in the large-N limit. This fact has an astonishing consequence. If we would imagine carrying out a numerical simulation of $SU(N)$ gauge theory at some enormous value of N, such that all non-leading powers of N could be neglected, then a gauge-invariant observable, evaluated in any set of thermalized configurations, would have the same value in each configuration. This means that it is not necessary to average over many configurations to obtain the expectation value of a gauge-invariant observable; a single configuration (if it is the right configuration) would be sufficient. This led Witten [6] to the idea of the large-N *master field*; i.e. there should exist a field configuration A_μ^{master} such that for any gauge invariant observable O, up to $1/N^2$ corrections,

$$\langle O \rangle = O[A_\mu^{master}], \tag{11.8}$$

where the right hand side is simply the operator evaluated at the master field configuration. This is very different from what happens at finite N. For example, In a Monte Carlo evaluation of a large Wilson loop $\text{Tr}[U(C)]/N$ in $SU(2)$ or $SU(3)$ gauge theory, the trace of the loop is usually not small, when evaluated in a typical thermalized configuration, and the trace divided by N varies wildly between ± 1. The expectation value of this quantity becomes exponentially small, for large loops, due to near-perfect cancellations between the large positive and negative values obtained in different configurations. In contrast, as $N \to \infty$, the same quantity is tiny, for large loops, in *any* thermalized configuration. For loops evaluated in such configurations, or in the master field configuration, the wild fluctuations between positive and negative values, and the near-perfect cancellation, really takes place among the eigenvalues of the loop holonomy $U(C)$, which nearly cancel among themselves when taking the trace.[3]

It would be wonderful to find a master field for $N = \infty$ gauge theory which has such marvelous properties. Unfortunately, although the master field is known

---

[2] For a discussion of the center vortex theory at large but finite N, cf. [5].

[3] In $D = 2$ dimensions and $N = \infty$, the spectral density of the eigenvalues of $U(C)$ was worked out by Durhuus and Olesen [7].

explicitly for certain models with matrix-valued fields in very low dimensions [8], where knowing the master field amounts to knowing its eigenvalue spectrum, we do not know the master field for $U(\infty)$ gauge theories in the interesting cases of $D = 3$ and $D = 4$ dimensions. Despite this fact, one has the feeling that the large-N limit is still a huge simplification in non-abelian gauge theory, and a number of important advances in the field are based on it. Perhaps first among these is the AdS/CFT correspondence, to be discussed in Chap. 13 Another intriguing idea is known as *large-N reduction*, in which the *D*-dimensional theory is replaced by a zero-dimensional theory; i.e. a theory defined at a single point. The original proposal, due to Eguchi and Kawai [9] was to replace the periodic lattice in *D*-dimensional lattice gauge theory by a single (periodic) hypercube, and this turns out to work at strong, but not at weak couplings. More sophisticated versions of large-N reduction were introduced soon after [10, 11], but it has been shown recently that these models also encounter technical difficulties at weak couplings [12], at least for pure gauge theories. In any case, it does not seem to be any easier analytically to solve the reduced models than to solve the full theory.

A further feature of the large-N expansion, pointed out by 't Hooft in his seminal work [1], is more suggestive than quantitative. Suppose that the diagrammatic expansion would somehow be relevant to large Wilson loops, and that the expansion is dominated by very high order planar diagrams, as indicated schematically (in double-line notation) in Fig. 11.4. This high-order diagram *looks* like a discretized surface, and if we would take a time-slice, the particle content is a static quark and antiquark source, with a series of virtual gluons between the sources. This picture suggests a discretized string of some kind, leading to the idea that the color-electric flux tube between a quark–antiquark pair might be thought of as a chain of gluons, bound together by attractive forces. This is known as the
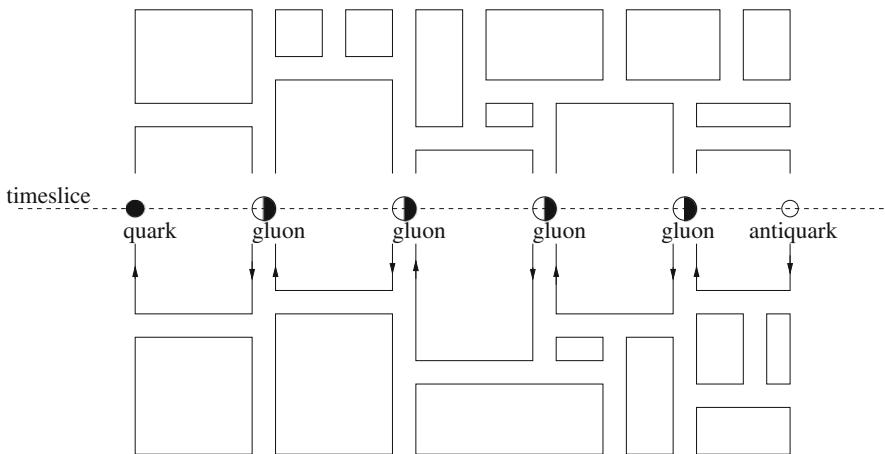
Fig. 11.4 The gluon chain as a time slice of a planar diagram, which is shown here in double-line notation for propagators introduced by 't Hooft. A *solid hemisphere* indicates a quark color index, and an *open hemisphere* an antiquark color index

gluon-chain model (cf. [13] and references therein, and ref. [14]), and it has a number of attractive features. In the next section we will discuss some tests of this idea, as formulated in Coulomb gauge.

11.2 The Gluon Chain Model

We have seen in Chap. 9 that the color Coulomb potential is linear in pure $SU(2)$ gauge theory, but there are (at least) two serious objections to claiming that the Coulomb potential explains confinement. First, the Coulomb string tension seems to be about three times larger than the asymptotic string tension of the static quark potential. Secondly, a Coulombic electric field, which depends on the charge distribution ρ in this way:

$$\mathbf{E}^L = \nabla \frac{1}{\nabla \cdot D} \rho, \tag{11.9}$$

will result in long-range Coulomb dipole fields, and long-range van der Waals forces among hadrons. This problem is really generic to any model of confinement based on ladder diagrams or (dressed) one-gluon exchange, as found, e.g., in the Dyson-Schwinger approach discussed in the previous chapter.

The underlying problem is that Coulomb confinement, while providing a linear potential, does not involve the collimation of the color-electric field into a flux tube. This raises the question, given the existence of a linear Coulomb potential, of how a flux tube forms in Coulomb gauge.

Let us recall that the Coulomb potential is simply the R-dependent part of the energy expectation value in the quark–antiquark physical state

$$\Psi_{q\bar{q}} = \bar{q}^a(0)q^a(R)\Psi_0. \tag{11.10}$$

But there is no reason that this state should be the minimal energy state out of all states containing a static quark–antiquark pair of separation R, and it should be possible to construct lower-energy states with the help of additional gluon operators, creating "constituent" gluons. Schematically, we are looking for lower energy states of the form

$$\Psi'_{q\bar{q}} = \bar{q}^a(0)\{c_0\delta^{ab} + c_1A^{ab} + c_2A^{ac}A^{cb} + \cdots\}q^b(R)\Psi_0. \tag{11.11}$$

States of this form may also have a very much smaller, or non-existent, long range dipole field.[4]

---

[4] In fact, a Wilson line running between the quark–antiquark pair has exactly this form, if one expands the path-ordered exponential in powers of the A_μ-field, and has no associated color electric dipole field.
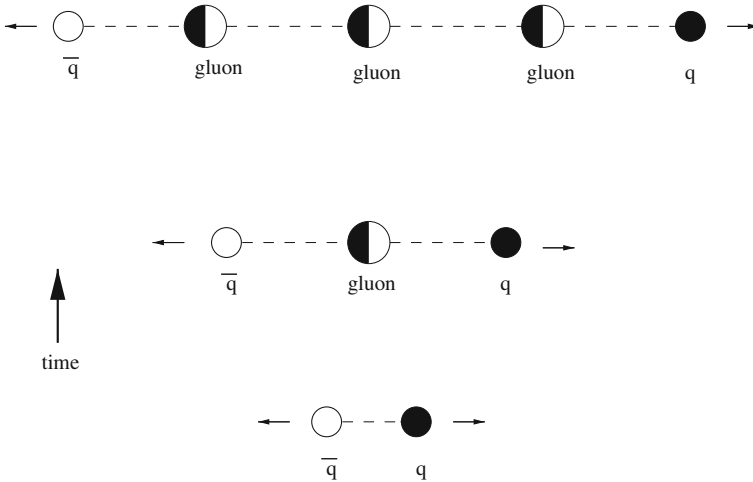
Fig. 11.5 Formation of a gluon chain, as a quark–antiquark pair separates

The gluon-chain model is the idea that a quark–antiquark pair, as they separate, pull out a chain of gluons (Fig. 11.5), with nearest neighbors in the chain bound by attractive forces between them. Thus the minimal energy state $\Psi'_{q\bar{q}}$ will be dominated, for large quark–antiquark separations, by components with many constituent gluons. In its original form, it was supposed that as the quark and antiquark separate, the field energy increases faster than linearly, and at some point it is energetically favorable to insert a gluon between the quark–antiquark charges to reduce the effective charge separation. In fact, we have seen that the Coulomb energy grows only linearly, but it may still be the case that the energy of physical states with widely separated color charges can be lowered by the inclusion of constituent gluons.

One of the strongest motivations for the gluon chain model is that a gluon chain can be regarded as a time-slice of a high-order planar Feynman diagram, as we have already seen in Fig. 11.4. The gluon chain also has string-like properties (e.g. a Lüscher term) due to fluctuations in the position of its gluon constituents, it features Casimir scaling at large N, and also has the right N-ality properties due to string-breaking [13]. For example, at large N there are two chains between heavy sources in the adjoint representation, giving rise to twice the string tension as in the fundamental representation, which is the correct Casimir ratio at large N. Interaction between the two chains is $1/N^2$ suppressed, but at finite N the chains can interact and rearrange themselves (Fig. 11.6) into two "gluelumps," having negligible energy dependence on the quark separation R. In general, for a quark and antiquark in representation r satisfying (11.6), there will be $n + \bar{n}$ chains running between the static sources. Neglecting string-breaking or other interactions between the chains, this leads to the large-N Casimir scaling result $\sigma_r = (n + \bar{n})\sigma_F$, and when string-breaking is taken into account, the number of remaining chains depends only on the N-ality of representation r.

Fig. 11.6 Adjoint string-breaking in the gluon chain model. Two gluons in separate chains (**I**) scatter by a contact interaction, resulting in the re-arrangement of color indices indicated in (**II**). This corresponds to chains starting and ending on the same heavy source. The chains then contract down to smaller "gluelumps" (**III**)
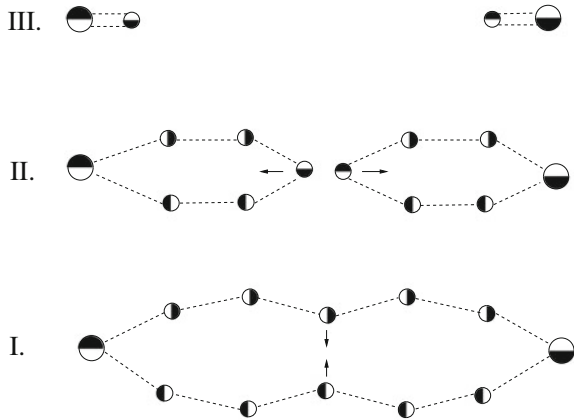
11.2.1 Numerical Investigation in Coulomb Gauge

In Euclidean lattice gauge theory, the transfer matrix $T = \exp[-Ha]$ is the Euclidean time version of the Minkowski space time evolution operator, where a is lattice spacing in the time direction. It is useful to define the rescaled transfer matrix

$$T = \exp[-(H - E_0)a], \tag{11.12}$$

where E_0 is the vacuum energy. To compute the static quark potential of a quark–antiquark pair separated by a distance R, one would ideally diagonalize the transfer matrix in the infinite-dimensional subspace of states which contain a single massive quark, and a single massive antiquark, located at sites $\mathbf{x}$ and $\mathbf{y}$ with $R = |\mathbf{x} - \mathbf{y}|$. The minimal energy eigenstate of the transfer matrix, in this subspace, is the state with the largest eigenvalue λ_{max} of T, and the energy of the state in lattice units is given by

$$V(R) = -\log(\lambda_{max}). \tag{11.13}$$

In practice, we seek to diagonalize T in the subspace spanned by a finite number of $q\bar{q}$ states of the form

$$|k\rangle = \bar{q}^a(\mathbf{x})Q_k^{ab}q^b(\mathbf{y})|\Psi_0\rangle, \quad k = 1, 2, \ldots, M, \tag{11.14}$$

where the Q_k operators are functionals of the lattice link variables. This program has recently been carried out in ref. [15], with a set of $M = 6$ operators involving up to two gluon fields on the lattice, creating up to two "constituent" gluons in the region between the quark and antiquark sources. The Q_k operators specify the set of $\{|k\rangle\}$ states, and from these an orthonormal set $\{|\varphi_k\rangle\}$ is constructed by the Gram-Schmidt procedure. The transfer matrix is diagonalized in the subspace spanned by the $\{|\varphi_k\rangle\}$, and the lowest energy eigenstate in this subspace

Fig. 11.7 The color Cou-
lomb potential $V_{coul}(R)$, the
"gluon-chain" potential $V(R)$
derived from the minimal-
energy variational state, and
the "true" static quark
potential $V_{true}(R)$ obtained by
standard methods. Results are
shown at lattice coupling
$\beta = 2.4$. Continuous lines are
from a fit of data points to
$\sigma R - \pi/(12 R) + c$
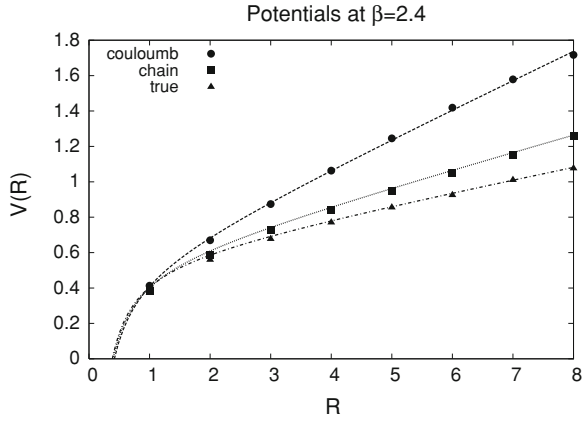
$$|\psi(R)\rangle = \sum_{k=1}^{6} a_k(R)|\varphi_k\rangle \tag{11.15}$$

is determined. The corresponding energy depends on a variational parameter in the
Q_k operators, which is then selected to minimize the energy of the lowest energy
state. Since φ_1 is just the vacuum (no gluon) state, φ_2 is a one-gluon state, and the
remainder are two-gluon states, it follows that a_1^2 is the fraction of the norm of
$\psi(R)$ due to the zero (constituent) gluon component $|\varphi_1\rangle = |1\rangle$, a_2^2 is the fraction of
the norm $\psi(R)$ from the one-gluon component $|\varphi_2\rangle \propto |2\rangle$, and $1 - a_1^2 - a_2^2$ is the
fraction of the norm due to the remaining two-gluon component states.[5]
The energy of this state is

$$V_{chain}(R) = -\log(\lambda_{max}), \tag{11.16}$$

which can be compared to the Coulomb (self + interaction) energy

$$V_C(R) = -\log(T_{11}), \tag{11.17}$$

where $T_{11} = \langle \varphi_1|T|\varphi_1 \rangle = \langle 1|T|1 \rangle$ is the matrix element of T in the zero constit-
uent gluon state. Note that $V_C(R) = V(R, 0)$, in the notation of Sect. 9.2 in Chap. 9

The resulting potentials, for $SU(2)$ lattice gauge theory at $\beta = 2.4$ are shown in
Fig. 11.7, where the static quark potential, computed by standard methods, is also
displayed. Two features worth emphasizing in this figure is that $V_{chain}(R)$ remains
linear, and its slope is much closer to that of the asymptotic string tension, as
compared to the Coulomb string tension.

Fig. 11.8 is a comparison of the zero, one, and two constituent-gluon content of
the minimal energy state at each R, where by "content" is meant the fraction of the
norm of the minimal energy state. The zero gluon content obviously dominates at

---

[5] For the explicit form of the $\{Q_k\}$ cf. [15].

Fig. 11.8 Zero, one, two-gluon content (fraction of the norm of the variational state) versus quark separation R in fermis, at $\beta = 2.2, 2.3, 2.4$
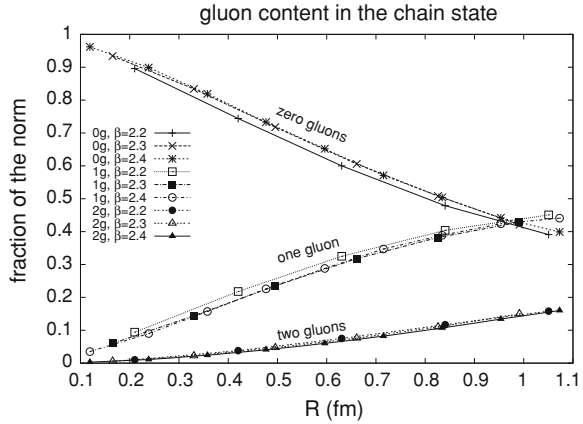
Fig. 11.9 Sensitivity of the Coulomb potential $V_{coul}(R)$ (*solid symbols*), and insensitivity of the chain potential $V_{chain}(R)$ (*open symbols*), to lattice volume. Data is for the gauge coupling $\beta = 2.4$, and lattice volumes $L^4 = 12^4, 16^4, 22^4$. Quark–antiquark separation R is in lattice units
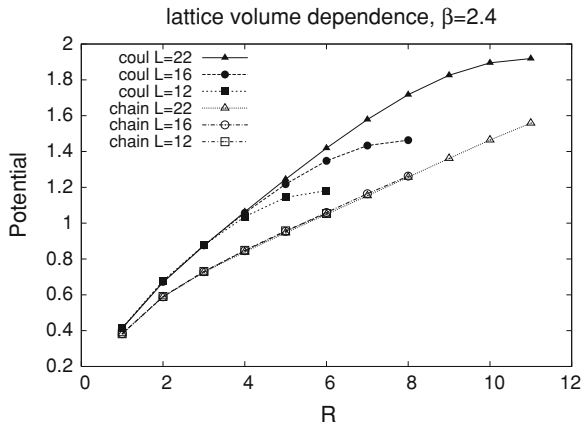
small $q\bar{q}$ separation. This is of course expected, because placing a gluon between the quark and antiquark comes with a price in kinetic energy. Around $R \approx 1$ fm, however, the zero and one-gluon contributions to the minimal energy state are about equal. It is important to note also that the gluon content versus R in physical units is almost coupling independent, which serves as a check of the whole procedure.

As already mentioned, the color Coulomb field is not expected to be collimated into a flux tube. This means that there should be strong sensitivity to lattice volume, on a lattice of spatial extension L, for quark–antiquark separations close to $R = L/2$. The reason is that for separations of that size, the finite volume cuts off a region where the field energy is still significant. If, instead, the field energy were collimated into a flux tube of diameter d, then so long as $L > d$, there would not be a similar sensitivity to the finite volume. Fig. 11.9 shows the Coulomb energy $V_C(R)$, and the multi-gluon variational state energy $V_{chain}(R)$ computed at $\beta = 2.4$, at a variety of lattice sizes. The Coulomb energy shows a strong sensitivity to

lattice extension, with the potential flattening out as $R \to L/2$. In contrast, $V_{chain}(R)$ appears to be rather insensitive to the lattice size. This suggests that the chain state has no long-range dipole field, or at least that the long-range field is greatly suppressed relative to the color dipole field of the zero-gluon state. It is possible that, in the multi-gluon state, we are beginning to see the formation of a color electric flux tube.

I have already noted that one of the main motivations of the gluon chain model is the resemblance of high-order planar Feynman diagrams to a discretized surface, and the time-slices of such diagrams to a discretized string; i.e. a chain. There is an approach due to Bardakci and Thorn [16], which explicitly formulates the planar Feynman diagrams of matrix ϕ^3 field theory as the diagrams of a discretized string theory in light-cone coordinates. This approach has been extended to gauge theories by Thorn in ref. [17]. Since this is a string formulation of any planar diagram, including low-order Feynman diagrams, the discretized string formulation is not sufficient in itself to establish the existence of confinement, the QCD flux tube, or the gluon-chain nature of the QCD flux tube. On the other hand, the "stringy" properties of a gluon-chain seem rather natural in this approach; perhaps gluon-chain excitations will emerge in some new approximation scheme, specifically adapted to the Bardakci-Thorn reformulation of planar QCD.

References

1. 't Hooft, G.: A planar diagram theory for strong interactions. Nucl. Phys. B **72**, 461–473 (1974)
2. Lucini, B., Teper, M.L.: SU(N) gauge theories in four-dimensions: Exploring the approach to N–infinity. JHEP **0106**, 050-1–050-36 (2001) [arXiv:hep-lat/0103027]
3. Coleman, S.: Aspects of symmetry. Cambridge University Press, Cambridge (1995)
4. Greensite, J., Halpern, M.B.: Suppression of color screening at large N. Phys Rev D **27**, 2545–2547 (1983)
5. Greensite, J., Olejnik, S.: K string tensions and center vortices at large N. JHEP **0209**, 039-1–039-22 (2002) [arXiv:hep-lat/0209088]
6. Witten, E.: The 1/N expansion in atomic and particle physics. In: 't Hooft, G. (ed.) Recent Developments in Gauge Theories 1979 Cargese Lectures, Plenum, New York (1980)
7. Durhuus, B., Olesen, P.: The spectral density for two-dimensional continuum QCD. Nucl. Phys. B **184**, 461–475 (1981)
8. Brezin, E., Itzykson, C., Parisi, G., Zuber, J. B.: Planar diagrams. Commun. Math. Phys. **59**, 35–51 (1978)
9. Eguchi, T., Kawai, H.: Reduction of dynamical degrees of freedom in the large N gauge theory. Phys. Rev. Lett. **48**, 1063–1066 (1982)
10. Bhanot, G., Heller, U.M., Neuberger, H.: The quenched Eguchi-Kawai model. Phys. Lett. B **113**, 47–50 (1982)
11. Gonzalez-Arroyo, A., Okawa, M.: A twisted model for large N lattice gauge theory. Phys. Lett. B **120**, 174–178 (1983)
12. Bringoltz, B., Sharpe, S.R.: Breakdown of large-N reduction in the quenched Eguchi-Kawai model. PoS **LATTICE2008**, 055-1–055-7 (2008) [arXiv:0810.1239 [hep-lat]]
13. Greensite, J., Thorn, C.B.: Gluon chain model of the confining force. JHEP **0202**, 014-1–014-23 (2002) [arXiv:hep-ph/0112326]

14. Tiktopoulos, G.: Gluon chains. Phys. Lett. B **66**, 271–275 (1977)
15. Greensite, J., Olejnik, S.: Constituent gluon content of the static quark-antiquark state in Coulomb gauge. Phys. Rev. D **79**, 114501-1–114501-7 (2009) [arXiv:0901.0199 [hep-lat]]
16. Bardakci, K., Thorn, C.B.: A world sheet description of large-N quantum field theory. Nucl. Phys. B **626**, 287–306 (2002) [arXiv:hep-th/0110301]
17. Thorn, C.B.: A world sheet description of planar Yang-Mills theory. Nucl. Phys. B **637**, 272–292 (2002) (Erratum-ibid B **648**, 457 (2003) [arXiv:hep-th/0203167])

Chapter 12
The Vacuum Wavefunctional

The static quark potential arises from vacuum fluctuations of the gauge fields; this is clear from the fact that the potential is extracted from the vacuum expectation value of a Wilson loop. In a Euclidean functional integral, the orientation of a rectangular $R \times T$ loop is obviously irrelevant to the expectation value, and in particular it can be oriented in a plane at a fixed time. In that case

$$W(R, T) = \langle \Psi_0 | \text{Tr}[U(C)] | \Psi_0 \rangle, \tag{12.1}$$

where, in a physical gauge such as temporal ($A_0 = 0$) or Coulomb gauge, the ground state $\Psi_0[A]$ of the Hamiltonian can be expressed in path-integral form

$$\Psi_0[A(\mathbf{x}, 0)] = \frac{1}{\sqrt{Z}} \int DA(\mathbf{x}, t<0) \; \delta(F[A]) \det(M) \exp\left[- \int_{-\infty}^{0} dt \; L[A(\mathbf{x}, t)]\right]. \tag{12.2}$$

Here $L[A]$ is the Lagrangian, and $F[A] = 0$ is the gauge-fixing condition with $\det(M)$ the corresponding Faddeev–Popov determinant. In Coulomb gauge an additional restriction of gauge fields to the Gribov region (or some subspace thereof) is understood, as discussed in Chap. 9. A physical gauge is a gauge with a ghost-free Hamiltonian (which requires, among other things, no time-derivatives in the gauge-fixing condition) and in such gauges Ψ_0 satisfies[1]

$$H\Psi_0[A_k(\mathbf{x})] = E_0 \Psi_0[A_k(\mathbf{x})]. \tag{12.3}$$

Then, given (12.1), it seems possible that something could be learned about confinement if we knew the form of the vacuum wavefunctional $\Psi_0[A]$.

---

[1] As already mentioned in Chapter 9, it may still be useful in a physical gauge to introduce Faddeev-Popov ghosts in the path integral, but the essential point is that in a physical gauge these ghosts do not propagate in time.

12.1 Dimensional Reduction

It was proposed long ago [1, 2] that for gauge-field configurations in which the field strength is weak, and which are "slowly varying" in the sense that gauge-invariant observables vary slowly with respect to the characteristic scale (set by Λ_{QCD}), the vacuum wavefunctional in $3 + 1$ dimensions has the simple form

$$\Psi_0[A] \approx \Psi_0^{eff}[A] = \mathcal{N} \exp\left[-\frac{1}{2}\mu \int d^3x \, \mathrm{Tr}[F_{ij}^2]\right]. \tag{12.4}$$

If this form for $\Psi_0^{eff}[A]$ correctly describes the properties of the vacuum at large scales, then the vacuum has the property of *dimensional reduction*, in the sense that the computation of the string tension of a large Wilson loop, in $D = 4$ Euclidean dimensions, reduces to the same calculation in $D = 3$ dimensions with an effective coupling $1/2g_{eff}^2 = \mu$, i.e.

$$\begin{aligned}
W(R, T) &= \langle \mathrm{Tr}[U(C)]\rangle_{D=4} \\
&= \langle \Psi_0 | \mathrm{Tr}[U(C)] | \Psi_0 \rangle \\
&\approx \int DA(\mathbf{x}) \, \delta(F[A]) \det(M) \mathrm{Tr}[U(C)] \\
&\quad \times \exp\left[-\mu \int d^3x \, \mathrm{Tr}[F_{ij}^2]\right] \\
&= \langle \mathrm{Tr}[U(C)]\rangle_{D=3}.
\end{aligned} \tag{12.5}$$

It was also suggested [3] that the vacuum state of Yang–Mills theory in $2 + 1$ dimensions has, at large scales, the same form (12.4) as in $3 + 1$ dimensions, but in one less spatial dimension. If this is so then dimensional reduction can be applied one more time, and we would find, for large Wilson loops

$$\begin{aligned}
W(R, T) &\approx \langle \mathrm{Tr}[U(C)]\rangle_{D=2} \\
&= \exp[-\sigma \mathrm{Area}(C)],
\end{aligned} \tag{12.6}$$

where the last step follows from the fact that Wilson loops have an area-law falloff in $D = 2$ dimensions.

The proposal (12.4) has been supported in two ways. First of all, it can be shown to be true in strong-coupling lattice gauge theory. In temporal gauge ($U_0 = 1$), the ground state of the lattice Hamiltonian has the form

$$\Psi_0[U] = \mathcal{N} \exp[-R(U)], \tag{12.7}$$

and $R(U)$ can be computed systematically in a power series in $1/g^2$ (the procedure is worked out in ref. [2]). At leading order, in $SU(N)$ gauge theory

$$\Psi_0[U] = \mathcal{N} \exp\left[\frac{N}{g^4(N-1)} \sum_P \mathrm{Tr}U(P) + \text{c.c.}\right], \tag{12.8}$$

where the sum is over spacelike plaquettes P, and $U(P)$ is the product of links around plaquette P. We see that the exponent has the form of the Wilson action in one lower dimension, and therefore at strong-coupling the dimensional reduction form of the vacuum wavefunctional is seen to be correct.

The second source of support comes from direct measurement, by numerical methods, of the wavefunctional evaluated on a finite set of configurations [4–6]. This method begins with the identity, in temporal gauge

$$|\Psi_0[U_i(\mathbf{x})]|^2 = \frac{1}{Z} \int DU_i \; \delta[U_i(\mathbf{x}) - U_i(\mathbf{x}, 0)] e^{-S}. \tag{12.9}$$

Suppose, in $d + 1$ dimensions, we pick any set of Md-dimensional lattice configurations $\{U^m_i(\mathbf{x}), \; m = 1 \ldots M\}$ that may be of interest, and define

$$\widetilde{Z} = \sum_{m=1}^{M} \int DU \; \delta[U^m(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}. \tag{12.10}$$

Then $\widetilde{Z}$ defines a lattice system in which the configurations at time $t = 0$ are constrained to be member of the given set of M configurations. The statistical probability associated with the k-th configuration in the set is given by

$$\text{Prob}[U^k] = \frac{1}{\widetilde{Z}} \int DU(\mathbf{x}, t) \; \delta[U^k(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}. \tag{12.11}$$

In a Monte Carlo simulation of the lattice system described by $\widetilde{Z}$,

$$\text{Prob}[U^k] = \lim_{N_{tot} \to \infty} \frac{N_k}{N_{tot}}, \tag{12.12}$$

where N_{tot} is the number of times the configuration on the $t = 0$ timeslice is updated by the Metropolis algorithm, and N_k is the number of times the k-th configuration, out of the set $\{U^m_i(\mathbf{x})\}$, $m = 1, 2 \ldots M$, is generated by the updates. In practice, links at $t \neq 0$ are calculated by some efficient algorithm, such as heat bath for the $SU(2)$ gauge group, while the configuration at $t = 0$ is updated by selecting one of the M configurations at random, and accepting or rejecting the choice via the Metropolis algorithm. With this technique we can also compute the ratio

$$\begin{aligned}
\frac{|\Psi_0[U^{m_1}]|^2}{|\Psi_0[U^{m_2}]|^2} &= \frac{\int DU \delta[U^{m_1}(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}}{\int DU \delta[U^{m_2}(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}} \\
&= \frac{\int DU \delta[U^{m_1}(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}/\widetilde{Z}}{\int DU \delta[U^{m_2}(\mathbf{x}) - U(\mathbf{x}, 0)] e^{-S}/\widetilde{Z}} \\
&= \lim_{N_{tot} \to \infty} \frac{N_{m_1}}{N_{m_2}},
\end{aligned} \tag{12.13}$$

and in this way it is possible to measure, via a slightly non-standard Monte Carlo simulation, the relative values of the Yang–Mills vacuum wavefunctional in any

given set of configurations. The computation was carried out long ago for non-abelian constant configurations (the $U_i(\mathbf{x})$ are independent of $\mathbf{x}$, and $[U_i, U_j] \neq 0$ if $i \neq j$), and also for slowly varying abelian plane wave configurations, with wavelengths comparable to the lattice size. In every case, the results were consistent with the dimensional reduction form

$$\Psi_0[U] = \exp\left[\mu(\beta) \sum_P \mathrm{Tr} U(P) + \mathrm{c.c.}\right], \tag{12.14}$$

where $\mu(\beta)$ in lattice units varies with β in such a way that, when converted to physical units, μ is constant at any β. This behavior was found in both three and four dimensions [4–6].

The above results also apply to Coulomb gauge, since the Coulomb gauge wavefunctional Ψ_0^{coul} is identical to the temporal gauge wavefunctional Ψ_0

$$\Psi_0^{coul}[A^{tr}] = \Psi_0[A^{tr}], \tag{12.15}$$

when the latter is restricted to transverse gauge fields satisfying the Coulomb gauge condition $\nabla \cdot A^{tr} = 0$ [7]. This means, in particular, that if the temporal gauge vacuum wavefunctional is, at large scales, gaussian in the field strength F_{ij}, then the Coulomb gauge vacuum wavefunctional is *also* gaussian in the field strengths. There is, however, a suggestion by Szczepaniak and Swanson [8, 9] that the vacuum wavefunctional in Coulomb gauge is gaussian in the gauge fields instead, i.e.

$$\Psi_0^{coul}[A] = \mathcal{N} \exp\left[-\frac{1}{2}\int \frac{d^3k}{(2\pi)^3} A_i^a(k)\omega(k)A_i^a(-k)\right], \tag{12.16}$$

with some kernel $\omega(k)$ to be determined by minimization of the vacuum energy. Reinhardt and Feuchter [10, 11] have suggested that this expression should be multiplied by a factor of the inverse square-root of the F–P determinant $M^{-1/2}$. The advantage of these proposals is that a wavefunctional which is gaussian in the gauge fields, rather than field strengths, lends itself to analytical methods, and an infrared enhancement of the ghost propagator, and a confining Coulomb potential, have been obtained. The disadvantage is that it is difficult to see how such a wavefunctional could ever lead to an area law for spatial Wilson loops,[2] and the proposal also seems to be at odds with the existing evidence, obtained in temporal gauge, for the dimensional reduction form at large scales.

However, the dimensional reduction form of the wavefunctional cannot be exact. If it were, then Green's functions in four-dimensional Yang–Mills theory would be identical to those in three-dimensional Yang–Mills theory, which certainly cannot be true at short distances (it would violate perturbation theory, at distance scales where perturbative calculations are valid). Moreover, while

---

[2] For a recent attempt to overcome this difficulty by inclusion of monopole fields, cf. [12].

reduction to two dimensions seems fine if the Wilson loop is in the fundamental representation, it is clearly wrong for Wilson loops in the adjoint representation, where Wilson loops must ultimately fall off with a perimeter law, due to color screening. The problem is that in two-dimensional Yang–Mills theory the string tension is proportional to the quadratic Casimir of the color representation of the Wilson loop, and therefore we have Casimir scaling rather than N-ality dependence asymptotically. For a more accurate picture of the Yang–Mills vacuum state, it is certainly necessary to go beyond the dimensional reduction form. Most of the recent progress in this area has been in $D = 2 + 1$ dimensions.

12.2 Temporal Gauge Vacuum State in 2 + 1 Dimensions

The Yang–Mills Hamiltonian in $D = 2 + 1$ dimensions, in temporal gauge, is

$$H = \int d^2x \left\{ -\frac{1}{2} \frac{\delta^2}{\delta A_k^a(x)^2} + \frac{1}{2} B^a(x)^2 \right\}, \tag{12.17}$$

where $B = F_{12}$, and we will consider, for simplicity, the $SU(2)$ gauge group. A special feature of $A_0 = 0$ gauge is that the Gauss law $D_k^{ab} E_k^b = 0$, which normally follows from stationarity of the action with respect to variations in A_0, must be imposed in the quantum theory as a constraint on physical states, i.e. $D_k^{ab} E_k^b \Psi = 0$. In Schrodinger representation this constraint is

$$\left(\delta^{ac} \partial_k - g \epsilon^{abc} A_k^b \right) \frac{\delta}{\delta A_k^c} \Psi = 0, \tag{12.18}$$

which is equivalent to requiring the invariance of $\Psi[A]$ under infinitesimal gauge transformations.

The Yang–Mills Schrödinger equation $H\Psi_0 = E_0\Psi_0$ is analytically soluble in two limits: the free-field limit $g \to 0$, in which the vacuum state is found to be

$$\Psi_0[A] = \exp\left[-\frac{1}{2} \int d^2x d^2y \, B^a(x) \left(\frac{\delta^{ab}}{\sqrt{-\nabla^2}} \right)_{xy} B^b(y) \right], \tag{12.19}$$

and also in the zero-mode limit, where one considers only gauge fields which are constant in space.[3] In this limit the Lagrangian is

$$L = \frac{1}{2} \int d^2x \left[\partial_t A_k \cdot \partial_t A_k - g^2 (A_1 \times A_2) \cdot (A_1 \times A_2) \right]$$

$$= \frac{1}{2} V \left[\partial_t A_k \cdot \partial_t A_k - g^2 (A_1 \times A_2) \cdot (A_1 \times A_2) \right] \tag{12.20}$$

---

[3] In quantum gravity this subspace of the set of all configurations is known as "minisuperspace."

with corresponding Hamiltonian

$$H = -\frac{1}{2}\frac{1}{V}\frac{\partial^2}{\partial A_k^a \partial A_k^a} + \frac{1}{2}g^2 V (A_1 \times A_2) \cdot (A_1 \times A_2), \tag{12.21}$$

where V is the volume of 2-space, and the cross-product and dot-product are defined with respect to $SU(2)$ color indices, i.e. $A_i = A_i^a \sigma_a/2$, and

$$(A_i \times A_j)^a = \varepsilon_{abc} A_i^b A_j^c. \tag{12.22}$$

In this case it can be shown that the ground state wavefunction is

$$\Psi_0 = \exp\left[-\frac{1}{2}gV \frac{(A_1 \times A_2) \cdot (A_1 \times A_2)}{\sqrt{|A_1|^2 + |A_2|^2}}\right]. \tag{12.23}$$

It is not hard to find an expression for $\Psi_0[A]$ which satisfies the Gauss Law constraint (12.18), and also agrees with both soluble limits of the Yang–Mills Schrodinger equation. It was suggested in [13] that

$$\Psi_0[A] = \exp\left[-\frac{1}{2}\int d^2x d^2y\, B^a(x)\left(\frac{1}{\sqrt{-D^2 - \lambda_0 + m^2}}\right)_{xy}^{ab} B^b(y)\right]. \tag{12.24}$$

could be a good approximation to the ground state wavefunctional in temporal gauge, where D^2 is the covariant Laplacian, λ_0 is the lowest eigenvalue of $-D^2$, and m^2 is a parameter which vanishes as $g \to 0$.[4] Agreement with the free-field expression in the $g^2 = 0$ limit is obvious. If we instead consider the limit in which the zero modes of the gauge field are much larger than all other modes, then

$$(-D^2)_{xy}^{ab} = g^2 \delta^2(x - y)\left[(A_1^2 + A_2^2)\delta^{ab} - A_1^a A_1^b - A_2^a A_2^b\right] \tag{12.25}$$

and m^2 is negligible. It is then found, after some algebra, that the proposed wavefunctional (12.24) reduces to the zero-mode solution (12.23).

The dimensional reduction form is obtained in the following way: Let

$$(-D^2)^{ab}\phi_n^b(x) = \lambda_n \phi^a(x) \tag{12.26}$$

denote the eigenvalue equation of the covariant Laplacian, and expand the field strength

$$B^a(x) = \sum_{n=0}^{\infty} b_n \phi_n^a(x). \tag{12.27}$$

---

[4]   The subtraction of λ_0 is introduced so that spectrum of $D^2 - \lambda_0 + m^2$ begins at m^2, rather than infinity in the continuum limit. Apart from this subtraction, the proposal is the same as an earlier suggestion by Samuel [14].

Let us define the "slowly varying" component of the B-field by inserting a cutoff in the mode sum

$$B^{a,\text{slow}}(x) = \sum_{n=0}^{n_{max}} b_n \phi_n^a(x) \tag{12.28}$$

such that $\lambda_{n_{max}} - \lambda_0 \ll m^2$. Then

$$\int d^2x d^2y \, B^{a,\text{slow}}(x) \left(\frac{1}{\sqrt{-D^2 - \lambda_0 + m^2}} \right)_{xy}^{ab} B^{b,\text{slow}}(y)$$

$$\approx \frac{1}{m} \int d^2x \, B^{a,\text{slow}}(x) B^{a,\text{slow}}(x), \tag{12.29}$$

and the part of the (squared) vacuum wavefunctional involving only $B^{a,\text{slow}}$ is

$$|\Psi_0|^2 = \exp\left[-\frac{1}{m} \int d^2x \, B^{\text{slow}} B^{\text{slow}} \right]. \tag{12.30}$$

This has the dimensional reduction form $\Psi_0^2 \sim e^{-S}$, where S is a two-dimensional gauge theory, and the string tension can be computed, in terms of m, analytically. For the $SU(2)$ gauge group this is

$$\sigma = \frac{3}{16} mg^2, \tag{12.31}$$

or, in terms of the lattice coupling $\beta = 4/g^2$, $\sigma = 3m/(4\beta)$. If we turn this around, and write $m(\beta) = 4\beta\sigma(\beta)/3$, then we have a complete proposal for the Yang–Mills vacuum wavefunctional, in 2 + 1 dimensions.

   To test this proposal, we need to use it to calculate some quantities which can or have been calculated by other means. The string tension is not a test, of course, it is an input. But we could look at other observables, such as the mass gap, computed from the connected equal-times correlator

$$\mathcal{G}(x - y) = \langle (B^a B^a)_x (B^b B^b)_y \rangle - \langle (B^a B^a)_x \rangle^2 \tag{12.32}$$

in the probability distribution

$$P[A] = |\Psi_0[A]|^2 = \exp\left[-\int d^2x d^2y \, B^a(x) K_{xy}^{ab}[A] B^b(y) \right], \tag{12.33}$$

where

$$K_{xy}^{ab}[A] = \left(\frac{1}{\sqrt{-D^2 - \lambda_0 + m^2}} \right)_{xy}^{ab}. \tag{12.34}$$

Numerically, the computation of $G(x - y)$ looks hopeless. Not only is the kernel $K_{xy}^{ab}[A]$ non-local, it is not even known explicitly for arbitrary $A_i^a(x)$, which would seem to rule out any lattice Monte Carlo approach. But suppose, after eliminating variance along gauge orbits by a further gauge choice (this is allowed in temporal gauge on a given time-slice at $t = 0$), that $K_{xy}^{ab}[A]$ has very little variance among thermalized configurations (i.e. the configurations which are generated by Monte Carlo simulation). In that case, the situation is more promising.

Let us define a probability distribution for gauge fields A which is controlled by a second, independent gauge field A'

$$P[A; K[A']] = \exp\left[- \int d^2x d^2y \, B^a(x) K_{xy}^{ab}[A'] B^b(y)\right], \qquad (12.35)$$

where B is computed from A, not A', and $P[A] = P[A;K[A]]$. Then, assuming the variance of K is small,

$$P[A] \approx P[A, \langle K \rangle]$$
$$= P\left[A, \int DA' K[A'] P[A']\right]$$
$$\approx \int DA' \, P[A, K[A']] P[A']. \qquad (12.36)$$

This expression is amenable to an iterative solution:

$$P^{(1)}[A] = P[A; K[0]]$$
$$P^{(n+1)}[A] = \int DA' \, P[A; K[A']] P^{(n)}[A']. \qquad (12.37)$$

The numerical simulation then proceeds in this way: Fix the gauge in the $D = 2$ time-slice to an axial $A_1 = 0$ gauge, and introduce a lattice regularization. Initially, set also $A_2 = 0$. Then

1. Set $A'_2(x) = A_2(x)$.
2. $P[A; K[A']]$ is gaussian in B. Diagonalize $K_{xy}^{ab}[A']$, and generate a new B-field stochastically.
3. Given B, calculate A_2 in the $A_1 = 0$ gauge, and compute observables.
4. Return to step 1, repeat as many times as necessary.

This procedure will generate configurations with probability distribution $P[A]$ if the kernel K_{xy}^{ab} converges to a fixed matrix (in lattice regularization) in the color-position indices, as the iterations proceed. The lattices which are generated in this way are called *recursion lattices*. Whatever observables are calculated in this way can be compared to a computation of the same (equal-time) observables on lattices generated by the usual Monte Carlo simulation in temporal gauge, with the remaining (time-independent) gauge symmetry used to fix to $A_1 = 0$ gauge on a
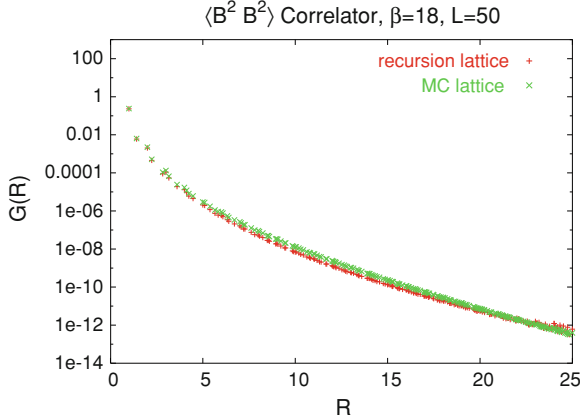
Fig. 12.1 The correlator $G(R)$ computed (i) on two-dimensional lattice configurations generated from the vacuum wavefunctional by the method described in the text; and (ii) on constant-time slices of three-dimensional lattice configurations generated by the usual lattice Monte Carlo method. Lattices generated by the first method are denoted "recursion," and by the second as "MC." In each case, the lattice extension is 50 sites at $\beta = 18$. From [13]

time-slice at $t = 0$. These lattices at time $t = 0$, are referred to as *MC* lattices, and are drawn from a probability distribution $|\Psi_0^{true}[A]|^2$, where Ψ_0^{true} is true lowest energy eigenstate of the lattice Hamiltonian.[5]

An important quantity to compute is the correlator

$$G(x - y) = \left\langle (K^{-1})_{xy}^{ab}(K^{-1})_{yx}^{ba} \right\rangle$$
$$K^{-1} = \sqrt{-D^2 - \lambda_0 + m^2},$$

(12.38)

which is of interest for two reasons. First of all, the underlying assumption of the numerical simulation is that the variance of K is negligible among recursion lattices; this means that the variance of $K^{-1}K^{-1}$ should also be negligible. Secondly, in the probability distribution governing recursion lattices,

$$\langle B^2(x)B^2(y) \rangle_{conn} \propto G(x - y),$$

(12.39)

and the mass gap is extracted from the exponential falloff of $G(R)$ with $R = |x - y|$. Figure 12.1 shows the results for $G(R)$ obtained from ten recursion lattices, and ten MC lattices. The two data sets match almost exactly, but what is equally important is that we find accurate values for $G(R)$ down to magnitudes on the order of 10^{-12}. The only way that this can be possible is that there is virtually no dispersion at all, from one lattice to the next, in the value of $G(x - y)$, and the same must therefore

---

[5]   More precisely, it is the eigenstate of the lattice transfer matrix $T = \exp[-Ha]$ with the highest eigenvalue.

be true for K_{xy}^{ab}ab. This absence of dispersion verifies the basic assumption under-lying the simulation method.

In order to extract the mass gap M as a function of β, $G(R)$ is fit to the analytic form ($R = |\mathbf{x} - \mathbf{y}|$)

$$G(R) = \delta^{ab}\delta^{ba}\left[\left(\sqrt{-\nabla^2 + (M/2)^2}\right)_{xy}\right]^2$$
$$= \frac{3}{4\pi^2}(1 + \frac{1}{2}MR)^2 \frac{e^{-MR}}{R^6},$$

(12.40)

which in fact fits the recursion lattice result very well, as shown in Fig. 12.2. The value of the mass gaps obtained are plotted in Fig. 12.3, and compared to lattice results for the mass gap obtained by Meyer and Teper [15] in three-dimensional $SU(2)$ lattice gauge theory by standard methods. The agreement is quite good.

It is also possible to use the vacuum wavefunctional to compute the ghost propagator in Coulomb gauge

Fig. 12.2 Best fit (dashed line) of the recursion lattice data for $G(R)$ by the analytic form given in eq. (12.40). From [13]
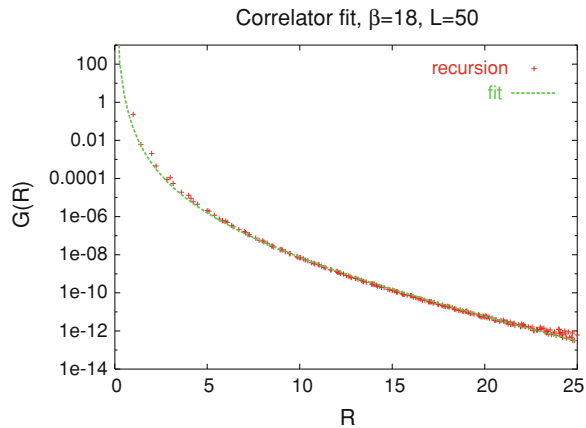
Fig. 12.3 Mass gaps extracted from recursion lattices at various lattice couplings, compared to the 0^+ glueball masses in $2 + 1$ dimensions obtained in ref. [15] (denoted "expt") via standard lattice Monte Carlo methods. Errorbars are smaller than the symbol sizes. From [13]
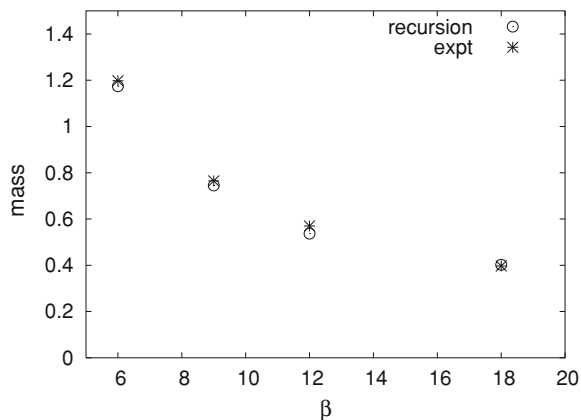
Fig. 12.4 The Coulomb ghost propagator evaluated on both recursion and MC lattices, at lattice coupling $\beta = 6$ and lattice extension $L = 24$ in $D = 3$ dimensions. From [16]

and the color Coulomb potential, which is proportional to
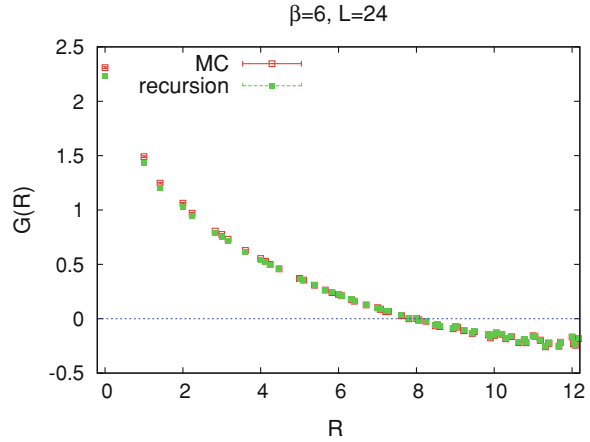
$$G_{gh}(x - y) = \left\langle \left(\frac{1}{-\nabla \cdot D} \right)^{aa}_{xy} \right\rangle, \tag{12.41}$$

and the color Coulomb potential, which is proportional to

$$\mathscr{V}(x - y) = -\left\langle \left(\frac{1}{\nabla \cdot D}(-\nabla^2)\frac{1}{\nabla \cdot D} \right)^{aa}_{xy} \right\rangle, \tag{12.42}$$

and to compare these with the values obtained from MC lattices [16]. In either case, the generated lattice configurations are transformed to Coulomb gauge, and the observables evaluated in the transformed configurations. The result for the ghost propagator, obtained from both recursion and MC lattices, is shown in Fig. 12.4. It is evident that there is almost no difference between the two data sets. The Coulomb potential is very sensitive to "exceptional" configurations with very small eigenvalues of the Faddeev-Popov operator $-\nabla \cdot D$; these lead to huge errorbars. To compare recursion and MC results, we impose cuts on the data, throwing away these rare configurations. The *same* cuts are applied to both the recursion and MC data. The aim here is not to get an accurate result for the Coulomb potential, since a cut on the data can obviously affect the outcome, but rather to see whether the recursion and MC lattices will yield the same value for the observable, when the same cut is applied to each set. The result is shown in Fig. 12.5. Obviously, the results for both the ghost propagator and Coulomb potential obtained from our proposed wavefunctional closely agree with those obtained by standard methods.

12.3 New Variables

Although the numerical results just presented suggest that the proposal (12.24) is on the right track, it is really only an educated guess for the vacuum

Fig. 12.5 Coulomb potential evaluated from data sets with a cut on lattices with $|V(0)| < -100$, for $\beta = 6$ and lattice extension $L = 24$. Results (with the same cut) are shown for both MC and recursion lattices. From [16]
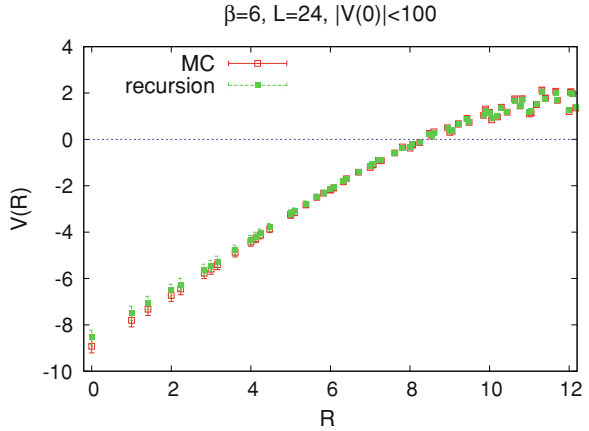
wavefunctional in $2 + 1$ dimensions, and the value of the string tension is an input, rather than emerging as an output. A much more ambitious program, initiated by Karabali, Kim, and Nair (KKN) in ref. [17, 19], aims at computing the $2 + 1$ dimensional wavefunctional, and the string tension, systematically. An essential feature of their approach is the transformation to a set of new, gauge-invariant variables.

KKN begin in temporal gauge and combine the remaining gauge fields A_1 and A_2 into a complexified gauge field

$$A \equiv A_z = \frac{1}{2}(A_1(x) + iA_2(x)), \quad \overline{A} \equiv A_{\overline{z}} = \frac{1}{2}(A_1(x) - iA_2(x)), \qquad (12.43)$$

where $z = x_1 - ix_2$, and $\overline{z} = x_1 + ix_2$ are the usual holomorphic variables in the complex plane. The complexified field A can be related to a matrix field M taking values in the group $SL(2, C)$,

$$A = -(\partial_z M)M^{-1}, \quad \overline{A} = M^{\dagger -1}\partial_{\overline{z}}M^{\dagger}, \qquad (12.44)$$

which transforms covariantly, $M \rightarrow GM$, under a gauge transformation G. Next, define the gauge-invariant variables

$$\begin{aligned} \mathscr{H} &= M^{\dagger}M \\ J &= \frac{C_A}{\pi} \frac{\partial \mathscr{H}}{\partial z} \mathscr{H}^{-1}, \end{aligned} \qquad (12.45)$$

where C_A is the quadratic Casimir in the adjoint representation. In terms of these gauge invariant variables, the Hamiltonian becomes

$$H_{KKN} = T + V, \qquad (12.46)$$

where T is derived from the E^2 term in the standard Hamiltonian

$$T = m\left(\int_u J^a(u) \frac{\delta}{\delta J^a(u)} + \int_{u,v} \Omega_{ab}(u,v) \frac{\delta}{\delta J^a(u)} \frac{\delta}{\delta J^b(v)} \right), \tag{12.47}$$

with

$$\Omega_{ab}(u,v) = \frac{C_A}{\pi^2} \frac{\delta_{ab}}{(u-v)^2} - if_{abc} \frac{J^c(v)}{\pi(u-v)} \tag{12.48}$$

and $(\overline{\partial} \equiv \partial_{\overline{z}})$

$$V = \frac{1}{2g^2} \int B^2(x) = \frac{\pi}{mC_A} \int \overline{\partial} J^a \overline{\partial} J^a \tag{12.49}$$

and also

$$m = \frac{g^2 C_A}{2\pi}. \tag{12.50}$$

Inner products are evaluated with respect to the integration measure

$$\langle \Psi_1 | \Psi_2 \rangle = \int d\mu(\mathcal{H}) e^{2C_A S_{WZW}(\mathcal{H})} \Psi_1^*(\mathcal{H}) \Psi_2(\mathcal{H}), \tag{12.51}$$

where $d\mu(\mathcal{H})$ is the standard integration measure for a matrix-valued variable (known as the *Haar measure*), and S_{WZW} is the Wess-Zumino-Witten action [20], whose precise form will not concern us here.

It is important to note that, although the theory is now expressed in terms of gauge-invariant variables $\mathcal{H}$ and J, a new local invariance, *holomorphic invariance*, has appeared in the problem. This is the transformation

$$M(z,\overline{z}) \rightarrow M(z,\overline{z})h^\dagger(\overline{z}), \quad M^\dagger(z,\overline{z}) \rightarrow h(z)M^\dagger(z,\overline{z}), \tag{12.52}$$

under which

$$\begin{aligned} H(z,\overline{z}) &\rightarrow h(z)H(z,\overline{z})h^\dagger(\overline{z}) \\ J &\rightarrow hJh^{-1} + \frac{C_A}{\pi}\partial h h^{-1}. \end{aligned} \tag{12.53}$$

It can be seen that J transforms under a holomorphic transformation $h(z)$ much like a gauge field under a gauge transformation. The Hamiltonian H_{KKN} is invariant under these local transformations, and it is crucial that the eigenstates of H_{KKN} are also invariant. In effect, by going to the new variables one trades gauge invariance (in temporal gauge formulation) for holomorphic invariance.

In these new variables, it is possible to carry out a systematic strong-coupling expansion in the continuum formulation. However, since the coupling g^2 has dimensions of mass and simply sets the scale, the question is what parameter

controls the convergence of the expansion. KKN suggest that the effective parameter is k/g^2 or (equivalently) k/m, where k is some characteristic momentum in the problem. To leading order, KKN find

$$
\begin{aligned}
\Psi_0 &= \exp\left[-\frac{\pi}{m^2 C_A}\int d^2x \mathrm{Tr}[\bar{\partial}J\bar{\partial}J]\right] \\
&= \exp\left[-\frac{1}{2mg^2}\int d^2x \mathrm{Tr}[B^2]\right].
\end{aligned}
\tag{12.54}
$$

In the usual variables, this ground state has the dimensional reduction form, and spacelike Wilson loops can be computed analytically. These have an area law falloff, and for Wilson loops in the fundamental representation of the $SU(N)$ gauge group, the string tension is

$$
\sigma = \frac{g^4}{8\pi}(N^2 - 1).
\tag{12.55}
$$

Very remarkably, this value is within a few percent of the value found by Teper and Bringoltz [18] in lattice Monte Carlo simulations of the the $2 + 1$ dimensional theory.[6]

On the other hand, we already know that the dimensional reduction form cannot be quite right. It leads to incorrect behavior at high momenta, and string tensions follow the Casimir scaling rule, rather than depending on N-ality. So the question is what is the effect of higher order terms in the perturbation series. KKN manage to sum up all terms in the series, for $\log(\Psi_0)$, which are bilinear in the J's. The result is

$$
\begin{aligned}
\Psi_0 &= \exp\left[-\frac{2\pi^2}{g^2 C_A^2}\int d^2x d^2y \bar{\partial}J^a(x)\left(\frac{1}{\sqrt{-\nabla^2 + m^2} + m}\right)_{xy}\bar{\partial}J^a(y)\right] \\
&= \exp\left[-\frac{1}{2g^2}\int d^2x d^2y\, B^a(x)\left(\frac{1}{\sqrt{-\nabla^2 + m^2} + m}\right)_{xy}B^a(y)\right],
\end{aligned}
\tag{12.56}
$$

which has the attractive feature of agreeing with the known result of the free theory at high momenta. The problem with this expression, however, is that in terms of the new variables (top line) it is not holomorphic invariant, and in terms of the old variables in temporal gauge (bottom line) it is not gauge invariant. Because physical states *must* be gauge invariant (this is demanded by the Gauss law constraint), it means that the KKN proposal is incomplete, in the sense that we don't really know what $\Psi_0[A]$ is equal to for an arbitrary gauge field configuration.

---

[6] Recently some corrections to σ have been calculated [19], and they are small. At present it is not entirely clear *why* the correction is so small, since it involves a sum of rather large (positive and negative) contributing terms, which for some reason nearly cancel.

There are an infinite number of ways in which (12.56) could be modified such that Ψ_0 becomes gauge-invariant, but reduces to (12.56) in the special case that the gauge field is abelian (i.e. $[A_i, A_j] = 0$). One possibility, suggested by KKN in their original paper, is that the neglected terms in the perturbation series, which involve higher powers of J, will sum up to convert the Laplacian operator ∇^2 to a Laplacian which is covariant under holomorphic transformations (when the wavefunctional is expressed in new variables), or covariant under gauge transformations, when the wavefunctional is converted to old variables. In that case

$$\Psi_0 = \exp\left[-\frac{1}{2g^2} \int d^2x d^2y \, B^a(x) \left(\frac{1}{\sqrt{-D^2 + m^2} + m}\right)_{xy} B^a(y)\right]. \quad (12.57)$$

In this form, the wavefunctional can be latticized and simulated by the technique outlined in the previous section. Unfortunately, when this calculation is performed [13], the resulting string tension differs substantially, by at least 50%, from the known value. The reason is easy to understand: Although the spectrum of $-\nabla^2$ begins at zero, in an infinite volume, the spectrum of the covariant Laplacian $-D^2$ begins at a positive value $\lambda_0 > 0$. Hence the approximation of simply dropping this operator when calculating large-scale Wilson loops, on the grounds that the operator is dominated at long wavelengths by its lowest eigenvalue—namely zero—is unjustified, as shown explicitly by the numerical calculation. This fact implies that the summation of the perturbation series, if done correctly, presumably does *not* lead to the form (12.57). But there are other reasonable gauge-invariant extensions of (12.56) which would avoid this problem, e.g.

$$\Psi_0 = \exp\left[-\frac{1}{2g^2} \int d^2x d^2y \, B^a(x) \left(\frac{1}{\sqrt{-D^2 - \lambda_0 + m^2} + m}\right)_{xy} B^a(y)\right], \quad (12.58)$$

in which case the KKN wavefunction is rather close in form to (12.24). At present, however, the gauge-invariant extension of (12.56) is not known, so the amplitude of an arbitrary vacuum gauge field is likewise unknown, and the proposal is, in this sense, incomplete.

An ingenious approach to deriving a holomorphic-invariant wavefunctional in the new variables was put forward by Leigh, Minic, and Yelnikov [21]. These authors start from the approximation

$$\Psi_0 = \exp\left[-\frac{\pi}{2C_A m^2} \int d^2x d^2y \, \bar{\partial}J^a(x) K_{xy}(L) \bar{\partial}J^a(y)\right], \quad (12.59)$$

where $L = -\Delta/m^2$, and Δ is the holomorphic-covariant Laplacian. They then derive and solve a differential equation for $K(L)$, where L is treated as a number, rather than an operator, and by solving this equation they arrive at

$$K(L) = \frac{1}{\sqrt{L}} \frac{J_2(4\sqrt{L})}{J_1(4\sqrt{L})}. \quad (12.60)$$

If the infrared limit means $L \to 0$, then $K \to 1$, and Ψ_0 has the dimensional reduction form (12.54), leading to a very accurate estimate of the string tension. The correct free-field limit is also obtained. By computing various equal-time correlators of the J field in the vacuum state (12.59), Leigh et al. obtain predictions for the glueball mass spectrum in $2 + 1$ dimensions, which appear to be in remarkable agreement with standard lattice Monte Carlo results.

Although the Leigh et al. approach seems quite successful, two cautionary remarks are in order. First, the differential equation for $K(L)$ is derived by assuming the validity of the operator identity

$$TO_n = (2 + n)mO_n, \tag{12.61}$$

where T is the kinetic term of the KKN Hamiltonian, and

$$O_n = \int \bar{\partial} J \Delta^n \bar{\partial} J. \tag{12.62}$$

This identity has not been proven, however. Secondly, the results for the correlators (and the string tension) are obtained by replacing the holomorphic covariant Laplacian by the ordinary Laplacian, and the effect of this replacement on the values of observables is not entirely clear.

References

1. Greensite, J.: Calculation of the Yang–Mills vacuum wave functional. Nucl. Phys. B **158**, 469–496 (1979)
2. Greensite, J.: Large scale vacuum structure and new calculational techniques in lattice SU(N) gauge theory. Nucl. Phys. B **166**, 113–124 (1980)
3. Halpern, M.B.: Field strength and dual variable formulations of gauge theory. Phys. Rev. D **19**, 517–530 (1979)
4. Greensite, J.: Computer measurement of the Yang–Mills vacuum wave functional in three-dimensions. Phys. Lett. B **191**, 431–436 (1987)
5. Greensite, J., Iwasaki, J.: Monte Carlo study of the Yang–Mills vacuum wave functional in D = 4 dimensions. Phys. Lett. B **223**, 207–212 (1989)
6. Arisue, H.: Monte Carlo measurement of the vacuum wave function for non-Abelian gauge theory in D = 3 dimensions. Phys. Lett. B **280**, 85–90 (1992)
7. Greensite, J., Olejnik, S., Zwanziger, D.: Coulomb energy, remnant symmetry, and the phases of non-Abelian gauge theories. Phys. Rev. D **69**, 074506-1–074506-18 (2004)
8. Szczepaniak, A.P., Swanson, E.S.: Coulomb gauge QCD, confinement, and the constituent representation. Phys. Rev. D **65**, 025012-1–025012-23 (2002) [arXiv:hep-ph/0107078]
9. Szczepaniak, A.P.: Confinement and gluon propagator in Coulomb gauge QCD. Phys. Rev. D **69**, 074031-1–074031-14 (2004) [arXiv:hep-ph/0306030]
10. Reinhardt, H., Feuchter, C.: On the Yang–Mills wave functional in Coulomb gauge. Phys. Rev. D **71**, 105002-1–105002-6 (2005) [arXiv:hep-th/0408237]
11. Feuchter, C., Reinhardt, H.: Variational solution of the Yang–Mills Schrodinger equation in Coulomb gauge. Phys. Rev. D **70**, 105021-1–105021-21 (2004) [arXiv:hep-th/0408236]
12. Szczepaniak, A.P., Matevosyan, H.H.: A model for QCD ground state with magnetic disorder. Phys. Rev. D **81**, 094007-1–094007-6 (2010) [arXiv:1003.1901 [hep-ph]]

13. Greensite, J., Olejnik, S.: Dimensional reduction and the Yang–Mills vacuum state in $2 + 1$ dimensions. Phys. Rev. D **77**, 065003-1–065003-15 (2008) [arXiv:0707.2860 [hep-lat]]
14. Samuel, S.: On the 0++ glueball mass. Phys. Rev. D **55**, 4189–4192 (1997) [arXiv:hep-ph/9604405]
15. Meyer, H.B., Teper, M.J.: Glueball Regge trajectories in $(2 + 1)$-dimensional gauge theories. Nucl. Phys. B **668**, 111–137 (2003) [arXiv:hep-lat/0306019]
16. Greensite, J., Olejnik, S.: Coulomb confinement from the Yang–Mills vacuum state in $2 + 1$ dimensions. Phys. Rev. D **81**, 074504-1–074504-8 (2010) [arXiv:1002.1189]
17. Karabali, D., Kim, C., Nair, V.P.: On the vacuum wave function and string tension of Yang–Mills theories in $(2 + 1)$-dimensions. Phys. Lett. B **434**, 103–109 (1998) [arXiv:hep-th/9804132]
18. Bringoltz, B., Teper, M.: A precise calculation of the fundamental string tension in SU(N) gauge theories in $2 + 1$ dimensions. Phys. Lett. B **645**, 383–388 (2007) [arXiv, hep-th/0611286]
19. Karabali, D., Nair, V.P., Yelnikov, A.: The Hamiltonian approach to Yang–Mills $(2 + 1)$: An expansion scheme and corrections to string tension. Nucl. Phys. B **824**, 387–414 (2010) [arXiv:0906.0783 [hep-th]]
20. Witten, E.: Global aspects of current algebra. Nucl. Phys. B **223**, 422–432 (1983)
21. Leigh, R.G., Minic, D., Yelnikov, A.: On the glueball spectrum of pure Yang–Mills theory in $2 + 1$ dimensions. Phys. Rev. D **76**, 065018-1–065018-23 (2007) [arXiv:hep-th/0604060]

Chapter 13
Anti-deSitter Space and Confinement

AdS/CFT, also known as the Maldacena conjecture, deserves a book in itself. The initials refer to an equivalence ("duality") between Type II superstrings propagating in a ten-dimensional background space, which is a product of anti-deSitter space (AdS) and some other manifold, with a conformal field theory (CFT). This duality, and its extension (AdS/QCD) to non-conformal, confining gauge theories, is one of the most important results obtained in the modern (i.e. post-1984) era of string theory, and the literature on this topic is vast. Unfortunately, the subject does require a substantial background in superstring theory and conformal field theory, both of which are beyond the scope of this book. Nevertheless, because of the great potential importance of AdS/CFT and AdS/QCD to non-abelian gauge theory and the confinement problem, this chapter aims to provide an inkling of the general idea, together with a sample calculation of Wilson loops via the AdS/CFT approach. The interested reader may wish to consult some of the recent texts on string theory, e.g. [1–3], or specialized review articles [4–6], for a deeper exposition of the subject.

Historically, string theory began as a theory of strings, which are one-dimensional objects propagating in time, but it is now recognized that the theory encompasses other dynamical objects, known as D p-branes (or "D-branes," for short), which are p-dimensional hypersurfaces in space, propagating in time. However, let's begin with strings. The first question is how to represent the motion of a string, and what action should be used, and here it is useful to back up a little further, and consider point particles. The motion of a point particle in $d + 1$ dimensions can be represented by a trajectory $\mathbf{x}(t)$, but this form treats the time coordinate $x^0 = t$ differently from the spatial coordinates x^i, $i = 1 \ldots d$. A more natural representation, from the point of view of relativity theory, is $x^\mu(\tau)$, $\mu = 0, 1 \ldots d$, where τ simply parametrizes the motion. The action of the trajectory $x^\mu(\tau)$ for a free particle of mass m is just the relativistically-invariant proper length L of the trajectory times the mass of the particle:

$$S_{particle} = -mL = -m \int ds$$

$$= -m \int d\tau \sqrt{-\partial_\tau x^\mu \partial_\tau x_\mu}. \tag{13.1}$$

The corresponding equations of motion, in flat Minkowski space,

$$p^\nu = \frac{m\partial_\tau x^\nu}{\sqrt{-\partial_\tau x^\mu \partial_\tau x_\mu}}, \quad \partial_\tau p^\nu = 0, \tag{13.2}$$

describe the motion of a free relativistic particle. Of course, there are an infinite number of ways of choosing the parameter τ, since for any *reparametrization* $\tau = \tau(\tau')$, the expressions $x^\mu(\tau)$ and $x^\mu(\tau') \equiv x^\mu(\tau(\tau'))$ describe the same trajectory. It is easy to check that the action $S_{particle}$ is invariant under arbitrary reparametrizations of τ.

A particle is a pointlike object, and traces out a trajectory, or "worldline" as it propagates in time. A string is a one-dimensional object, and therefore it traces out a two-dimensional manifold, or "worldsheet" as it propagates. To specify the location of the worldsheet in the $d + 1$-dimensional background (or "target") space requires two parameters, τ, σ, and it is useful to think of the coordinates $x^\mu(\tau, \sigma)$ of the worldsheet as a field in two-dimensional $\tau - \sigma$ space. The action of the corresponding two-dimensional field theory, i.e. the action of the worldsheet, is defined by analogy to the point particle. Introduce the notation $\sigma^0 = \tau$, $\sigma^1 = \sigma$ and

$$\dot{x}^\mu \equiv \frac{\partial x^\mu}{\partial \tau}, \quad x'^\mu \equiv \frac{\partial x^\mu}{\partial \sigma}. \tag{13.3}$$

Since the particle action is proportional to the length of a particle worldline, it is natural that the string action is proportional to the area of a particle worldsheet

$$S = -\mathscr{T} \times \text{Area}$$

$$= -\mathscr{T} \int d^2\sigma \sqrt{-\det[\gamma_{\alpha\beta}]}, \tag{13.4}$$

where the constant of proportionality $\mathscr{T}$ must have units of energy per unit length, or, in other words, $\mathscr{T}$ is a string tension. The metric $\gamma_{\alpha\beta}$ (with α, $\beta = 0$, 1) is the two-dimensional induced metric on the world sheet. This induced metric is simple to express in terms of the parametrized worldsheet $x^\mu(\tau, \sigma) = x^\mu(\sigma^\alpha)$. Suppose we have two nearby points on the worldsheet $x^\mu(\sigma^\alpha)$ and $x^\mu(\sigma^\alpha + d\sigma^\alpha)$. Then the squared distance between them is

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = g_{\mu\nu}\frac{\partial x^\mu}{\partial \sigma^\alpha}\frac{\partial x^\nu}{\partial \sigma^\beta}d\sigma^\alpha d\sigma^\beta$$

$$= \gamma_{\alpha\beta}[x(\tau, \sigma)]d\sigma^\alpha d\sigma^\beta, \tag{13.5}$$

where

$$\gamma_{\alpha\beta} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \sigma^\alpha} \frac{\partial x^\nu}{\partial \sigma^\beta}. \tag{13.6}$$

This induced metric gives us the Nambu action

$$S = -\mathscr{T} \int d\tau d\sigma \sqrt{(\dot{x})^2 (x')^2 - (\dot{x}^\mu x'_\mu)^2}. \tag{13.7}$$

which is the starting point of string theory.

The Nambu action is invariant under the reparametrizations

$$\tau \to \tau(\tau', \sigma'), \sigma \to \sigma(\tau', \sigma'), \tag{13.8}$$

which allow us to choose, if we wish, some special parametrization. This amounts fixing the reparametrization invariance by a special choice of (or conditions on) the coordinates on the worldsheet, and that choice, in analogy to what is done with ordinary gauge invariance, can be regarded as a choice of gauge. For the calculations in this chapter it is the static gauge, which identifies the x^0 coordinate of a point on the worldsheet with the parameter τ, that will be the most useful.

Strings can be open or closed, and in the case of open strings the equations of motion must be supplemented with boundary conditions. The endpoints of an open string can either be free to move in the d spatial dimensions (Neumann boundary conditions), or each end can be fixed to some definite point in space (Dirichlet boundary conditions). Neumann and Dirichlet boundary conditions may also be mixed, with p coordinates of an endpoint satisfying Neumann conditions, and the rest of the coordinates fixed by Dirichlet boundary conditions. More generally the endpoints of a string may be restricted to move only on a $p + 1$ dimensional hypersurface known as a *D p-brane*, as illustrated in Fig. 13.1.

If the degrees of freedom of a string are only the coordinates of the string in the $d + 1$-dimensional target space, as is the case for the Nambu action, then all of the quantum excitations of the string (whether open or closed) are bosonic. Because inner products of the vector operators x'^μ and $\dot{x}^\mu$ involve a spacetime metric $g_{\mu\nu}$ which has both positive and negative components, there is a danger of having negative norm states known as "ghosts." This situation is encountered in quantum electrodynamics, where the ghost states are eliminated by imposing a certain operator constraint, related to the Landau gauge condition, on physical states (the Gupta-Bleuler procedure). A similar strategy in string theory only works

Fig. 13.1 The endpoints of open strings may be restricted to lie on a D p-brane
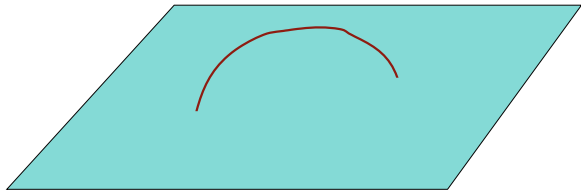
in $d + 1 = 26$ dimensions. In this critical dimension the spectrum of open bosonic strings (with Neumann boundary conditions) includes a tachyon, with $m^2 < 0$, massless vector particles which can be identified with abelian gauge bosons, and an infinite tower of massive states, lying along Regge trajectories (see Chap. 3) with slope $\alpha' = 1/(2\pi\mathscr{T})$. The spectrum of closed bosonic strings also includes a tachyon, as well as a massless spin-2 state that can be identified as the graviton, and another infinite tower of massive bosonic states. It is not hard to compute scattering amplitudes among these states up to one-loop level. The non-perturbative formulation of bosonic strings is problematic, however, because of the unphysical tachyon states.

Tachyon-free theories can be formulated by introducing fermionic degrees of freedom on the string worldsheet, i.e. there are fermionic fields $\psi(\tau, \sigma)$ as well as bosonic fields $x^\mu(\tau, \sigma)$ on the string worldsheet. Several consistent, tachyon-free string theories (type I, IIA, IIB and heterotic) with fermionic degrees of freedom are known; these are the superstring theories. While these theories were once considered to be entirely independent, they are now understood to be related by various duality transformations, and are believed to be only particular vacua of an underlying theory known as "M-theory." The theory relevant for AdS/CFT is the type IIB theory, which is formulated in terms of interacting closed strings, and is consistent (no negative norm states) in the critical dimension $d + 1 = 10$. If we take the string tension $\mathscr{T} \to \infty$ (or, equivalently, the zero Regge-slope limit $\alpha' \to 0$) in the type IIB theory, then the mass of the massive string excitations goes to infinity, and only the zero mass particles remain. These include the graviton, and some other massless fermions and bosons which fill out the supermultiplet of type IIB supergravity in ten dimensions.

D-branes enter the picture in an interesting and unexpected way. They are not merely allowable boundary conditions restricting the motion of open string endpoints, but turn out to be physical objects in their own right, with their own dynamics. There are several reasons for this. In the first place, string interactions include gravitational interactions, and having an object at a fixed point in space, insensitive to the passage of gravitational waves or any other gravitational phenomena, is simply inconsistent with general relativity. D-branes must, therefore, be dynamical. Secondly, the zero-slope limit of superstring theories, which are various types of supergravity theories, contain p-branes as classical, solitonic solutions of the equations of motion (much as the Georgi-Glashow theory contains magnetic monopoles), and these possess a certain type of charge which can be identified with the charge of D p-branes in string theory. Finally, it was discovered in the 1990s that all the known superstring theories with spatial directions compactified on a circle ("toroidal compactification") are related by various duality transformations, and D p-branes arise inevitably as a consequence of some of these transformations.

D-branes also have an interesting effective action at low energies [7]. Among the quantum states of open strings whose endpoints lie on D-branes are massless vector particles, i.e. photon excitations. Therefore, electromagnetic fields can live on D-branes. Let us identify the first $p + 1$ coordinates $\sigma^a = x^a$, $a = 0, 1 \ldots p$ as

coordinates along the brane worldvolume, and the remaining coordinates x^m, $m > p$ are in directions transverse to the brane. This is the static gauge for the brane. Then in a bosonic theory, the low-energy D-brane dynamics is described by the Dirac-Born-Infeld action

$$S_{DBI} = -\mathscr{T}_{brane} \int d^{p+1}\sigma \sqrt{-\det(\eta_{ab} + \partial_a x^m \partial_b x_m + 2\pi\alpha' F_{ab})}, \qquad (13.9)$$

where η_{ab} and F_{ab} are the Minkowski metric and electromagnetic field strength, respectively, on the brane, so the argument of the determinant is a $(p+1) \times (p+1)$ matrix. Expanding the square root determinant in powers of $\partial_a x^m$ and F_{ab}, one finds to leading order

$$S_{DBI} \approx -\frac{1}{4}\mathscr{T}_{brane}(2\pi\alpha')^2 \int d^{p+1}\sigma \left(F_{ab}F^{ab} + \frac{2}{(2\pi\alpha')^2} \partial_a x^m \partial^a x_m \right)$$
$$- \mathscr{T}_{brane}V_{p+1} + O(F^4), \qquad (13.10)$$

where V_{p+1} is the worldvolume of the p-brane. To this lowest order, the D-brane action is simply Maxwell electrodynamics plus some real scalar fields in $p + 1$ dimensions.

13.1 The Maldacena Conjecture

Consider some set of N D 3-branes whose locations exactly coincide. Each endpoint of an open string can lie on any one of the N D p-branes, which makes N^2 massless vector particles in all. In bosonic string theory, the low-energy dynamics of the brane turns out to be $U(N)$ gauge theory,[1] and in the case of superstrings it is $\mathscr{N} = 4$ super Yang–Mills theory in $3 + 1$ dimensions. The latter is a gauge theory with N colors and matter content such that the beta function is zero, which means that the theory is conformally invariant, and non-confining. Conformal invariance is a symmetry which implies, among other things, that there are no dimensionful constants in the theory, such as the QCD Λ-parameter, or the gauge coupling constant itself in $d + 1 \neq 4$ dimensions.

The relevant constants, in this setup of coinciding D 3-branes, are the following:

- The string tension (energy/unit length) of strings in type IIB string theory $\mathscr{T} = 1/(2\pi\alpha')$. This sets the scale for the masses of the massive string excitations.
- The string coupling constant g_s of type IIB string theory. This constant controls the strength with which strings interact with each other.

---

[1] Which is really $U(1) \times SU(N)$ gauge theory, since one gauge boson decouples from the rest.

- The coupling constant for the $\mathcal{N} = 4$ super Yang–Mills theory, $g^2 = 4\pi g_s$;
- The number of N of coinciding D 3-branes;
- The $d = 10$ dimensional Newton constant $16\pi G = (2\pi)^7 g_s^2 \alpha'^4$.

The AdS/CFT connection begins with an observation about the low-energy excitations of type IIB string theory, in $d = 10$ dimensions, containing a stack of coinciding D 3-branes. The observation is that there are two quite different descriptions of these low-energy excitations in two different limits, depending on the strength of the string coupling constant g_s. Let's begin with the weak-coupling limit. D-branes are massive, and since the low-energy limit of type IIB string theory is type IIB supergravity, these D-branes should (like any massive source), curve spacetime. However, there are N of these branes, and Newton's constant is proportional to g_s, so in the limit that $g_s N \ll 1$ the effect of the branes on the spacetime background should be negligible. Also, in the limit that g_s is small, strings hardly interact with each other. So at low energies the excitations in the theory are of two kinds (see Fig. 13.2):

1. the massless excitations of the strings tied to the D-branes, and their interactions, which is described by $\mathcal{N} = 4$ super Yang–Mills theory;
2. very weakly interacting gravitons, and their superpartners (which make up the spectrum of type IIB supergravity), propagating in flat $D = 10$ dimensional Minkowski space.

Now consider the opposite limit, with $g_s N \gg 1$, but still $g_s \ll 1$. Here the effect of the N D-brane sources on the background geometry cannot be ignored, and in this limit that effect can be treated semiclassically, just by solving the field equations of supergravity (a generalization of Einstein's equations) with the static

Fig. 13.2 First description, at $g_s N \ll 1$, of the low energy excitations in a ten-dimensional space containing a stack of N parallel D 3-branes. These are $\mathcal{N} = 4$ super Yang-Mills theory (from strings on the branes) and non-interacting massless particles (a supergravity multiplet) in the "bulk" 10-dimensional Minkowski space
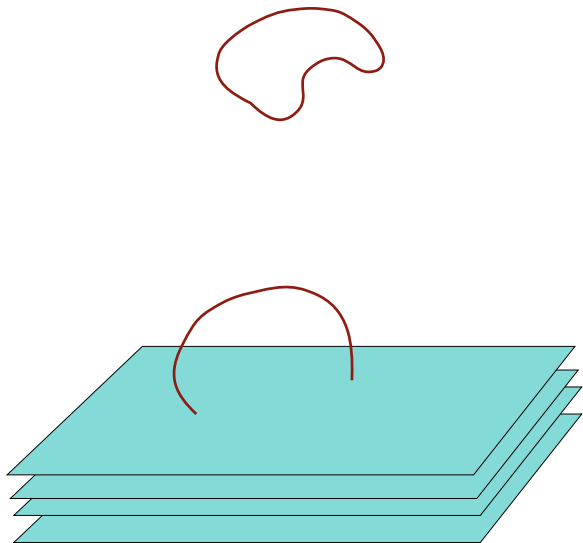
D-brane source. The solution for the $d = 10$ dimensional spacetime metric turns
out to be

$$ds^2 = f^{-1/2}\left(-dt^2 + \sum_{i=1}^{3} dx_i^2\right) + f^{1/2}(dr^2 + r^2 d\Omega_5^2)$$

$$f = 1 + \left(\frac{R}{r}\right)^4 \tag{13.11}$$

$$R^2 = \alpha'\sqrt{4\pi g_s N},$$

whose salient features are represented in Fig. 13.3. At $r \to \infty$, the geometry is
simply flat Minkowski space. But as $r \to 0$ (where the stack of D-branes are
located), the geometry develops a long "throat", and the circles around the throat
represent a five-dimensional sphere S^5, whose radius R, far down the throat, is
almost constant. The type IIB string excitations of this system, which have very
low energy with respect to an observer in the flat Minkowski region (i.e. at very
large r), are of two sorts. First, there are the massless excitations of the string in the
large r Minkowski region. On dimensional grounds, if the strings have energy E,
they would interact with strength GE^8, where G is the 10-dimensional Newton
constant. Therefore, the low-energy closed strings are essentially non-interacting,
and moreover, because their wavelengths are much greater than the radius of the
throat, they are insensitive to the excitations deep inside the throat region.
The other type of low energy excitations in this geometry, as seen by the observer
in the asymptotic region, are the string excitations deep within the throat region.
These excitations are so strongly red-shifted, with respect to the outside observer
in the asymptotic, Minkowski-space region, that they appear to have an arbitrarily
low energy, as seen by that observer, providing the string is sufficiently near the

Fig. 13.3 Second description, at $g_s N \gg 1$, of the low energy excitations in a ten-dimensional space with a p = 3 brane source. These are type IIB superstrings close to the $r = 0$ horizon, and non-interacting massless particles (a supergravity multiplet) in the asymptotic ten-dimensional Minkowski space
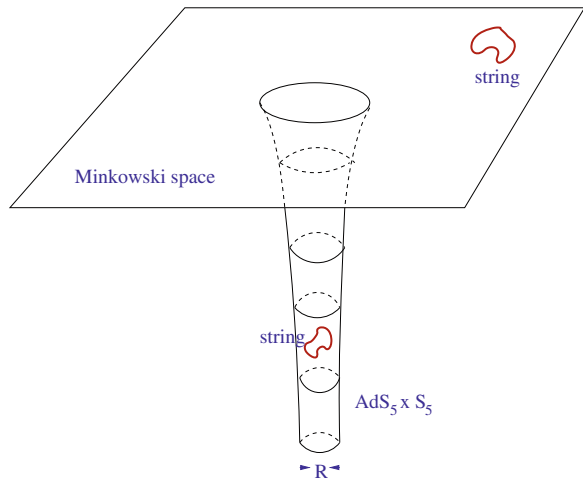
horizon at $r = 0$. Therefore, in this opposite $g_s N \gg 1$ limit, the low energy excitations of the theory again fall into two categories:

1. Type IIB superstrings close to the $r = 0$ horizon;
2. Very weakly interacting gravitons, and their superpartners, propagating in the asymptotically flat Minkowski space.

Both the $g_s N \ll 1$ and $g_s N \gg 1$ limits include weakly interacting gravitons in the low energy limit. However, the other class of low energy excitations is $\mathcal{N} = 4$ super Yang–Mills theory, at $g_s N \ll 1$, and type IIB superstrings in the near-horizon (or "throat") region of the metric in (13.11), at $g_s N \gg 1$. Maldacena [8] made the important conjecture that $\mathcal{N} = 4$ super Yang–Mills theory is equivalent ("dual") to type IIB superstrings propagating in the throat region, so that the latter theory, at $g_s N \gg 1$, is somehow describing $\mathcal{N} = 4$ super Yang–Mills theory at strong couplings. The metric in the throat region, far from the asymptotic Minkowski space region, is obtained by simply neglecting the constant 1 in the definition of $f(r)$, i.e.

$$ds^2 = \frac{r^2}{R^2}\left(-dt^2 + \sum_{i=1}^{3} dx_i^2\right) + \frac{R^2}{r^2}\left(dr^2 + r^2 d\Omega_5^2\right), \qquad (13.12)$$

which is the product of anti-deSitter space in five spacetime dimensions AdS_5 and the five-sphere S_5. Thus the conjecture: $\mathcal{N} = 4$ *super Yang–Mills theory in $3 + 1$ dimensions (which is a conformal field theory) is dual to type IIB string theory in an $AdS_5 \times S_5$ spacetime background*. This is actually just one particular example of the AdS/CFT correspondence; other dualities follow from other arrangements of D p-branes.

To make this conjecture precise, it is still necessary to specify what is meant by the correspondence; i.e. what does "duality" mean, exactly? The basic idea, first of all, is that to every string field $\Phi(x, r)$ (one of the string excitation modes such as the graviton, dilaton, etc.) there corresponds a local, gauge-invariant observable $\mathcal{O}(x)$ in the gauge theory. Let $\phi(x)$ be a source term for the operator $\mathcal{O}(x)$. The AdS/CFT correspondence then says that

$$\left\langle \int d^4x \phi(x) \mathcal{O}(x) \right\rangle_{CFT} = Z_{string}[\Phi(x, r = \infty) = \phi(x)], \qquad (13.13)$$

where the left hand side is the expectation value in the gauge theory, while the right hand side is the partition function for the string theory, with the string field $\Phi(x, r)$ held fixed and equal to $\phi(x)$ on the boundary. There is a kind of dictionary relating operators $\mathcal{O}(x)$ in $\mathcal{N} = 4$ super Yang–Mills theory to string fields $\Phi(x, r)$ on $AdS_5 \times S_5$, which we cannot go into here.

In order to use the AdS/CFT correspondence to calculate observables in the gauge theory, it is necessary to be able to solve the string theory in the $AdS_5 \times S_5$ background spacetime. This is a difficult thing to do, in general. In order to make the string theory tractable, we require that the string coupling constant is small, i.e. $g_s \to 0$, and that the radius of curvature R of $AdS_5 \times S_5$ spacetime is very large

compared to the length scale characteristic of strings, which is set by the Regge slope α'. In these two limits, the string theory is well approximated by its supergravity limit, which can be evaluated semiclassically, i.e.

$$Z_{string} \sim e^{-S_{sugra}}, \tag{13.14}$$

where S_{sugra} is the (on-shell) supergravity action. But then, since

$$\frac{R^4}{\alpha'^2} \sim g_s N \gg 1, \tag{13.15}$$

it means that this semiclassical approximation also requires taking the large-N limit. However, when one analyzes the leading (planar) diagrams in any $SU(N)$ gauge theory, it turns out that the effective coupling is not g^2, but actually $\lambda = g^2 N$, as discussed in Chap. 11. Because $g^2 \sim g_s$, the condition (13.15) means that a calculation on the string theory side in the supergravity limit provides information about the gauge theory in the strong 't Hooft coupling $\lambda \gg 1$ regime, which is not easily accessible by any other method.

13.2 Wilson Loops in AdS Space

Consider a loop C at the boundary of AdS_5 space, i.e. at $r = \infty$, and the worldsheet of a superstring which has C as its boundary. Let $S_{sheet}(C)$ be the action of this worldsheet at the saddlepoint (i.e. the minimal area string configuration). It was argued by Maldacena [9] that in the limit of large N and large 't Hooft coupling λ, where the semiclassical, supergravity approximation to string theory is valid, the gauge/string correspondence for Wilson loops would be

$$\left\langle \mathrm{Tr} P \exp[i \oint_C dx^\mu A_\mu] \right\rangle \approx \exp[-S_{sheet}(C)]. \tag{13.16}$$

Choosing rectangular $L \times T$ loops, with $T \gg L$, the static quark potential can be extracted from Wilson loops in the usual way. On the gauge theory side, one can calculate the left hand side of (13.16) perturbatively at small 't Hooft coupling $\lambda \ll 1$, and determine the potential $V(L) \sim -\lambda/L$. On the string theory side, the right hand side of (13.16) can be computed at large $\lambda \gg 1$, with the result that $V(L) \sim -\sqrt{\lambda}/L$. In either limit the potential is non-confining, as expected in a conformal theory such as $\mathcal{N} = 4$ super Yang–Mills. Since there is no dimensionful scale in a conformal theory, there is nothing to set the scale of the string tension; it can only be zero. Therefore, on dimensional grounds, the only possibility is that $V(L) \sim -1/L$. In order to use the gauge/string correspondence to investigate confining theories, conformal invariance has to be broken in some way, and a dimensionful scale introduced.

Conformal invariance can be broken in a number of ways, and one possibility, suggested by Witten [10], is to introduce a finite temperature. On the gauge theory side, a finite temperature can be introduced by going to a Euclidean metric signature and toroidally compactifying the time variable, so that the periodic time direction has some finite extension β. Then it is necessary to impose periodic boundary conditions on the bosons, and antiperiodic boundary conditions on the fermions. While this leaves some zero-energy bosonic modes, there are no zero energy fermion modes, and supersymmetry is completely broken. At scales much larger than the time compactification scale the theory is effectively three, rather than four dimensional, and the scalars acquire a large mass from one-loop quantum effects. Effectively, the $\mathcal{N} = 4$ super Yang–Mills theory will appear, at large scales, like a pure, non-supersymmetric Yang–Mills theory in three Euclidean dimensions, and this theory is, of course, confining.

On the string theory side it is also possible to go to a Euclidean signature and compactify the time direction. Then, as shown by Hawking and Page [11], there exist black hole solutions in AdS spacetime. There is a temperature on the string side as well: it is the Hawking temperature which is associated with any black hole horizon. The relevant metric is

$$ds^2 = \alpha' \left\{ \frac{U^2}{\widetilde{R}^2}[h(U)dt^2 + dx_i^2] + \frac{\widetilde{R}^2}{U^2}h^{-1}(U)dU^2 + \widetilde{R}^2 d\Omega_5^2 \right\}, \qquad (13.17)$$

where we have rescaled

$$U = \frac{r}{\alpha'}, \quad \widetilde{R} = \frac{R}{\sqrt{\alpha'}} \qquad (13.18)$$

and

$$h(U) = 1 - \frac{U_T^4}{U^4}, \quad \widetilde{R}^2 = \sqrt{4\pi g_s N}. \qquad (13.19)$$

The black-hole horizon is at $U = U_T$, and the corresponding Hawking temperature is $U_T/(\pi\widetilde{R}^2)$. The conjecture is that type IIB string theory in this background is dual to $\mathcal{N} = 4$ super Yang–Mills theory at high temperature, which at distance scales much larger than β should be equivalent to $D = 3$ dimensional (non-supersymmetric) Yang–Mills theory at zero temperature. Then the computation of a large Wilson loop in $D = 3$ Yang–Mills theory, at large 't Hooft coupling, can be solved by finding the minimal area of the string worldsheet bounded by the curve C at the boundary $U = \infty$ of the AdS space.

The calculation goes as follows [12]: Consider rectangular $L \times T$ contours C, which lie in the $x - x'$ plane (e.g. $x = x^1$, $x' = x^2$) at some fixed point in S_5, and at the boundary $U = \infty$ of the AdS space. It is convenient to adopt a static gauge in which worldsheet coordinates σ, τ are identified with the spacetime coordinates $\sigma = x, \tau = x'$. It is assumed that $T \gg L$, and the string is static in the sense that the transverse position $U(\sigma, \tau)$ of the worldsheet depends only on x. Using the

10-dimensional metric $g_{\mu\nu}$ from the line element (13.17) to contract spacetime indices in the Nambu action, and integrating over $\tau = x'$

$$S = \frac{T}{2\pi} \int_{-L/2}^{L/2} \frac{dx}{\sqrt{h(U)}} \sqrt{(\partial_x U)^2 + (U^4 - U_T^4)/\widetilde{R}^4}. \qquad (13.20)$$

Since the Lagrangian in (13.20) does not depend on x explicitly, there is a constant of motion (if we think of x as "time," the constant of motion is analogous to energy)

$$\frac{U^4 - U_T^4}{\sqrt{h(U)}\sqrt{(\partial_x U)^2 + (U^4 - U_T^4)/\widetilde{R}^4}} = \text{constant.} \qquad (13.21)$$

If we evaluate this constant at $x = 0$, where by $U(x) = U(-x)$ symmetry we have $\partial_x U = 0$, and let $U_0 = U(x = 0)$ denote the minimal value of $U(x)$ in the interval $x \in [-\frac{1}{2}L, \frac{1}{2}L]$, then

$$\frac{U^4 - U_T^4}{\sqrt{h(U)}\sqrt{(\partial_x U)^2 + (U^4 - U_T^4)/\widetilde{R}^4}} = \frac{\widetilde{R}^2}{\sqrt{h(U_0)}} \sqrt{U_0^4 - U_T^4}. \qquad (13.22)$$

Solving for $\partial_x U$, and writing

$$y = U/U_0, \quad \varepsilon = h(U_0) = 1 - \frac{U_T^4}{U_0^4}, \qquad (13.23)$$

one finds

$$x = \frac{\widetilde{R}^2}{U_0} \int_1^{U/U_0} \frac{dy}{\sqrt{(y^4 - 1)(y^4 - 1 + \varepsilon)}}, \qquad (13.24)$$

which determines $U(x)$ implicitly, given U_0. But $U_0 = U(0)$, while the ends of the string are at $x = \pm L/2$, where $U = \infty$, so that

$$L = 2\frac{\widetilde{R}^2}{U_0} \int_1^{\infty} \frac{dy}{\sqrt{(y^4 - 1)(y^4 - 1 + \varepsilon)}} \qquad (13.25)$$

gives U_0 as a function of quark separation L.

The energy of the static worldsheet $E = S/T$ is calculated from (13.20) by using the relationships (13.22), and changing integration variable from x to y via (13.24)

$$E(L) = \frac{U_0}{\pi} \int_1^{\infty} dy \frac{y^4}{\sqrt{(y^4 - 1)(y^4 - 1 + \epsilon)}}. \qquad (13.26)$$

It is clear that this energy is infinite, but we should recall that the expectation value of a Wilson loop in continuum gauge theory vanishes (and hence the static quark potential is infinite), due to the infinite self-energy of a static charge. In order to extract the L-dependence of the potential, it is necessary to discard the infinite, and L-independent, self-energy. Following Maldacena [9] we regularize by integrating y only up to $y_{max} = U_{max}/U_0$, and then subtract

$$
E_0 = \frac{U_T}{\pi} \int\limits_{1}^{U_{max}/U_T} dy 1
$$
$$
= \frac{U_{max} - U_T}{\pi},
$$
(13.27)

so that in the $U_{max} \to \infty$ limit

$$
E_{int} = E(L) - E_0
$$
$$
= \frac{U_0}{\pi} \int\limits_{1}^{\infty} dy \left(\frac{y^4}{\sqrt{(y^4 - 1)(y^4 - 1 + \varepsilon)}} - 1 \right) - \frac{U_0 - U_T}{\pi}.
$$
(13.28)

At any finite L there is a corresponding finite $U_0 > U_T$, with $U_0 \to U_T$ in the $L \to \infty$ limit, and with energies given by the expression in (13.26). Some typical solutions for $U(x)$ at various U_0/U_T, are shown in Fig. 13.4. Note that as $\varepsilon \to 0$ and $L \to \infty$ the string worldsheet flattens out along the horizon $U(x) = U_T$, in the finite interval $|x| < L/2$.

For small interquark separation L, the mimimal value U_0 is much larger than U_T, and $\varepsilon \approx 1$. In the $\varepsilon = 1$ limit all of the integrals can be carried out analytically,

Fig. 13.4 String contours $U(x)$ for various $\varepsilon = 1 - (U_T^4/U_0^4)$ and $\widetilde{R} = 1$, with units chosen such that $U_T = 1$. The asymptotes of each curve lie at $x = \pm L/2$. Note the approach to the horizon (here at $U = 1$), as $\varepsilon \to 0$
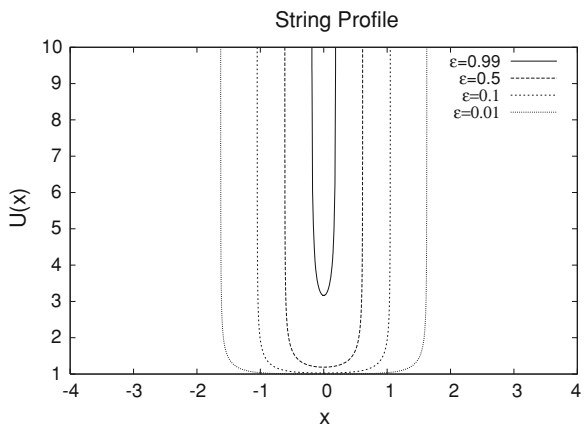
and $L \propto 1/U_0$, while $E_{int}(L) \propto -U_0$. From this, it is clear that $E_{int}(L) \propto -1/L$. The precise result is

$$E(L) = -\frac{1}{L}\frac{4\pi^2(2g^2N)^{1/2}}{\Gamma(\frac{1}{4})^4}. \qquad (13.29)$$

For large interquark separation, we have $U_0 \to U_T$ (and $\varepsilon \to 0$) as $L \to \infty$ for spacelike loops. Both L and $E(L)$ diverge, for spacelike loops, as $-\log(\varepsilon)$. The main contribution to the integrals in (13.25) and (13.26) comes from the integration region near $y = 1$, where the integrands are identical. Extracting a constant of proportionality by inspection, we find

$$E_{int}(L) = \frac{U_T^2}{2\pi\widetilde{R}^2}L. \qquad (13.30)$$

The result is easy to interpret: the coefficient of L is simply the action density for a worldsheet lying on the horizon, in the $x - x'$ plane, at $U = U_T$, i.e.

$$E_{int}(L) = \frac{1}{T}\lim_{U \to U_T} S[U, \partial_x U = 0] = \frac{U_T^2}{2\pi\widetilde{R}^2}L. \qquad (13.31)$$

This makes sense because, as is clear from Fig. 13.4, the minimal area string worldsheet falls from $U = \infty$ to $U = U_T$, and flattens out along the horizon as $L \to \infty$ ($\varepsilon \to 0$). As a result, the string tension for spacelike loops is essentially the action density of a worldsheet on the horizon.

13.3 AdS/QCD

We have arrived at a truly remarkable conclusion: confinement on the gauge theory side can be derived from the corresponding AdS black hole metric on the string theory side. The main requirement was to work with an AdS geometry, on the string side, dual to a non-conformal gauge theory, and the high temperature approach suggested by Witten is only one possibility. Other setups of branes in the $D = 10$ dimensional space, suggested by Polchinski and Strassler [13], Klebanov and Strassler [14], Maldacena and Nunez [15], DiVecchia et al. [16], and others, are also dual to confining gauge theories. The proposal by Sakai and Sugimoto [17], which is dual to a gauge theory including quarks, is perhaps the closest to realizing QCD itself. This general approach, which begins with a particular setup of D-branes and then works out the ten-dimensional geometry and the corresponding gauge theory, is sometimes referred to as the "top-down" approach to AdS/QCD. The "CFT" initials are no longer really appropriate, since the dual gauge theories are non-conformal and confining.

There is an alternative philosophy, known as the "bottom-up" approach to AdS/QCD. Here the idea is just to cook up a field theory, in some truncated version

of AdS space, from which one could "solve" QCD, using the field-operator correspondence. The fields that are taken to propagate in the AdS space are chosen based on the known properties of QCD, and the AdS geometry is modified (by hand) in such a way, for example by truncation of the space near $r = 0$, that the string worldsheet calculation will lead to an area law for Wilson loops. This is obviously a phenomenological, rather than fundamental approach, but it has had some impressive numerical successes. For an assessment of the promise and the drawbacks of this approach, cf. [18, 19].

Returning to the top-down models: do these prove confinement in non-abelian gauge theories? The answer is no, or at least, not yet. First of all, the Maldacena conjecture itself has not yet been proved, although there is a great deal of evidence in its favor (cf. the cited texts and review articles). Apart from that, the main limitation seems to be that a controlled calculation of the Yang–Mills string tension on the string theory side requires a strong 't Hooft coupling $\lambda = g^2 N \gg 1$. In this sense, the situation is reminiscent of the strong-coupling expansion in lattice gauge theory. In that case we can also derive confinement analytically, from first principles, but not asymptotic freedom, which can only be seen at weak lattice couplings. Relaxing the $\lambda = g^2 N \gg 1$ condition on the string side would require solving the string theory beyond the supergravity approximation, and this is a difficult task, perhaps as difficult (or more difficult) than solving the gauge theory directly, without recourse to duality. One advantage that the AdS/CFT approach does have, over the lattice strong-coupling expansion, is that the universal "stringy" corrections to the static quark potential (i.e. the Lüscher and higher-order terms observed in numerical simulations) arise in a natural way, already in the supergravity approximation [20].

The AdS/CFT approach remains under very active investigation, and an important recent application is to the study of gauge theories in the deconfined phase. It has been argued that strongly coupled $\mathcal{N} = 4$ super Yang–Mills theory at high temperatures, which can be treated via AdS/CFT, may be a reasonable approximation to non-supersymmetric gauge theories in the deconfinement regime. An important result, derived from the gauge/string duality, is the (very low) viscosity of the quark gluon plasma [21], which seems to be in accord with experimental results [22]. This is perhaps the first time that string theory has made contact with accelerator data, and demonstrates the potential power and importance of this very intriguing approach to the physics of gauge field theories.

References

1. Zwiebach, B.: A First Course in String Theory. Cambridge University Press, New York (2009)
2. Becker, K., Becker, M., Schwarz, J.H.: String Theory and M-Theory: A Modern Introduction. Cambridge University Press, New York (2007)
3. Kiritsis, E.: String Theory in a Nutshell. Princeton University Press, Princeton (2007)

4. Aharony, O., Gubser, S.S., Maldacena, J.M., Ooguri, H., Oz, Y.: Large N field theories, string theory and gravity. Phys. Rep. **323**, 183–386 (2000) [arXiv:hep-th/9905111]
5. Horowitz, G.T., Polchinski, J.: Gauge/gravity duality. arXiv:gr-qc/0602037
6. Mateos, D.: String theory and quantum chromodynamics. Class Quant. Grav. **24**, S713–S740 (2007) [arXiv:0709.1523 [hep-th]]
7. Leigh, R.G.: Dirac-Born-Infeld action from Dirichlet sigma model. Mod. Phys. Lett. A **4**, 2767–2772 (1989)
8. Maldacena, J.M.: The large N limit of superconformal field theories and supergravity. Adv. Theor. Math. Phys. **2**, 231–252 (1998) [Int. J. Theor. Phys. 38:1113 (1999)] [arXiv:hep-th/9711200]
9. Maldacena, J.M.: Wilson loops in large N field theories. Phys. Rev. Lett. **80**, 4859–4862 (1998) [arXiv:hep-th/9803002]
10. Witten, E.: Anti-deSitter space, thermal phase transition, and confinement in gauge theories. Adv. Theor. Math. Phys. **2**, 505–532 (1998) [arXiv:hep-th/9803131]
11. Hawking, S.W., Page, D.: Thermodynamics of Black Holes in anti-deSitter space. Commun. Math. Phys. **87**, 577–588 (1983)
12. Brandhuber, A., Itzhaki, N., Sonnenschein, J., Yankielowicz, S.: Wilson loops, confinement, and phase transitions in large N gauge theories from supergravity. JHEP **9806**, 001-1–001-19 (1998) [arXiv:hep-th/9803263]
13. Polchinski, J., Strassler, M.J.: The string dual of a confining four-dimensional gauge theory. arXiv:hep-th/0003136
14. Klebanov, I.R., Strassler, M.J.: Supergravity and a confining gauge theory: Duality cascades and chiral symmetry breaking resolution of naked singularities. JHEP **0008**, 052-1–052-35 (2000) [arXiv:hep-th/0007191]
15. Maldacena, J.M., Nunez, C.: Towards the large N limit of pure $N = 1$ super Yang–Mills. Phys. Rev. Lett. **86**, 588–591 (2001) [arXiv:hep-th/0008001]
16. Di Vecchia, P., Liccardo, A., Marotta, R., Pezzella, F.: The gauge/gravity correspondence for non-supersymmetric theories. Fortsch. Phys. **53**, 450–455 (2005) [arXiv:hep-th/0412234]
17. Sakai, T., Sugimoto, S.: Low energy hadron physics in holographic QCD. Prog. Theor. Phys. **113**, 843–882 (2005) [arXiv:hep-th/0412141]
18. Erlich, J.: How well does AdS/QCD describe QCD? Int. J. Mod. Phys. A **25**, 411–421 (2010) [arXiv:0908.0312 [hep-ph]]
19. Csaki, C., Reece, M., Terning, J.: The AdS/QCD correspondence: Still undelivered. JHEP **0905**, 067-1–067-33 (2009) [arXiv:0811.3001 [hep-ph]]
20. Aharony, O., Karzbrun, E.: On the effective action of confining strings. JHEP **0906**, 012-1–012-69 (2009) [arXiv:0903.1927 [hep-th]]
21. Policastro, G., Son, D.T., Starinets, A.O.: The shear viscosity of strongly coupled $N = 4$ supersymmetric Yang–Mills plasma. Phys. Rev. Lett. **87**, 081601-1–081601-4 (2001) [arXiv:hep-th/0104066]
22. Shuryak, E.: Why does the quark gluon plasma at RHIC behave as a nearly ideal fluid? Prog. Part. Nucl. Phys. **53**, 273–303 (2004) [arXiv:hep-ph/0312227]

Chapter 14
Concluding Remarks

The ideas surveyed in previous chapters do not exhaust all of the suggestions that have been made over the years, regarding the confinement mechanism. To include every proposal would have required an encyclopedia, not an introduction to the subject. I have included those ideas which seem to me to be the most promising, and the most likely to survive as part of our eventual understanding of confinement. But the selection comes without any guarantees. The solution of the confinement problem may eventually emerge from a further development of ideas which have been presented here. It may also come from some other direction entirely.

I have tried to stress, in this volume, a few main themes. Among these are the importance of carefully defining the problem that we are actually trying to solve, and the relevance of the center of the gauge group. The word "confinement" means different things to different people, and the existence of a color singlet spectrum and a mass gap is a feature not just of QCD, but also of Higgs theories having purely Yukawa-like forces. A color-singlet spectrum may have little to do with linear potentials and color electric flux tube formation, and if we are interested in the latter properties, it is better to focus on the asymptotic behavior of Wilson loops in the vacuum state. A vacuum state in which Wilson loops have an area-law falloff for arbitrarily large loops is qualitatively different, in its large-scale structure, from a vacuum state which does not have this property, and the symmetry which distinguishes between the two different types is an unbroken, and non-trivial, global center symmetry.

A strong clue about the nature of magnetic disorder responsible for the confining force is the dependence of the string tension on the color representation of the static charges. Asymptotically, the string tension depends only on the N-ality of the group representation, suggesting that vacuum fluctuations must arrange themselves, at large scales, in such a way that a Wilson loop is affected only via its N-ality. The existence of center vortices, or at least of domains whose color magnetic flux is quantized in units of the center subgroup, seems for this reason almost inevitable. At intermediate distance scales, extending to infinity as the

number of colors $N \to \infty$, the string tension appears instead to be proportional to the quadratic Casimir of the color charge representation. A theory of confinement, in terms of the field configurations which dominate the vacuum state, must ultimately account for not only the asymptotic dependence on N-ality, but also the Casimir scaling found at intermediate distances, as well as the transition between the two regimes.

Not every proposed explanation of confinement is based on the nature of vacuum field configurations (such as vortices, monopoles, and calorons). The Green's function approach, which relies on solutions of the Dyson–Schwinger equations, the Gribov–Zwanziger scenario, and especially the AdS/CFT approach, relating gauge theory to a dual theory of strings in 10-dimensional space–time, are important counter-examples. It is, of course, possible that confinement can be understood in more than one way, and it could happen that more than one approach will ultimately succeed. We would then expect the successful explanations to be related to one another, and somehow interdependent.

Confinement, in any case, is the hard problem of hadronic physics. It is also the hard problem for non-abelian gauge theories in general. Asymptotic freedom at short distances has been understood since the 1970's, but a similar grasp of the confining property of non-abelian gauge theories at large distances has so far eluded us. A number of ideas about confinement, introduced in previous chapters, seem well-motivated and promising in various degrees. But until non-abelian gauge theories are solved analytically, there is likely to be disagreement about the structure of the vacuum state, the origin of the mass gap, and the origin of the confining force. While there has been a great deal of progress, as outlined in this volume, the confinement problem is still open. As long as this problem remains open, a large gap will also remain in our understanding of the strongest of the known fundamental forces in Nature.

Index